The 51-state beeper M_{51}: every battle ends, and what Peano Arithmetic can prove about it

Self-contained proof document (Markdown with LaTeX). Apart from the two items listed below under "What lies outside this document", it depends on nothing else. Status: all results are proved in full. These are Theorem 1 (every battle ends, so M_{51} does not quasihalt), Theorem 2 (PA does not prove that the process beeps forever), Theorem 3 (quasihalting of M_{51} is independent of PA) and Corollary 8.1. The growth lemma that the unprovability half needs (LEMMA G) is proved in §6. §10 lists the dependencies.

What lies outside this document. Two things, and nothing else:

- the classification theorem of Buchholz–Wainer 1987 (Theorem 5), cited as Theorem A.13 and checked against the source in §A.8;
- the machine-checked simulation theorem (§2.3), which is a Rocq proof, distributed as `rocq51.tar.gz`. Its statements are quoted verbatim in §2.3, and every definition they depend on is listed and explained there. Only the truth of those statements is used.

The process R is given in full in §2.1, and the machine table of M_{51} is printed in Appendix B. The numerical checks of §6.11 and §9 are not part of any proof. Their scripts are distributed as `checks51.tar.gz`; file names below refer to that archive (or, for `.v` files, to `rocq51.tar.gz`).

Background facts about ordinals below ε_0, the Buchholz–Wainer fundamental sequences $\alpha[n]$, the Hardy functions H_α, the norm lemma and the classification theorem are collected, with proofs, in **Appendix A**. They are cited as "(F1)", "Lemma A.6" and so on. The only result used without proof is the classification theorem of Wainer and Buchholz–Wainer (Theorem A.13); §A.8 checks it against the source. Sources are listed in the References at the end.

hexagon:2611.00281v1

---

0. Results at a glance

Let R be the dispatch-level process `step` of §2.1, started at $[1,0]$, and M_{51} the 51-state, 2-symbol beeping machine of Appendix B (start state 30, beep state 50).

Theorem 1 (termination; proved). Every battle of R ends. Hence R beeps infinitely often, and so does M_{51}, by the machine-checked simulation theorem of §2.3 (Rocq, `M51_run30` / `M51_run_fun`, `Axioms: <none>`). So M_{51} **does not quasihalt.** The proof assigns to each list $[1,0]+X$ a worm ordinal $\mathrm{ord}(X) < \varepsilon_0$ and shows that every step lowers it (§4). The battle from $[1,0,k]$ starts at τ_{k+1}, a tower of k omegas.

Theorem 2 (unprovability for R; proved). The start letters k_j of the battles, and the beep times of M_{51}, eventually dominate every function that is provably computable in PA. PA does not prove the Π_2 sentence "R beeps infinitely often".

Theorem 3 (the machine; proved). (1) M_{51} does not quasihalt. (2) PA does not prove that M_{51} quasihalts. (3) PA does not prove that M_{51} does not quasihalt. So **quasihalting of M_{51} is independent of PA.** The only input about the machine is the truth of the Rocq statements. The time bound $\tau(j+2) \geq 2k_{j+1}+2$ is derived from the tape content at the dispatch before the beep (Corollary 2.3).

LEMMA G (proved, §6). For every $\beta < \varepsilon_0$, $\Psi(k) \geq H_\beta(k+1)$ for all large k. Here $\Psi(k)$ is the counter returned by the battle from $[1,0,k]$. §6 proves more:

- **Theorem 6.20.** $\Phi_{(k)}(z) \geq \frac{1}{2} H_{\tau_{k+1}}\big(\lfloor (2z+k+1)/(8(k+1)) \rfloor\big)$ whenever $2z+k+1 \geq 8(k+1)^2+1000$. Here $\Phi_{(k)}(z)$ is the counter returned by the battle of the one-letter worm (k) with counter z. So, at a large counter, this battle is at least the Hardy function of its own ordinal τ_{k+1}, at a linearly slowed argument.
- **Corollary 6.22.** $\Psi(k) \geq H_{\tau_{\lfloor k/2 \rfloor}}(k)$ for all large k.

The method (§6).

- Every list that R reaches is in a *normal form* (Lemma 6.9). Each ordinal has exactly one normal-form worm, and no worm with that ordinal is smaller (Lemma 6.4).
- On normal forms, each step of R removes the z trailing zeros of the list and then takes the largest ordinal below the ordinal of what remains whose normal-form worm fits into the new total size of the list (Proposition 6.12). The removed zeros are accounted for as z successor steps of the same descent (Theorem 6.14).
- Hence a battle is the longest strictly decreasing sequence of ordinals in which the term at each time has a worm of size at most that time (Theorem 6.14).
- Such sequences can be built along the Buchholz–Wainer fundamental sequences, regenerating time at each limit step (Theorem 6.19).

The exact structure laws of §§3–5 are used throughout. They are prefix intactness, the composition law $\Phi_{A0Y}(z) = \Phi_A(1 + \Phi_Y(z))$, lifts, and the counter lemmas.

Verdict. No flaw was found in the process or the machine. Every battle ends (Theorem 1), and quasihalting of M_{51} is independent of PA (Theorem 3).

---

1. Plain-language overview

The process. The process R manipulates a finite list of natural numbers that always starts with $1, 0$. The rest of the list, call it the *worm* X, changes at every step. The number z of zeros at the end of the list is a *counter*. If the worm is all zeros, the process *beeps* and starts a new *battle* from the one-letter worm (z), so the counter becomes the size of the next opponent. Otherwise the zeros at the end are removed, the last letter c is lowered by one, and a certain infinite periodic continuation e of the list is formed. From the continuation one reads off an infinite, ordered list of *candidates* $C_0, C_1, C_2, \ldots$ (initial pieces of e with their last letter possibly lowered), and the process jumps to candidate number $2z + 1$. The counter is not carried over: whatever zeros the new candidate happens to end with become the new counter.

What is proved here.

1. *Every battle ends* (Theorem 1). Each worm gets an ordinal below ε_0, a "worm ordinal" of the kind used in Beklemishev's worm calculus, and every step lowers it (§4). So R beeps infinitely often. A machine-checked Rocq theorem (§2.3) says that the 51-state Turing machine M_{51} carries out R, so M_{51} enters its beep state infinitely often.
2. *Counters never get lost* (§5). Although the counter is reset at every step, its value survives in the shape of the candidate. Three exact inequalities make this precise: a size bound, monotonicity in the counter, and a *doubling lemma*: lowering the last letter of a worm by one at least doubles the counter that the rest of the battle receives. From these: the battle from the one-letter worm (k) returns a counter $\Psi(k) \geq 2^{k+1} - 1$, and the j-th battle starts from a worm (k_j) with $k_{j+1} \geq 2^{k_j+1}$.
3. *How fast* Ψ *grows* (§6). PA-unprovability needs Ψ to outgrow every function that PA proves total. That is LEMMA G, proved in §6.
   - The counter is reset at every step, but the *total size* of the list is not; the size of an entry x is $x + 1$. A step with counter z adds exactly $z + 1$ to the total size.
   - Every ordinal has exactly one worm in *normal form*, and no worm with that ordinal is smaller. The run of R never leaves normal forms. On normal forms, the candidate chosen by R is the largest ordinal below the ordinal of the list without its trailing zeros whose normal form fits into the new total size.
   - So a battle is a *greedy* descent through the ordinals under a size budget that grows by one per unit of time. A greedy descent under such a budget is the longest possible one. To bound a battle from below, it is therefore enough to exhibit *one* long descent that respects the budget.
   - Such a descent is built along the Buchholz–Wainer fundamental sequences, slowed down so that the size budget is respected. At each limit step it first spends a cheap summand ω^2, which buys an exponential amount of time.

   This gives Hardy-function lower bounds at every level below ε_0.

4. *Unprovability and independence* (§§7–8). PA does not prove that R beeps forever, and whether M_{51} quasihalts is independent of PA. These theorems follow from LEMMA G, the machine-checked simulation, and the classification theorem of Wainer and Buchholz–Wainer.

Why this is harder than the Kirby–Paris hydra. In the Kirby–Paris hydra battle under their recursive strategy ([KP] p. 292), each step is exactly one fundamental-sequence step $\alpha \mapsto \alpha[c]$ (Definition A.1), and the stage counter c grows by one at every step and is never reset. Here neither holds. The candidate number $2z+1$ is compared with the *width* of the periodic part, which grows as the worm grows, so one step can descend much further than $\alpha[z]$ would. And the counter is reset at every step and rebuilt by the battles of sub-worms. A step-by-step comparison with a Buchholz–Wainer descent therefore breaks down. §6 avoids it. It never follows the battle step by step: it compares the battle with all budgeted descents at once (Theorem 6.14), and then builds one good descent by hand (Theorem 6.19).

---

2. The process R, the machine M_{51}, and the sentences

2.1 The process

The process is the following Python function `step` (the reference implementation `ref51.py`, verbatim):

```
def cand(S0, n):
    c = S0[-1]
    r = max(i for i in range(len(S0)) if S0[i] < c)
    Sp = S0[:-1] + [c - 1]
    B = Sp[r + 1:]
    def E(i):
        return Sp[i] if i < len(Sp) else B[(i - len(Sp)) % len(B)]
    idx = 0; L = len(S0) + 1
    while True:
        pre = [E(i) for i in range(L - 1)]
        for v in range(E(L - 1) + 1):
            if idx == n: return pre + [v]
            idx += 1
        L += 1
def step(S):
    z = 0
    while z < len(S) and S[len(S) - 1 - z] == 0: z += 1
    S0 = S[:len(S) - z]
    if S0 == [1]: return True, [1, 0, z]
    return False, cand(S0, 1 + 2 * z)
```

The run starts at $S_0 = [1,0]$ and S_{j+1} is the second component of `step`(S_j). Step j *beeps* if the first component is `True`. The Rocq file `Proc.v` transcribes these two functions line by line (§2.3), so the process of this section is literally the one in the machine-checked theorem.

Worm notation. Every S_j has the form $[1,0]$ followed by a finite sequence of naturals (by induction on j: see the description of a step below). A **worm** is a finite sequence $X = (x_1, \ldots, x_p)$ of naturals; $\varnothing$ is the empty worm, juxtaposition is concatenation, 0^m is a block of m zeros, and $|X| = p$. Write

$$\langle X, z\rangle := [1,0] + X + [0]^z.$$

Different pairs can denote the same list: $\langle X0, z\rangle = \langle X, z+1\rangle$. A worm is **stripped** if it is nonempty and its last letter is ≥ 1. Every list $[1,0]+X$ is either $\langle \varnothing, m\rangle$ (all zeros after $[1,0]$) or $\langle Y, z\rangle$ for a unique stripped Y and $z \geq 0$.

One step, in worm notation. Let $S = [1,0] + X$.

- *Beep.* If $X = 0^m$, the loop counts $z = m+1$ zeros (the 0 of $[1,0]$ counts too), $S_0 = [1]$, and the step beeps and returns $[1,0,m+1] = \langle (m+1), 0\rangle$.
- *Non-beep.* Otherwise $S = \langle Y, z\rangle$ with $Y = (y_1, \ldots, y_p)$ stripped, and the loop counts exactly the z final zeros (it stops at $y_p \geq 1$). Then $S_0 = [1,0] + Y$ and $c = y_p$. Put $y_0 := 0$, the 0 of $[1,0]$, which plays the

role of a *base letter*. In `cand`, $S_0[0] = 1$, $S_0[1] = y_0 = 0 < c$ and $S_0[i+1] = y_i$, so `r` $= r(Y) + 1$ with

$$r(Y) := \max\{\, i \in \{0, \dots, p-1\} : y_i < c \,\} \qquad \text{(well defined, since } y_0 = 0 < c\text{)}.$$

Write $Y^- := (y_1, \dots, y_{p-1}, c-1)$ (the last letter lowered),

$$B(Y) := (y_{r+1}, \dots, y_{p-1}, c-1) \quad (r = r(Y);\ |B(Y)| = p - r \geq 1),$$

and let $e = e(Y) = (e_1, e_2, \dots)$ be the infinite sequence $Y^- B(Y) B(Y) B(Y) \cdots$. Then `Sp` $= [1,0] + Y^-$, `B` $= B(Y)$, and `E(i)` is the i-th entry of $[1,0] + e$. In the loop, `L` $= i+3$ with $i \geq p$, `pre` $= [1,0] + (e_1, \dots, e_i)$ and `E(L-1)` $= e_{i+1}$. So the **candidates of** Y, in the order of the loop, are

$$C(i, v) := (e_1, \dots, e_i, v) \qquad \text{for } i = p, p+1, p+2, \dots \text{ and, within each } i,\ v = 0, 1, \dots, e_{i+1}.$$

Number them $C_0(Y), C_1(Y), \dots$ in this order. The step returns $[1,0] + C_{2z+1}(Y)$, without a beep. Call the candidates with the same i the **group** i. It has $e_{i+1} + 1$ members, and its last member is $(e_1, \dots, e_{i+1})$.

So, in short: $[1,0] + 0^m = \langle \varnothing, m \rangle$ beeps and is followed by $\langle (m+1), 0 \rangle$; and $\langle Y, z \rangle \mapsto [1,0] + C_{2z+1}(Y)$ for stripped Y.

Battles. A **battle** is the part of the run from a list $[1, 0, k] = \langle (k), 0 \rangle$ up to the next beep. By the first case, after every beep the run is at some $\langle (k), 0 \rangle$ with $k \geq 1$. The run starts with a beep: $S_0 = [1,0]$ beeps and $S_1 = [1,0,1]$. If the battle from $\langle (k), 0 \rangle$ ends (Theorem 1 shows that every battle does), it ends at a list $[1,0] + 0^m$, which beeps, and the next battle starts from $\langle (m+1), 0 \rangle$. Here $m \geq 2$, because every candidate has at least two letters ($i \geq p \geq 1$). Write

$$\Psi(k) := m \quad \text{(the number of zeros at the end of the battle from } \langle (k), 0 \rangle\text{)}, \qquad k_1 := 1, \quad k_{j+1} := 1 + \Psi(k_j).$$

Battle j ($j \geq 1$) starts from $\langle (k_j), 0 \rangle$. For example $C_1((1)) = (0,0,0)$, so the first battle is $[1,0,1] \to [1,0,0,0,0]$, which beeps: $\Psi(1) = 3$ and $k_2 = 4$. This matches the Rocq check `Sseq_check` ($S_0, \dots, S_6 = [1,0], [1,0,1], [1,0,0,0,0], [1,0,4], [1,0,3,1], [1,0,3,0,1], [1,0,3,0,0,0,0]$).

Beep indices. Let $b_1 < b_2 < \cdots$ be the indices j at which step j beeps. Then $b_1 = 0$, and for $j \geq 1$, $S_{b_{j+1}} = [1,0] + 0^{\Psi(k_j)}$ is the last list of battle j. In particular

$$S_{b_{j+1}} = \langle \varnothing,\ k_{j+1} - 1 \rangle. \tag{2.1}$$

2.2 The machine

M_{51} is the machine of Appendix B (file `beeper51.txt`): 51 states, 2 symbols, start state 30, beep state 50. It is Racheline's 51-state machine with one change: the original halting transition (state 50 reading 1) is replaced by `0L48`, the same as state 50 reading 0. All 102 transitions of `beeper51.txt` are defined (checked: every row has two entries of the form write–move–next), so M_{51} **never halts**. The table and the start states are used only through the Rocq development of §2.3, whose machine file `Machine51.v` is generated from `beeper51.txt`. Regenerating it (in `rocq51.tar.gz`: `python3 -I gen/gen_machine.py sources/beeper51.txt <out dir>`) gives a file byte-identical to `Machine51.v`. An independent parse of `Machine51.v` agrees with `beeper51.txt` entry by entry, all 102 entries.

Beeps of the machine. M_{51} *beeps at time* t if its configuration after t steps is in state 50. Since the start state is not 50, this means an *entry* into state 50 at step t. Let $\tau_{30}(1) < \tau_{30}(2) < \cdots$ be the beep times of the run from the blank tape in state 30, and $\tau_{31}(\cdot)$ those of the run started in state 31 on the blank tape (used as an intermediate object in the Rocq development).

2.3 The machine-checked simulation theorem

The connection between M_{51} and R is a theorem of the Rocq development `rocq51.tar.gz`. Its README states: complete, nothing admitted. The independent kernel checker, re-run separately for this document and by a referee, `rocqchk -o -Q . M51 M51.Main M51.ProcCheck` reports `Modules were successfully`

checked and Axioms: <none> (no type-in-type, no unsafe fixpoints, no assumed positivity). grep finds no Admitted, admit, Axiom, Abort or Parameter in the .v files.

Statements used (from Main.v, verbatim).

```
Corollary M51_run_fun (j : nat) :
  exists t, run delta t C31 = Some (Dconf (Sseq j)) /\ nbeeps_run delta t C31 = 1 + nbeep_before j.

Theorem M51_run30 (j : nat) :
  exists t, Steps delta isbeep t (1 + nbeep_before j) C30 (DconfJ (Sseq j) (tail30 (nbeep_before j))).
```

The definitions these statements rest on (the trusted base, besides the Rocq kernel). Each was read against its source.

- *Turing machines* (TM.v). A configuration is (q, L, x, R): state, the cells left of the head as an infinite stream listed outward, the symbol under the head, and the cells to the right likewise. One step is

  ```
  Definition step (C : Conf) : option Conf :=
    match C with
    | (q, L, x, R) =>
      match delta q x with
      | None => None
      | Some (w, Lft, q') => Some (q', stl L, shd L, w >> R)
      | Some (w, Rgt, q') => Some (q', w >> L, shd R, stl R)
      end
    end.
  ```

  so a step writes w in the head cell, moves the head one cell, and changes no other cell. run n C iterates step n times (None if a step is undefined). Steps n b C C' holds iff C reaches C' in exactly n defined steps and exactly b of the n configurations from which a step is taken are in a beep state (constructors steps0 C : Steps 0 0 C C and stepsS : step C = Some C' -> Steps n b C' C'' -> Steps (S n) (bp C + b) C C'', with bp C $= 1$ iff the state of C is a beep state). Steps n b C C' implies run n C = Some C' (Steps_run).

- *The machine* (Machine51.v, generated from beeper51.txt; see §2.2). Sym = s0 | s1 (s0 blank), states Q0…Q50, delta the 102 transitions, isbeep q true only for Q50.

- *The process* (Proc.v). tz, cand, step, Sseq, nbeep_before, pw, cells, enc, Dconf, C31. Compared tz, maxlt, Efun, candloop, cand and step with ref51.py line by line: tz S is the number of trailing zeros; step S returns (true, [1;0;z]) if the stripped list is [1], else (false, cand S0 (1+2z)); cand computes c, r $= \max\{i : S_0[i] < c\}$, Sp, B = skipn (S r) Sp and runs the candidate loop from L = |S0|+1, returning candidate number n. (The fuel $n+1$ of candloop is proved sufficient: candloop_returns, candloop_fuel.) ProcCheck.v also checks step, enc and Sseq against ref51.py by vm_compute on 500 random lists. Sseq 0 = [1;0], Sseq (S j) = snd (step (Sseq j)), so Sseq j $= S_j$ of §2.1. nbeep_before j is the number of $i < j$ such that step i beeps.

- *The encoding.* enc S = cells (pw (rev S)), where pw writes $S = [s_0, \ldots, s_m]$ as the pair word $X^{s_m} U X^{s_{m-1}} U \cdots U X^{s_0}$ and cells maps $X \mapsto$ s1;s0, $U \mapsto$ s1;s1. Dconf S = (Q22, blank, s0, enc S *> blank): state 22, head on a blank cell, blank to the left, enc S to the right, then blank. C31 = (Q31, blank, s0, blank) and C30 = (Q30, blank, s0, blank) (Main.v): the blank tape in state 31, resp. 30. DconfJ S J = (Q22, blank, s0, enc S *> J) (Cycle.v), and tail30 k is the stream of $11 - k$ blanks, then one s1 (the end marker written by the start-30 run), then blanks, if $k \le 10$, and all blank if $k \ge 11$.

- *Beep counts* (Main.v). nbeeps_run d n C is the number of the first n steps of the run from C that are taken from state 50. For a run that starts outside state 50 this is the number of times $t \in \{1, \ldots, n-1\}$ at which the configuration is in state 50.

Facts derived from the statements. The step function of TM.v is a function, so the configuration after t steps is unique; we write $\mathrm{Conf}_q(t)$ for it (start state $q \in \{30, 31\}$, blank tape).

(Sim1) For every j there is t such that $\mathrm{Conf}_{31}(t)$ = Dconf (Sseq j) and exactly 1+nbeep_before j beeps of the start-31 run occur at times $< t$ (M51_run_fun). For every j there is t such that $\mathrm{Conf}_{30}(t)$ = DconfJ

`(Sseq j) (tail30 (nbeep_before j))` and exactly 1+`nbeep_before j` beeps of the start-30 run occur at times $< t$ (`M51_run30` and `Steps_run`). In both cases the configuration at time t is in state 22, so it is not a beep.

(Sim2) *If R beeps infinitely often, so does M_{51}, from either start state.* By (Sim1) the number of beeps before the time at which `Sseq j` is displayed is 1+`nbeep_before j`, which is unbounded in j.

Lemma 2.1 (ones). Let `ones(C)` be the number of cells containing `s1` in a configuration C with finitely many such cells. Then for both start states, $\texttt{ones}(\mathrm{Conf}_q(t)) \le t$.

Proof. `ones(C30) = ones(C31) = 0`. By the definition of `step` quoted above, a step changes only the content of one cell (the old head cell, from `x` to `w`); every other cell keeps its content, and the head moves. So `ones` grows by at most 1 per step. □

Lemma 2.2 (ones of an encoding). `enc S` contains exactly $\sum_i s_i + 2(|S|-1)$ cells `s1`. In particular `enc` $([1,0]+0^m)$ contains $2m+3$ of them.

Proof. The pair word has $|S|-1$ letters U (two `s1` each) and $\sum_i s_i$ letters X (one `s1` each). For $[1,0]+0^m$: $\sum s_i = 1$ and $|S| = m+2$. □

Corollary 2.3 (beep times). Assume R beeps infinitely often (Theorem 1). For $q \in \{30, 31\}$ and every $j \ge 1$,

$$\tau_q(j+2) \;\ge\; 2k_{j+1}+2 \;>\; k_{j+1}.$$

Proof. Let $i = b_{j+1}$, the index of the beep that ends battle j. The beeps of R before step i are those at $b_1, \ldots, b_j$, so `nbeep_before i` $= j$. By (Sim1) there is a time t at which the start-q run shows `Sseq i` $= \langle \varnothing, k_{j+1}-1\rangle$ (by (2.1)), with exactly $j+1$ beeps at times $< t$ and none at time t. By Lemma 2.2 the tape then contains at least $2(k_{j+1}-1)+3 = 2k_{j+1}+1$ cells `s1` (for $q=30$ the marker in `tail30` can only add one), so $t \ge 2k_{j+1}+1$ by Lemma 2.1. The $(j+2)$-th beep exists by (Sim2) and comes after time t. So $\tau_q(j+2) \ge t+1$. □

Remarks. (a) The development also proves `M51_beeps_forever_iff` (M_{51} from state 31 beeps infinitely often iff R beeps infinitely often), `M51_run_times` (the display times of $S_0, S_1, \ldots$ increase strictly), and two facts about Racheline's unmodified table (`orig_never_halts`, `M51_state50_reads_0`). None of them is needed below. (b) Only the *truth* of the Rocq statements is used, never their provability in PA. (c) The Rocq statements say nothing about the growth of Ψ; all growth facts below are about R alone.

2.4 The sentences

The language of PA is $\{0, S, +, \times\}$. Fix the Cantor pairing $\langle x, y\rangle = \frac{1}{2}(x+y)(x+y+1)+x$ (so $\langle x,y\rangle \ge x, y$) and Gödel's $\beta(a,b,i) = a \bmod (1+(i+1)b)$ (so $\beta(a,b,i) \le a$), as is standard. A finite list $L = (l_0, \ldots, l_s)$ is coded by any $\langle s, \langle a, b\rangle\rangle$ with $\beta(a,b,i) = l_i$ for $i \le s$; every code of L is $\ge$ every entry of L and $\ge s$.

Runs of R. $\mathrm{Run}_R(j,w)$ is the Δ_0 formula expressing: w codes a list $(D_0, \ldots, D_s)$ of records; each record D_i codes the list S_i, a *running beep count* n_i, and the auxiliary data of step i (its z, r, the first $|C|$ entries of e, and the partial sums of the group sizes up to the chosen candidate); $S_0 = [1,0]$; for $i < s$ the data of step i are correct and S_{i+1} is the result; $n_0 = 1$ and $n_{i+1} = n_i + [\text{step } i+1 \text{ beeps}]$; $n_s = j$ and step s beeps. Every condition is a check of finitely many entries against each other with all quantifiers bounded by w. Storing the count and the intermediate sums in the witness keeps the formula Δ_0 without any Δ_0 counting or summation (counting the beeps inside a Δ_0 formula would otherwise need a separate argument). By determinism of `step`:

- **(R1)** For $j \ge 1$, $\exists w\, \mathrm{Run}_R(j,w)$ holds iff R has a j-th beep, and then every witness codes $S_0, \ldots, S_{b_j}$. For $j \ge 2$ it contains a code $\langle s, \langle a, b\rangle\rangle$ of $S_{b_j} = \langle \varnothing, k_j - 1\rangle$, a list of length $k_j + 1$ (by (2.1)), so $s = k_j$. This list contains the entry 1, so $a \ge 1$ and $\langle a, b\rangle \ge 1$; hence the code is $\langle k_j, \langle a, b\rangle\rangle \ge k_j + 1$, since $\langle x, y\rangle \ge x + 1$ when $y \ge 1$. A witness is $\ge$ every code it contains, so **every witness is $\ge k_j + 1$.**

Put $\theta_R(j,w) :\equiv (j = 0 \wedge w = 0) \vee \mathrm{Run}_R(j,w)$ and $\mathsf{Inf}_R :\equiv \forall j\, \exists w\, \theta_R(j,w)$, "$R$ beeps infinitely often", a Π_2 sentence.

Runs of M_{51}. $\mathrm{Run}_M(k,w)$ is the Δ_0 formula expressing: w codes records $D_0,\dots,D_s$, where D_t codes the configuration of M_{51} at time t (state, head position, and the tape on the window $[-t,t]$ around the start cell) and a running count m_t of the times $\le t$ in state 50; D_0 is the blank tape in state 30 with $m_0=0$; each D_{t+1} follows from D_t by the table; $m_{t+1}=m_t+[\text{state at } t+1 \text{ is } 50]$; $m_s=k$; the state at time s is 50. Then:

- **(R2)** For $k\ge 1$, $\exists w\,\mathrm{Run}_M(k,w)$ holds iff M_{51} has a k-th beep, and then every witness has $s=\tau_{30}(k)$, so **every witness is $\ge \tau_{30}(k)$.**

A *run code of length t* is a code of records $D_0,\dots,D_t$ exactly as in Run_M (configurations and running state-50 counts, D_0 the start, each record following from the previous one), without the two final conditions. Let $B(t)$ be the Σ_1 formula "there is a run code of length t whose state at time t is 50", and $\mathsf{QH} :\equiv \exists t_0\,\forall t\ge t_0\,\neg B(t)$, "$M_{51}$ quasihalts" (from state 30), a Σ_2 sentence.

Lemma 2.4. $\mathrm{PA}\vdash \neg\mathsf{QH}\to\forall k\,\big(k=0\vee\exists w\,\mathrm{Run}_M(k,w)\big)$.

Proof (in PA). First, any two run codes agree at every time where both are defined: induction on the time, since the table is a function. Also every run code of length t can be truncated to a run code of any length $t'\le t$ (the β-function lemma, provable in PA, gives a code of the first $t'+1$ records). Now assume $\neg\mathsf{QH}$ and prove the claim by induction on k; the formula is Σ_1. For $k=0$ it is trivial. For the step, put $t^*=0$ if $k=0$; otherwise let w witness $\mathrm{Run}_M(k,w)$, with end time s, and put $t^*=s+1$. By $\neg\mathsf{QH}$ there is $t\ge t^*$ with $B(t)$; let w' be a run code of length t ending in state 50. By the bounded least-number principle let t' be the least time in $[t^*,t]$ at which w' is in state 50. Truncate w' at t'. If $k\ge 1$, then by agreement with w the running count of w' at s is k; if $k=0$, there is no state-50 time before t' at all (time 0 is in state 30). There are no state-50 times in $[t^*,t')$, so the count at t' is $k+1$, and the truncation witnesses $\mathrm{Run}_M(k+1,\cdot)$. □

This is the only direction used. It needs neither the existence of long runs nor non-halting inside PA. (Non-halting is true and easy anyway: all 102 transitions are defined.)

---

3. The structure of one step

Throughout this section $Y=(y_1,\dots,y_p)$ is stripped, $c=y_p$, $r=r(Y)$, $B=B(Y)$, $e=e(Y)$, as in §2.1. The **size** of a worm is $\mathrm{size}(X)=\sum_i(x_i+1)=|X|+\sum_i x_i$.

Lemma 3.1 (the candidate list).

- (a) Every candidate of Y has $Y^-=(e_1,\dots,e_p)$ as a prefix and has at least $p+1\ge 2$ letters.
- (b) *Consecutive candidates.* Let $C_n=(e_1,\dots,e_i,v)$. If $v<e_{i+1}$ then $C_{n+1}=(e_1,\dots,e_i,v+1)$; if $v=e_{i+1}$ then $C_{n+1}=(e_1,\dots,e_{i+1},0)=C_n0$.
- (c) *Size.* $\mathrm{size}(C_n(Y))=\mathrm{size}(Y)+n$.
- (d) *The first group.* $C_v(Y)=Y^-v$ for $0\le v\le e_{p+1}=B_1$ (the first letter of B).
- (e) *Expansions.* For $j\ge 0$ let $Y[j]:=(e_1,\dots,e_{r+(j+1)|B|})$, that is, Y^- followed by j more copies of B (note $Y^-=(y_1,\dots,y_r)B$, so $Y[0]=Y^-$). Put $W(Y):=\mathrm{size}(B)$. For $j\ge 1$, $Y[j]$ is the candidate $C_{jW(Y)-1}(Y)$.

Proof. (a) $i\ge p$ and $(e_1,\dots,e_p)=Y^-$ by definition of e. (b) Within group i, v runs through $0,\dots,e_{i+1}$; after its last member $(e_1,\dots,e_{i+1})$ comes the first member $(e_1,\dots,e_{i+1},0)$ of group $i+1$. (c) If $C_n=(e_1,\dots,e_i,v)$, the candidates before it are the e_t+1 members of each group $t-1$ for $t=p+1,\dots,i$, and v members of group i. So $n=\sum_{t=p+1}^{i}(e_t+1)+v$, and $\mathrm{size}(C_n)=\sum_{t=1}^{i}(e_t+1)+(v+1)=\mathrm{size}(Y^-)+n+1=\mathrm{size}(Y)+n$. (d) Group p is $Y^-0,\dots,Y^-e_{p+1}$, and e_{p+1} is the first entry after Y^-, which is B_1. (e) $Y[j]$ is the last member of group $i=r+(j+1)|B|-1\ge p$ (as $j\ge 1$). By (c), its index is $\mathrm{size}(Y[j])-\mathrm{size}(Y)=\mathrm{size}(Y^-)+j\,\mathrm{size}(B)-\mathrm{size}(Y)=jW(Y)-1$. □

($Y[j]$ is the expansion of Y with $j+1$ copies of B in Beklemishev's worm calculus. By (e) it is one of the candidates; R usually takes a candidate *between* two such expansions.)

Lemma 3.2 (prefix intactness). Let A be any worm and Y stripped. Then $r(A0Y)=|A|+1+r(Y)$, $B(A0Y)=B(Y)$, $(A0Y)^-=A0Y^-$ and $e(A0Y)=A\,0\,e(Y)$. Hence $C_n(A0Y)=A\,0\,C_n(Y)$ for every n, and

$(A0Y)[k] = A\,0\,(Y[k])$ for every $k \geq 0$ (Lemma 3.1(e)). Hence for every z, the step from $\langle A0Y, z\rangle$ goes to $[1,0] + A + [0] + C_{2z+1}(Y)$, and the step from $\langle Y, z\rangle$ goes to $[1,0] + C_{2z+1}(Y)$.

Proof. Write $X = A0Y = (x_1, \ldots, x_q)$ with $q = |A| + 1 + p$ and $x_0 := 0$. The last letter is c, and $x_{|A|+1} = 0 < c$. So $r(X) \geq |A| + 1$, and for $i > |A| + 1$ we have $x_i < c$ iff $y_{i-|A|-1} < c$, where position $|A| + 1$ of X corresponds to the base $y_0 = 0$ of Y. Hence $r(X) = |A| + 1 + r(Y)$ and $B(X) = B(Y)$. Also $X^- = A0Y^-$, so $e(X) = A\,0\,e(Y)$. The candidates of X are $(e(X)_1, \ldots, e(X)_i, v)$ for $i \geq q$ and $v \leq e(X)_{i+1}$, that is, $A0(e_1, \ldots, e_{i'}, v)$ with $i' = i - |A| - 1 \geq p$ and $v \leq e_{i'+1}$, in the same order. Finally $(A0Y)[k] = (A0Y)^- B(A0Y)^k = A0Y^- B(Y)^k = A\,0\,(Y[k])$. □

Lemma 3.3 (last letter 1). Let P be zero-free (possibly empty) and $f = (P0)(P0)(P0)\cdots$. List the **tails** $T_0(P), T_1(P), \ldots$ as $(f_1, \ldots, f_i, v)$ for $i = 0, 1, 2, \ldots$ and, within each i, $v = 0, \ldots, f_{i+1}$. Then $C_n(P1) = P\,0\,T_n(P)$ for every n, and $\mathrm{size}(T_n(P)) = n + 1$.

Proof. All letters of P are $\geq 1 = c$, so $r(P1) = 0$ and $B(P1) = (P, 0) = (P1)^-$. Hence $e(P1) = (P0)(P0)\cdots$, whose first $|P| + 1 = |P1|$ entries are $P0$ and the rest is f again. The candidates are $(e_1, \ldots, e_i, v)$ with $i \geq |P| + 1$, that is, $P0$ followed by $(f_1, \ldots, f_{i'}, v)$ with $i' = i - |P| - 1 \geq 0$, $v \leq f_{i'+1}$, in the same order. For the size: $\mathrm{size}(P1) + n = \mathrm{size}(P0T_n) = \mathrm{size}(P) + 1 + \mathrm{size}(T_n)$ by Lemma 3.1(c), and $\mathrm{size}(P1) = \mathrm{size}(P) + 2$. □

For example $T_v(P) = (v)$ for $v \leq p_1$ (the first letter of P), and $T_n(\varnothing) = 0^{n+1}$, so $C_n((1)) = 0^{n+2}$.

Lemma 3.4 (lifts). For a worm X let $X + 1$ ($X - 1$) add (subtract) 1 to (from) every letter. Let Y be zero-free with $c \geq 2$. Then $Y - 1$ is stripped, $r(Y - 1) = r(Y)$, $B(Y - 1) = B(Y) - 1$ and $e(Y - 1) = e(Y) - 1$. Each group i of Y consists of its **extra** $(e_1, \ldots, e_i, 0)$, followed by $C' + 1$ for the members C' of group i of $Y - 1$, in order.

Proof. The last letter of $Y - 1$ is $c - 1 \geq 1$. For $1 \leq i < p$: $y_i - 1 < c - 1$ iff $y_i < c$; and the base $0 < c - 1$. So $r(Y - 1) = r(Y)$, and the rest follows from the definitions. Group i of Y is $(e_1, \ldots, e_i, v)$ for $v = 0, \ldots, e_{i+1}$; for $v \geq 1$ this is $(e_1 - 1, \ldots, e_i - 1, v - 1) + 1$, and $v - 1$ runs through $0, \ldots, e_{i+1} - 1$, which are the values allowed in group i of $Y - 1$. □

---

4. Worm ordinals, and Theorem 1 (termination)

Ordinals below ε_0 are written in Cantor normal form, compared by (F2), and added by ordinary ordinal addition; $\tau_0 = 0$, $\tau_{s+1} = \omega^{\tau_s}$ ((F6)).

4.1 Definition

Every worm W is uniquely $W = A_0\,0\,A_1\,0\cdots 0\,A_m$ with zero-free (possibly empty) worms A_i; m is the number of zeros of W. The A_i are the **blocks** of W, and A_0 is its **first block**. By recursion on the largest letter define, for zero-free A,

$$\theta(A) := o(A - 1) \qquad (\theta(\varnothing) = o(\varnothing) = 0),$$

and for every worm $W = A_0\,0\cdots 0\,A_m$

$$o(W) := [A_0 \neq \varnothing] \cdot \omega^{\theta(A_0)} \;+\; \omega^{\theta(A_1)} + \cdots + \omega^{\theta(A_m)}$$

(left-to-right ordinal sum; the first block is **counted** only if it is nonempty, the others always). Finally the **ordinal of a worm** is

$$\mathrm{ord}(X) := o(0X) = \omega^{\theta(A_0)} + \omega^{\theta(A_1)} + \cdots + \omega^{\theta(A_m)} \qquad (X = A_0\,0\cdots 0\,A_m),$$

all blocks counted. (The recursion terminates because $A - 1$ has a smaller largest letter than A. The leading 0 in $0X$ is the base letter of the list $[1,0] + X$.)

Lemma 4.0 (basic facts). For all worms A, W, X, Y:

- (i) $o(W0) = o(W) + 1$ and $\mathrm{ord}(X0) = \mathrm{ord}(X) + 1$; hence $\mathrm{ord}(X0^z) = \mathrm{ord}(X) + z$.

- (ii) $\mathrm{ord}(A0Y) = \mathrm{ord}(A) + \mathrm{ord}(Y)$, and $o(A0Y) = o(A) + \mathrm{ord}(Y)$.
- (iii) $o(W) \le \mathrm{ord}(W)$, with equality if W starts with a nonzero letter.
- (iv) If Y is zero-free and nonempty, then $\mathrm{ord}(Y) = o(Y) = \omega^{\theta(Y)}$ with $\theta(Y) \ge 1$; in particular $\mathrm{ord}(Y) \ge \omega$ is a limit and is closed under addition (a sum of finitely many ordinals $< \omega^{\theta(Y)}$ is $< \omega^{\theta(Y)}$).
- (v) $o((k)) = \tau_{k+1}$ and $\mathrm{ord}((k)) = \tau_{k+1}$ for $k \ge 1$. If the largest letter of W is $\le h$, then $o(W) < \tau_{h+2}$ and $\mathrm{ord}(W) < \tau_{h+2}$. So all these ordinals are $< \varepsilon_0$.

Proof. (i) Appending 0 adds one block, the empty block after the new zero, which is counted and contributes $\omega^0 = 1$. (ii) The blocks of $A0Y$ are those of A followed by those of Y; the first block of Y now follows a zero and is counted. (iii) $\mathrm{ord}(W)$ also counts an empty first block, contributing a leading 1. (iv) Y is a single block. $Y - 1$ is nonempty, and the value o of a nonempty worm is ≥ 1 (it has a counted block: the block after a zero, or a nonempty first block). Every ordinal ω^β is closed under finite sums of smaller ordinals ((F2)). (v) $o((0)) = \omega^{\theta(\varnothing)} = 1 = \tau_1$ (one empty block after the zero), and $o((k)) = \omega^{\theta((k))} = \omega^{o((k-1))} = \omega^{\tau_k} = \tau_{k+1}$ by induction; $\mathrm{ord}((k)) = o((0,k)) = \omega^{\theta((k))}$ likewise. For the bound, induct on h: if all letters are 0, $o(W)$ and $\mathrm{ord}(W)$ are finite, $< \tau_2 = \omega$. For $h \ge 1$, every block A has $\theta(A) = o(A-1) < \tau_{h+1}$ by induction, so every term is $< \omega^{\tau_{h+1}} = \tau_{h+2}$, and τ_{h+2} is closed under finite sums. □

4.2 Three monotonicity facts

Lemma 4.1 (appending raises o). For every worm W and letter $a \ge 0$: $o(W) < o(Wa)$ and $\mathrm{ord}(W) < \mathrm{ord}(Wa)$. Hence o and ord strictly increase along prefixes, and $\theta(A') < \theta(A)$ if A' is a proper prefix of a zero-free A.

Proof. Induction on a. For $a = 0$ use Lemma 4.0(i). Let $a \ge 1$. If $W = \varnothing$: $o(\varnothing) = 0 < o((a))$. Otherwise let D be the last block of W; it is counted (it follows a zero, or it is a nonempty first block). Write $o(W) = \sigma + \omega^{\theta(D)}$; then $o(Wa) = \sigma + \omega^{\theta(Da)}$, and $\theta(Da) = o((D-1)(a-1)) > o(D-1) = \theta(D)$ by the induction hypothesis for the letter $a-1$ (applied to the worm $D-1$). Left addition is strictly increasing. For ord apply this to $0W$. The statement about θ is the same computation. □

Lemma 4.2 (lowering the last letter lowers o). For every worm W and letter $a \ge 1$: $o(W(a-1)) < o(Wa)$ and $\mathrm{ord}(W(a-1)) < \mathrm{ord}(Wa)$. Hence $o(Wu) < o(Wa)$ for all $u < a$; in particular $o(Y^-) < o(Y)$ and $\mathrm{ord}(Y^-) < \mathrm{ord}(Y)$ for stripped Y.

Proof. Induction on a. If $W = \varnothing$: $o((a-1)) < o((a))$, since $o((0)) = 1 < \omega = o((1))$ and, for $a \ge 2$, $\theta((a-1)) = o((a-2)) < o((a-1)) = \theta((a))$ by induction. Otherwise let D be the last (counted) block of W and $o(W) = \sigma + \omega^{\theta(D)}$. For $a = 1$: $o(W0) = \sigma + \omega^{\theta(D)} + 1$, while $o(W1) = \sigma + \omega^{\theta(D1)}$ with $\theta(D1) = o((D-1)0) = \theta(D) + 1$ by Lemma 4.0(i). Since $\omega^{\theta(D)} + 1 < \omega^{\theta(D)} \cdot 2 \le \omega^{\theta(D)+1}$, we get $o(W0) < o(W1)$. For $a \ge 2$: $\theta(D(a-1)) = o((D-1)(a-2)) < o((D-1)(a-1)) = \theta(Da)$ by induction. For ord apply this to $0W$. □

Lemma 4.3. Let A' be zero-free, and let A'' be a prefix of A', possibly with its last letter lowered (but still ≥ 1). Then $\theta(A'') \le \theta(A')$.

Proof. $A'' - 1$ is a prefix of $A' - 1$, possibly with its last letter lowered. Apply Lemmas 4.2 and 4.1 to $o(A'' - 1) \le o(A' - 1)$. □

4.3 Every step lowers the ordinal

Lemma 4.4. For every stripped Y and every n: $\mathrm{ord}(C_n(Y)) < \mathrm{ord}(Y)$ and $o(C_n(Y)) < o(Y)$.

Proof. Induction on the largest letter h of Y, for all stripped Y at once.

Reduction to the last block. If Y contains a zero, write $Y = P0A$ with A its last block. A is zero-free and nonempty, since Y is stripped. By Lemma 3.2, $C_n(Y) = P0C_n(A)$. By Lemma 4.0(ii), $\mathrm{ord}(C_n(Y)) = \mathrm{ord}(P) + \mathrm{ord}(C_n(A))$ and $\mathrm{ord}(Y) = \mathrm{ord}(P) + \mathrm{ord}(A)$, and the same with $o(P)$ in place of $\mathrm{ord}(P)$. Left addition is strictly increasing, so it suffices to prove $\mathrm{ord}(C_n(A)) < \mathrm{ord}(A)$ for zero-free A (whose largest letter is $\le h$).

Zero-free $A = (a_1, \dots, a_p)$, $c = a_p$. By Lemma 4.0(iii),(iv) it suffices to show $\mathrm{ord}(C) < \omega^{\theta(A)}$ for $C = C_n(A)$. Since $\mathrm{ord}(C)$ is the sum of $\omega^{\theta(D)}$ over all blocks D of C, and $\omega^{\theta(A)}$ is closed under finite sums, it suffices that **every block D of C has $\theta(D) < \theta(A)$.**

Case $c \geq 2$. All letters of $e(A)$ are ≥ 1 (those of A^-, and those of B, which are letters of A or $c-1$). Let $C = (e_1, \dots, e_i, v)$. If $v \geq 1$, C is a single block and, by Lemma 3.4, $C-1$ is a candidate of $A-1$, a stripped worm with largest letter $\leq h-1$. By the induction hypothesis $\theta(C) = o(C-1) < o(A-1) = \theta(A)$. If $v = 0$ (an extra), the blocks are $(e_1, \dots, e_i)$ and $\varnothing$. First, $\theta(\varnothing) = 0 < \theta(A)$. Next, $\theta((e_1, \dots, e_i)) = o((e_1-1, \dots, e_i-1))$. For $i = p$ this is $o((A-1)^-) < o(A-1)$, by Lemma 4.2. For $i > p$, $(e_1 - 1, \dots, e_i - 1)$ is the last member of group $i-1$ of $A-1$ (Lemma 3.4: $e(A-1) = e(A) - 1$), so the induction hypothesis applies.

Case $c = 1$. Write $A = A'1$ with A' zero-free. By Lemma 3.3, $C = A'0T$ with $T = T_n(A')$, a prefix of $(A'0)(A'0)\cdots$ with its last letter possibly lowered. So every block of C is a prefix of A', possibly with its last letter lowered: the first block A', full copies of A', a truncated copy, possibly an empty block. By Lemma 4.3 each has $\theta \leq \theta(A')$. And $\theta(A) = \theta(A'1) = o((A'-1)0) = \theta(A') + 1 > \theta(A')$ by Lemma 4.0(i). □

4.4 Theorem 1, and the battle function Φ

Theorem 1. From every list $[1,0] + X$, the run of R reaches a beep after finitely many steps. Hence every battle ends, R beeps infinitely often, and by (Sim2) of §2.3 M_{51} enters state 50 infinitely often from either start state: M_{51} **does not quasihalt.**

Proof. Assign to the list $[1,0] + X$ the ordinal $\mathrm{ord}(X) < \varepsilon_0$. A non-beep step goes from $[1,0] + Y0^z$ (Y stripped) to $[1,0] + C$ with $C = C_{2z+1}(Y)$, and $\mathrm{ord}(C) < \mathrm{ord}(Y) \leq \mathrm{ord}(Y) + z = \mathrm{ord}(Y0^z)$ by Lemmas 4.4 and 4.0(i). Every strictly decreasing sequence of ordinals is finite. So a run without beeps is finite, that is, every run reaches a list $[1,0] + 0^m$, which beeps. □

Definition 4.5. For a worm X and $z \geq 0$, let $\Phi_X(z)$ be the number m of zeros of the first list of the form $[1,0] + 0^m$ in the run from $\langle X, z\rangle$ (it exists by Theorem 1; if $X0^z$ is all zeros, $\Phi_X(z) = |X| + z$). Thus $\Psi(k) = \Phi_{(k)}(0)$.

Corollary 4.6 (composition law). For all worms A, X, Y, all stripped Y', and all z:

$$\Phi_\varnothing(z) = z, \qquad \Phi_{X0}(z) = \Phi_X(z+1), \qquad \Phi_{Y'}(z) = \Phi_{C_{2z+1}(Y')}(0), \qquad \Phi_{A0Y}(z) = \Phi_A\big(1 + \Phi_Y(z)\big).$$

Proof. The first two hold by definition, since $\langle X0, z\rangle = \langle X, z+1\rangle$. The third holds because the run from $\langle Y', z\rangle$ goes to $[1,0] + C_{2z+1}(Y') = \langle C_{2z+1}(Y'), 0\rangle$ in one non-beep step. For the fourth: if $Y0^z$ is all zeros, both sides are $\Phi_A(|Y| + z + 1)$. Otherwise let $[1,0] + W_0, \dots, [1,0] + W_T$ be the run from $\langle Y, z\rangle$ up to the first all-zero list $[1,0] + W_T = [1,0] + 0^m$, $m = \Phi_Y(z)$. By Lemma 3.2 and induction on t, the run from $\langle A0Y, z\rangle$ is $[1,0] + A + [0] + W_t$ for $t \leq T$. None of these lists is all zero for $t < T$ (each contains the nonzero letters of W_t), and the last one is $\langle A, m+1\rangle$. □

So the first steps of every battle are: $\langle (k), 0\rangle \to \langle (k-1, 1), 0\rangle \to \langle (k-1, 0, 1), 0\rangle$ for $k \geq 2$ (by Lemma 3.1(d) and Lemma 3.3), and $\Phi_{(1)}(0) = 3$ (Lemma 3.3: $C_1((1)) = 0^3$). Hence

$$\Psi(k) = \Phi_{(k-1)}(4) \qquad (k \geq 2), \tag{4.1}$$

and more generally $\Phi_{Pc}(0) = \Phi_{P(c-1)}(4)$ for zero-free P and $c \geq 2$, and $\Phi_{P1}(0) = \Phi_P(4)$ for zero-free nonempty P. (For Pc: $C_1(Pc) = (P, c-1, 1)$ by Lemma 3.1(d), since $e_{p+1} = B_1 \geq 1$; then $C_1(P(c-1)1) = (P, c-1, 0, 1)$ by Lemma 3.3; then apply Corollary 4.6.)

---

5. Counters are never lost

The counter is reset at every step, but it is rebuilt from the shape of the candidate. This section makes that exact.

Lemma 5.1 (size bound). For every stripped Y and every z: $\Phi_Y(z) \geq \text{size}(Y) + 2z + 1$. For every worm X: $\Phi_X(z) \geq \text{size}(X) + z$.

Proof. Induction on the length of the run from $\langle Y, z\rangle$ to its first all-zero list (finite by Theorem 1). The first step goes to $[1,0] + C$ with $C = C_{2z+1}(Y)$, and $\text{size}(C) = \text{size}(Y) + 2z + 1$ by Lemma 3.1(c). If C is all zeros, $\Phi_Y(z) = |C| = \text{size}(C)$. Otherwise $C = Y'0^{z'}$ with Y' stripped, the run from $\langle Y', z'\rangle$ is one step shorter, and $\Phi_Y(z) = \Phi_{Y'}(z') \geq \text{size}(Y') + 2z' + 1 > \text{size}(C)$. For a general X write $X = Y0^j$ (Y stripped) and use $\Phi_X(z) = \Phi_Y(z + j)$, or $X = 0^j$ and $\Phi_X(z) = j + z$. □

Lemma 5.2 (counter lemma). For all worms X, P, every stripped Y, every letter $v \geq 0$ and all $n, y, z \geq 0$:

- **(M)** $\Phi_X(z) < \Phi_X(z+1)$;
- **(D) (doubling)** $\Phi_{P(v+1)}(y) \geq \Phi_{Pv}(2y+2)$;
- **(I) (index for counter)** $\Phi_{C_{n+1}(Y)}(y) \geq \Phi_{C_n(Y)}(y+1)$.

Proof. To an instance attach the worm U and the number u given by: $U = X$, $u = z+1$ for (M); $U = P(v+1)$, $u = y$ for (D); $U = C_{n+1}(Y)$, $u = y$ for (I). Its **key** is the pair $(\text{ord}(U) + u,\ |U|)$, ordered lexicographically. This is a well-order. We prove all instances by induction on the key. As the proofs below show, an instance of (M) or (D) uses only instances with smaller keys, and an instance of (I) uses one instance of (D) with the same key and instances of (M) with smaller keys. So at each key, first (M) and (D) are proved, then (I). Two facts are used repeatedly: if λ is a limit and $\beta < \lambda$, then $\beta + x < \lambda$ for every finite x; and $\text{ord}(Q) \geq \omega$ is a limit for zero-free nonempty Q (Lemma 4.0(iv)).

(I) from (D). By Lemma 3.1(b), either $C_{n+1} = C_n 0$, and then $\Phi_{C_{n+1}}(y) = \Phi_{C_n}(y+1)$; or $C_n = Qv$ and $C_{n+1} = Q(v+1)$. In the second case (D) for (Q, v, y) has the same key and gives $\Phi_{C_{n+1}}(y) \geq \Phi_{C_n}(2y+2)$. Then (M) for C_n at the counters $y+1, \ldots, 2y+1$ gives $\geq \Phi_{C_n}(y+1)$. These keys have first component $\text{ord}(Qv) + x < \text{ord}(Q(v+1))$: by Lemma 4.2, $\text{ord}(Qv) < \text{ord}(Q(v+1))$, and the latter is a limit (its last block is nonempty).

(M) for (X, z). If $X = \varnothing$: $z < z+1$. If $X = X'0$: this is (M) for $(X', z+1)$, whose key has the same first component ($\text{ord}(X'0) = \text{ord}(X')+1$) and a shorter worm. If X is stripped and contains a zero, write $X = A0Q$ with Q its last block. By Corollary 4.6, $\Phi_X(z) = \Phi_A(1 + \Phi_Q(z))$ and $\Phi_X(z+1) = \Phi_A(1 + \Phi_Q(z+1))$. Apply (M) for (Q, z): its key has first component $\text{ord}(Q) + z + 1 \leq \text{ord}(X) + z + 1$ and a shorter worm. Then apply (M) for A at the counters in between: their keys have first component $\text{ord}(A) + x < \text{ord}(A) + \text{ord}(Q) = \text{ord}(X)$. If X is zero-free and nonempty, write $C_m = C_m(X)$. Then

$$\Phi_X(z+1) = \Phi_{C_{2z+3}}(0) \ \geq\ \Phi_{C_{2z+2}}(1) \ \geq\ \Phi_{C_{2z+1}}(2) \ >\ \Phi_{C_{2z+1}}(0) = \Phi_X(z),$$

by (I) twice and (M) twice. All these keys have first component $\text{ord}(C_m) + x < \text{ord}(X)$, by Lemma 4.4.

(D) for (P, v, y). If P contains a zero, write $P = A0P'$ with P' zero-free (possibly empty). By Corollary 4.6, $\Phi_{P(v+1)}(y) = \Phi_A(1 + \Phi_{P'(v+1)}(y))$ and $\Phi_{Pv}(2y+2) = \Phi_A(1 + \Phi_{P'v}(2y+2))$. Apply (D) for (P', v, y), whose key has first component $\leq \text{ord}(P(v+1)) + y$ and a shorter worm. Then apply (M) for A, at keys $< \text{ord}(A) + \text{ord}(P'(v+1)) = \text{ord}(P(v+1))$. Now let P be zero-free.

- $v = 0$, $P = \varnothing$: $\Phi_{(1)}(y) = 2y + 3 = \Phi_{(0)}(2y+2)$, since $C_{2y+1}((1)) = 0^{2y+3}$.
- $v = 0$, $P \neq \varnothing$: by Lemma 3.3, $C_{2y+1}(P1) = P0T$ with $T = T_{2y+1}(P)$ and $\text{size}(T) = 2y+2$. By Corollary 4.6 and Lemma 5.1, $\Phi_{P1}(y) = \Phi_P(1 + \Phi_T(0))$ with $\Phi_T(0) \geq 2y+2$. By (M) for P, this is $\geq \Phi_P(2y+3) = \Phi_{P0}(2y+2)$. The keys used have first component $\text{ord}(P) + x < \omega^{\theta(P)} \cdot 2 \leq \omega^{\theta(P)+1} = \text{ord}(P1)$.
- $v \geq 1$: $P(v+1)$ is zero-free with last letter ≥ 2. By Corollary 4.6 and (I) applied $2y+1$ times (keys $< \text{ord}(P(v+1))$ by Lemma 4.4),

$$\Phi_{P(v+1)}(y) = \Phi_{C_{2y+1}}(0) \geq \Phi_{C_{2y}}(1) \geq \cdots \geq \Phi_{C_0}(2y+1),$$

and $C_0 = C_0(P(v+1)) = (P, v, 0)$ by Lemma 3.1(d). So this is $\Phi_{Pv}(2y+2)$. □

Corollary 5.3.

- (a) *(Choosing a candidate.)* For stripped Y and $0 \le n \le 2z+1$: $\Phi_Y(z) \ge \Phi_{C_n(Y)}(2z+1-n)$. In particular, with $Y[j]$ and $W(Y)$ as in Lemma 3.1(e), $\Phi_Y(z) \ge \Phi_{Y[j]}(2z+2-jW(Y))$ whenever $1 \le j$ and $jW(Y) \le 2z+2$.
- (b) *(Appending.)* $\Phi_{Xa}(z) \ge \Phi_X(z+1) > \Phi_X(z)$ for every worm X and letter a.
- (c) *(Iterated doubling.)* $\Phi_{P(v+m)}(y) \ge \Phi_{Pv}(2^m(y+2)-2)$.
- (d) Ψ is strictly increasing, $\Psi(k) \ge 2^{k+1}-1$, $k_{j+1} \ge 2^{k_j+1}$, and $k_j \ge j$ for all $j \ge 1$.

Proof. (a) Corollary 4.6 and (I), $2z+1-n$ times; for the second form take $n = jW(Y)-1$ (Lemma 3.1(e)). (b) $\Phi_{X0}(z) = \Phi_X(z+1)$, and for $a \ge 1$ use (D) repeatedly and (M). (c) m applications of (D), with $y \mapsto 2y+2$ iterated m times. (d) $\Psi(k+1) = \Phi_{(k+1)}(0) \ge \Phi_{(k)}(2) > \Phi_{(k)}(0) = \Psi(k)$ by (D) and (M). By (c) with $P = \varnothing$, $v = 0$, $y = 0$: $\Psi(k) \ge \Phi_{(0)}(2^{k+1}-2) = 2^{k+1}-1$. Then $k_{j+1} = 1 + \Psi(k_j) \ge 2^{k_j+1} > k_j$, and $k_1 = 1$. □

These bounds are exact facts about R, with complete proofs. They are far too weak for PA-unprovability: 2^{k+1} is elementary. The next section proves the growth that PA-unprovability needs.

---

6. Growth: LEMMA G

Conventions are those of §§1–5 and of Appendix A: $\alpha[n]$ is the Buchholz–Wainer fundamental sequence (Definition A.1), H_α and F_α are the Hardy and fast-growing functions (§A.3), $N(\cdot)$ is the norm of (F5), and $\tau_0 = 0$, $\tau_{s+1} = \omega^{\tau_s}$. This section proves LEMMA G, the growth lemma used in §§7–8.

Main results.

- **Theorem 6.20.** For all $k \ge 1$ and $z \ge 0$ with $2z+k+1 \ge 8(k+1)^2 + 1000$:

$$\Phi_{(k)}(z) \;\ge\; \tfrac{1}{2}\, H_{\tau_{k+1}}\Big(\Big\lfloor \frac{2z+k+1}{8(k+1)} \Big\rfloor\Big).$$

  So the battle of a single letter k with a large counter is at least the Hardy function of its own ordinal $\mathrm{ord}((k)) = \tau_{k+1}$, with the argument divided by $4(k+1)$. No tower level is lost.
- **Corollary 6.21 (LEMMA G*).** For every $m \ge 1$, with $j = \max(m, 3)$: $\Phi_{(j)}(z) \ge H_{\tau_m}(z)$ for all large z.
- **Corollary 6.22 (LEMMA G).** For every $\beta < \varepsilon_0$: $\Psi(k) \ge H_\beta(k+1)$ for all large k. Moreover $\Psi(k) \ge H_{\tau_{\lfloor k/2 \rfloor}}(k)$ for all large k.

How the proof goes.

1. *Normal forms (§§6.2–6.5).* A worm is in normal form (NF) if no term of its ordinal is absorbed, at any level. Every ordinal below ε_0 has exactly one NF worm, and no worm with that ordinal is smaller (Lemma 6.4). On NF worms the ordinal order is the lexicographic order (Lemma 6.7). Every battle stays inside NF (Lemma 6.9). On NF worms, the candidate $C_n(Y)$ is the **size-greedy predecessor** of Y: among all ordinals below $o(Y)$ that have a worm of size at most $\mathrm{size}(Y)+n$, it has the largest one (Proposition 6.12).
2. *Battles are optimal plays (§6.6).* Call a strictly decreasing sequence of ordinals a *play* if its i-th term has a worm of size at most (start size)$+i$. The total size of the list of R grows by at least one per step, and by Proposition 6.12 the battle always takes the largest ordinal allowed. So no play outlasts the battle (Theorem 6.14). This is a comparison principle between the battle and arbitrary descents: to bound Φ from below, it is enough to exhibit **one** long play, and plays may use Buchholz–Wainer fundamental sequences freely.
3. *A Hardy bound for plays (§§6.7–6.8).* From (α, t) with $t \ge 8\|\alpha\|^2 + 1000$, where $\|\alpha\|$ is the size of the NF worm of α, some play from (α, t) ends at a time $\ge H_\alpha(\lfloor t/(8\|\alpha\|)\rfloor)$ (Theorem 6.19). The proof is a transfinite induction. At a limit λ the play moves to $\lambda[y] + \omega^2$ with $y = \lfloor t/(8\|\lambda\|)\rfloor$. It burns the summand ω^2 first, which raises the available time exponentially, and then continues from $\lambda[y]$ with ample time.

The key new idea, in brief. The counter of R is reset at every step, but the *total size* of the list is never reset: it grows by $z+1$ at a step with counter z, and the next candidate is the largest ordinal that fits in

that size. So a battle is a "budgeted" ordinal descent, with budget equal to elapsed time, that is greedy. The greedy descent is the longest descent under that budget. A slowed-down Hardy descent fits under the same budget once a cheap summand ω^2 is spent at each limit step to regenerate time.

6.1 The statement, and two reductions

LEMMA G. For every $\beta < \varepsilon_0$ there is K such that $\Psi(k) \geq H_\beta(k+1)$ for all $k \geq K$.

(Here H_β is the Buchholz–Wainer Hardy function, §A.3.) This is the only ingredient of Theorems 2 and 3 (§§7–8) beyond §§2–5. It reduces to a statement about the battle functions $z \mapsto \Phi_{(j)}(z)$ of the one-letter worms (j) with a *large* counter z:

LEMMA G*. For every $m \geq 1$ there is $j \geq 1$ such that $\Phi_{(j)}(z) \geq H_{\tau_m}(z)$ for all sufficiently large z.

Proposition 6.1. LEMMA G* implies LEMMA G.

Proof. Let $\beta < \varepsilon_0$. By (F6) there is m with $\beta < \tau_m$. Let j be given by LEMMA G* for $m+1$, and z_0 such that $\Phi_{(j)}(z) \geq H_{\tau_{m+1}}(z)$ for $z \geq z_0$. For $k > j$, Corollary 5.3(c) with $P = \varnothing$, $v = j$, $y = 0$ and $k-j$ doublings (the parameter called m there) gives

$$\Psi(k) = \Phi_{(k)}(0) \;\geq\; \Phi_{(j)}(Z_k), \qquad Z_k := 2^{k-j+1} - 2.$$

Take K so large that for $k \geq K$: $Z_k \geq z_0$, $Z_k \geq k+1$ and $Z_k \geq N(\beta)$. Then for $k \geq K$: $\Phi_{(j)}(Z_k) \geq H_{\tau_{m+1}}(Z_k)$; next $H_{\tau_{m+1}}(Z_k) \geq H_\beta(Z_k)$ by Corollary A.11, since $\beta < \tau_m < \tau_{m+1}$ and $N(\beta) \leq Z_k$; and $H_\beta(Z_k) \geq H_\beta(k+1)$ by Lemma A.6. □

Proposition 6.2. Let $2 \uparrow\uparrow 0 = 1$, $2 \uparrow\uparrow (i+1) = 2^{2 \uparrow\uparrow i}$ and $\log^* x = \min\{i : 2 \uparrow\uparrow i \geq x\}$. Suppose $\Psi(k) \geq H_{\tau_{f(k)}}(2)$ for all large k, where $f(k) \geq \log^*(k+1) + 3$. Then LEMMA G holds.

Proof. Put $a_s = H_{\tau_s}(2)$. For $s \geq 2$, $\tau_s \cdot 2 = \tau_s + \tau_s$ is a normal sum of norm 2 and $\tau_s \cdot 2 < \tau_{s+1}$. So Corollary A.11 and Lemma A.8(a) give $a_{s+1} \geq H_{\tau_s \cdot 2}(2) = H_{\tau_s}(a_s)$. Next, $a_2 = H_\omega(2) = 5$ and $a_3 \geq H_\omega(5) = 11$. For $s \geq 3$ we have $\omega^2 < \tau_s$ and $N(\omega^2) = 2 \leq a_s$, so $a_{s+1} \geq H_{\tau_s}(a_s) \geq H_{\omega^2}(a_s) = 2^{a_s+1}(a_s+1) - 1 \geq 2^{a_s}$ (Corollary A.11 again, and Lemma A.8(b) with $F_2(x) = 2^{x+1}(x+1) - 1$). Hence $a_s \geq 2 \uparrow\uparrow (s-1)$ for $s \geq 3$. Let $\beta < \tau_m$. For k so large that $f(k) - 1 \geq \max(m, 3)$ and $N(\beta) \leq k+1$:

$$\Psi(k) \geq a_{f(k)} \geq H_{\tau_{f(k)-1}}(a_{f(k)-1}) \geq H_{\tau_{f(k)-1}}(k+1) \geq H_\beta(k+1).$$

The third inequality uses $a_{f(k)-1} \geq 2 \uparrow\uparrow (f(k)-2) \geq k+1$ and Lemma A.6. The last uses Corollary A.11 with $\beta < \tau_m \leq \tau_{f(k)-1}$. □

LEMMA G* is proved in Corollary 6.21, and the hypothesis of Proposition 6.2 (with $f(k) = \lfloor k/2 \rfloor$) in Corollary 6.22. Either route gives LEMMA G.

6.2 Normal-form worms

Throughout, worms, blocks, θ, o and ord are as in §4.1. Every worm is uniquely $W = A_0\,0\,A_1\,0 \cdots 0\,A_m$ with zero-free (possibly empty) blocks A_i; $o(W) = [A_0 \neq \varnothing]\,\omega^{\theta(A_0)} + \omega^{\theta(A_1)} + \cdots + \omega^{\theta(A_m)}$ (left-to-right ordinal sum), and $\theta(A) = o(A-1)$ for nonempty zero-free A, $\theta(\varnothing) = 0$. From Lemma 4.0 we use: (i) $o(W0) = o(W) + 1$; (ii) $o(A0Y) = o(A) + \mathrm{ord}(Y)$; (iv) for zero-free nonempty A, $o(A) = \mathrm{ord}(A) = \omega^{\theta(A)}$ with $\theta(A) \geq 1$. Recall $\mathrm{ord}(U) = \sum_{\text{blocks } B \text{ of } U} \omega^{\theta(B)}$, all blocks counted. From Lemmas 4.1, 4.2 and 4.4 we use: appending a letter strictly raises o; lowering the last letter strictly lowers o; $o(C_n(Y)) < o(Y)$ for stripped Y.

In an ordinal sum $\omega^{e_1} + \cdots + \omega^{e_s}$ a term ω^{e_i} is **absorbed** if $e_i < e_l$ for some $l > i$. The non-absorbed terms, in order, form the Cantor normal form (CNF) of the sum. (Proof: if some term is absorbed, let i be the largest absorbed index. The terms after i are not absorbed, so their exponents are non-increasing, and $e_i < e_l \leq e_{i+1}$ for the $l > i$ that absorbs i. The standard identity $\omega^a + \omega^b = \omega^b$ for $a < b$ removes term i without changing the sum. A term $j < i$ that was absorbed stays absorbed, because if i absorbed j then so does l. Repeat until no term is absorbed. The exponents are then non-increasing, which is the CNF by (F1).)

Norms. For $j \geq 0$ and a worm $X = (x_1, \ldots, x_p)$ put $\nu_j(X) := \sum_{i=1}^{p}(x_i + 1 + j)$. So $\nu_0 =$ size, a zero costs $1 + j$, and for zero-free X

$$\nu_j(X) = \nu_{j+1}(X - 1). \tag{6.1}$$

Definition 6.3 (normal form). By recursion on the largest letter, a worm W is in **normal form (NF)** if either $W = 0^r$ for some $r \geq 0$, or

$$W = A_0\, 0\, A_1\, 0 \cdots 0\, A_q\, 0^r \qquad (q, r \geq 0)$$

with $A_0, \ldots, A_q$ nonempty and zero-free, $\theta(A_0) \geq \theta(A_1) \geq \cdots \geq \theta(A_q)$, and every $A_i - 1$ in NF.

So the blocks of an NF worm are nonempty, except for a final run of empty blocks. They are non-increasing in θ, and the same holds one level down inside every block. An NF worm begins with 0 only if it is 0^r. For W as displayed,

$$o(W) = \omega^{\theta(A_0)} + \cdots + \omega^{\theta(A_q)} + r, \qquad \theta(A_0) \geq \cdots \geq \theta(A_q) \geq 1, \tag{6.2}$$

and this *is* the CNF of $o(W)$: no term is absorbed.

Lemma 6.4 (canonical worms).

- (a) For every $\beta < \varepsilon_0$ there is exactly one NF worm $[\![\beta]\!]$ with $o([\![\beta]\!]) = \beta$. Explicitly, $[\![r]\!] = 0^r$ for finite r, and if $\beta = \omega^{\beta_0} + \cdots + \omega^{\beta_q} + r$ is in CNF with $q \geq 0$ and $\beta_q \geq 1$,

$$[\![\beta]\!] = \big([\![\beta_0]\!] + 1\big)\, 0\, \big([\![\beta_1]\!] + 1\big)\, 0 \cdots 0\, \big([\![\beta_q]\!] + 1\big)\, 0^r.$$

- (b) Put $m_j(\beta) := \nu_j([\![\beta]\!])$. Then $m_j(r) = r(1 + j)$ for finite r, and for β as in (a) with $q \geq 0$

$$m_j(\beta) = \sum_{i=0}^{q} m_{j+1}(\beta_i) + (q + r)(1 + j). \tag{6.3}$$

- (c) (*Minimality*) For every worm W and every $j \geq 0$: $\nu_j(W) \geq m_j(o(W))$.

Proof. (a) *Existence*, by recursion on β. Each $\beta_i < \beta$, because $\beta_i < \omega^{\beta_i} \leq \beta$ below ε_0. Each $[\![\beta_i]\!]$ is nonempty ($\beta_i \geq 1$), so $[\![\beta_i]\!] + 1$ is a nonempty zero-free block with $\theta([\![\beta_i]\!] + 1) = o([\![\beta_i]\!]) = \beta_i$, and $([\![\beta_i]\!] + 1) - 1 = [\![\beta_i]\!]$ is NF. The θ's are the non-increasing β_i, so the displayed worm is NF, and by (6.2) its o is β. *Uniqueness*, by induction on the largest letter. If W is NF with $o(W) = \beta$, then (6.2) is the CNF of β, so q, r and $\theta(A_i) = \beta_i$ are determined by β. Each $A_i - 1$ is NF with $o(A_i - 1) = \beta_i$ and has a smaller largest letter, so $A_i - 1 = [\![\beta_i]\!]$ by induction, and $W = [\![\beta]\!]$. If $\beta = r$ is finite, W has no nonempty block, because by (6.2) a nonempty block would contribute a CNF term ω^θ with $\theta \geq 1$. So $W = 0^r = [\![r]\!]$.

(b) $[\![\beta]\!]$ consists of the blocks $[\![\beta_i]\!] + 1$, with $\nu_j([\![\beta_i]\!] + 1) = \nu_{j+1}([\![\beta_i]\!]) = m_{j+1}(\beta_i)$ by (6.1), and of $q + r$ zeros, each costing $1 + j$.

(c) Induction on the largest letter of W, for all j at once. If $W = 0^m$ then $o(W) = m$ and $\nu_j(W) = m(1 + j) = m_j(m)$. Otherwise write $W = A_0\, 0 \cdots 0\, A_m$ (m zeros, at least one nonempty block), so $\nu_j(W) = \sum_i \nu_j(A_i) + m(1 + j)$. For a nonempty block, (6.1) and the induction hypothesis for $A_i - 1$ (smaller largest letter) give $\nu_j(A_i) = \nu_{j+1}(A_i - 1) \geq m_{j+1}(o(A_i - 1)) = m_{j+1}(\theta(A_i))$. Let $\beta := o(W)$. It is the sum of the terms $\omega^{\theta(A_i)}$ of the counted blocks ($i \geq 1$, and $i = 0$ if $A_0 \neq \varnothing$). A nonempty block has $\theta \geq 1$ (Lemma 4.0(iv)), and the last nonempty block is counted and its term is not absorbed, so $\beta \geq \omega$. Write its CNF as $\omega^{\beta_0} + \cdots + \omega^{\beta_q} + r$ with $q \geq 0$ and $\beta_q \geq 1$. It consists of the non-absorbed terms. The terms ω^{β_l} ($\beta_l \geq 1$) are the terms of a set I of $q + 1$ nonempty counted blocks. The final r ones are the terms of the r blocks after the last nonempty block; these blocks are empty, have index ≥ 1, and lie outside I. If $0 \in I$, the other $q + r$ of these $q + 1 + r$ blocks have indices in $\{1, \ldots, m\}$; if $0 \notin I$, all $q + 1 + r$ do. Either way $m \geq q + r$. Dropping the blocks outside I (each has $\nu_j \geq 0$), $\nu_j(W) \geq \sum_{i \in I} m_{j+1}(\theta(A_i)) + (q + r)(1 + j) = m_j(\beta)$ by (6.3). □

The worm norm. For $\beta < \varepsilon_0$ put

$$\|\beta\| := m_0(\beta) = \text{size}([\![\beta]\!]) = \min\{\text{size}(X) : X \text{ a worm with } o(X) = \beta\}$$

(the minimum by Lemma 6.4(c)). For an NF worm X, $X = [\![o(X)]\!]$ by uniqueness, so $\mathrm{size}(X) = \|o(X)\|$. There are finitely many ordinals of worm norm $\le s$, since there are finitely many worms of size $\le s$.

Lemma 6.5 (norm facts).

- (i) $\|\beta + r\| = \|\beta\| + r$ for every $\beta < \varepsilon_0$ and finite r; in particular $\|r\| = r$.
- (ii) Let $\rho \ge \omega$ have no finite part, and let $\zeta \ge 1$ have all CNF exponents $\le$ the last exponent of ρ (so the CNF of $\rho + \zeta$ is that of ρ followed by that of ζ). Then $\|\rho + \zeta\| = \|\rho\| + 1 + \|\zeta\|$ if $\zeta \ge \omega$, and $\|\rho + \zeta\| = \|\rho\| + \zeta$ if $\zeta < \omega$.
- (iii) $\|\omega^\gamma\| = m_1(\gamma)$ for $\gamma \ge 1$, and $m_1(k) = 2k$ for finite k; so $\|\omega\| = 2$, $\|\omega^2\| = 4$, $\|\omega \cdot k\| = 3k - 1$ ($k \ge 1$). For ρ as in (ii): $\|\rho + \omega \cdot k\| = \|\rho\| + 3k$, and $\|\rho + \omega^2\| = \|\rho\| + 5$ if the last exponent of ρ is ≥ 2.
- (iv) $[\![\tau_{k+1}]\!] = (k)$, so $\|\tau_{k+1}\| = k + 1$ ($k \ge 0$).
- (v) For every limit $\mu < \varepsilon_0$, every $y \ge 0$ and every $j \ge 0$: $m_j(\mu[y]) \le (y+1)\, m_j(\mu)$. In particular $\|\lambda[y]\| \le (y+1)\|\lambda\|$ for every limit λ.

Proof. (i)–(iii) are read off from (6.3). For (i): if β is finite, $\|\beta + r\| = m_0(\beta + r) = \beta + r$; otherwise $\beta + r$ has the infinite CNF terms of β and r more finite units, and each finite unit costs 1 in (6.3) at $j = 0$. For (ii), if ρ has $q_\rho + 1$ terms and $\zeta = \omega^{\zeta_0} + \cdots + \omega^{\zeta_p} + r'$ with $p \ge 0$, (6.3) gives $\|\rho + \zeta\| = \sum m_1(\rho_i) + \sum m_1(\zeta_i) + (q_\rho + p + 1 + r') = \|\rho\| + 1 + \|\zeta\|$; if $\zeta = r'$ is finite, $\|\rho + r'\| = \sum m_1(\rho_i) + q_\rho + r' = \|\rho\| + r'$. For (iii), $\|\omega^\gamma\| = m_1(\gamma)$ is (6.3) with $q = r = 0$, and $m_1(k) = 2k$, $m_1(1) = 2$, $m_1(2) = 4$; then $\|\omega \cdot k\| = 2k + (k-1)$, and the two sums follow from (ii). For (iv): $[\![\tau_1]\!] = [\![1]\!] = (0)$, and $[\![\tau_{k+1}]\!] = [\![\omega^{\tau_k}]\!] = [\![\tau_k]\!] + 1 = (k-1) + 1 = (k)$ by induction.

(v) Induction on μ, for all j at once. Write $\mu = \sigma + \omega^\eta$ (last CNF term, $\eta \ge 1$), where σ has $s \ge 0$ terms, all with exponents $\ge \eta$, and no finite part. By (6.3), $m_j(\mu) = \Sigma + m_{j+1}(\eta) + s(1+j)$ with $\Sigma := \sum_{\text{terms } \omega^{\sigma_i} \text{ of } \sigma} m_{j+1}(\sigma_i)$.

- $\eta = \theta + 1$ with $\theta \ge 1$: $\mu[y] = \sigma + \omega^\theta \cdot (y+1)$ (Definition A.1). Its CNF is that of σ followed by $y+1$ copies of ω^θ, because $\theta < \eta$ and every exponent of σ is $\ge \eta$. So $m_j(\mu[y]) = \Sigma + (y+1)m_{j+1}(\theta) + (s+y)(1+j)$. Since $\theta + 1$ has the infinite terms of θ and one more finite unit, $m_{j+1}(\eta) = m_{j+1}(\theta) + (j+2)$. Hence $(y+1)m_j(\mu) \ge \Sigma + s(1+j) + (y+1)m_{j+1}(\theta) + (y+1)(j+2) \ge m_j(\mu[y])$, the last step because $(y+1)(j+2) \ge y(1+j)$.
- $\eta = 1$: $\mu[y] = \sigma + (y+1)$. If $s \ge 1$, (6.3) with $q = s-1$ and $r = y+1$ gives $m_j(\mu[y]) = \Sigma + (s+y)(1+j)$, while $m_j(\mu) = \Sigma + (j+2) + s(1+j)$ (as $m_{j+1}(1) = j+2$). So $(y+1)m_j(\mu) \ge \Sigma + s(1+j) + (y+1)(j+2) \ge \Sigma + (s+y)(1+j) = m_j(\mu[y])$. If $s = 0$, then $\mu = \omega$ and $m_j(y+1) = (y+1)(1+j) \le (y+1)(j+2) = (y+1)m_j(\omega)$.
- η a limit: $\mu[y] = \sigma + \omega^{\eta[y]}$. Here $1 \le \eta[y] < \eta$: $\eta[y]$ has a CNF term by Definition A.1, and $\eta[y] < \eta$ by Lemma A.2(a). So the CNF of $\mu[y]$ is that of σ followed by the single term $\omega^{\eta[y]}$, with a positive exponent, and (6.3) applies. Since $\eta < \omega^\eta \le \mu$, the induction hypothesis for η (at level $j+1$) gives $m_j(\mu[y]) = \Sigma + m_{j+1}(\eta[y]) + s(1+j) \le \Sigma + (y+1)m_{j+1}(\eta) + s(1+j) \le (y+1)m_j(\mu)$. □

6.3 On normal forms, the ordinal order is lexicographic

For worms $U \ne V$ write $U <_{\mathrm{lex}} V$ if U is a proper prefix of V, or if $u_k < v_k$ at the first index k with $u_k \ne v_k$. This is a total order on worms.

Lemma 6.6. Let $V = W\,a\,U$ be an NF worm, where W, U are worms and a is a letter. Then $o(V) < o(W(a+1))$.

Proof. Induction on the largest letter of V. Write $W = W_{\mathrm{pre}} D$, where D is the longest zero-free suffix of W (possibly empty) and W_{pre} is empty or ends with 0. Put $\sigma := 0$ if $W_{\mathrm{pre}} = \varnothing$, and $\sigma := o(W')$ if $W_{\mathrm{pre}} = W'0$.

Case $a = 0$. If $W = \varnothing$, then $V = 0U$ is NF and begins with 0, so $V = 0^r$ and $o(V) = r < \omega = o((1))$. Suppose $W \ne \varnothing$. By Lemma 4.0(ii), $o(V) = o(W0U) = o(W) + \mathrm{ord}(U)$. If $D \ne \varnothing$: D is the last block of W, so $o(W) = \sigma + \omega^{\theta(D)}$ (Lemma 4.0(ii),(iv)). The blocks of U are the blocks of V after D, so by NF each of them has $\theta \le \theta(D)$, and $\mathrm{ord}(U) \le \omega^{\theta(D)} \cdot \ell$, where ℓ is the number of blocks of U. Hence $o(V) \le \sigma + \omega^{\theta(D)}(\ell + 1) < \sigma + \omega^{\theta(D)+1} = \sigma + \omega^{\theta(D1)} = o(W1)$, using $\theta(D1) = o((D-1)0) = \theta(D) + 1$

(Lemma 4.0(i)). If $D = \varnothing$, then $W = W'0$ ends with an empty block. By NF all later blocks are empty, so $U = 0^s$, $V = W'0^{s+2}$ and $o(V) = o(W') + s + 2 < o(W') + \omega = o(W'0\,1) = o(W1)$.

Case $a \geq 1$. Write $U = E\,U_{\mathrm{post}}$, where E is the longest zero-free prefix of U and U_{post} is empty or equals $0U'$. Then DaE is a block of V, and $V = W_{\mathrm{pre}}\,(DaE)\,U_{\mathrm{post}}$. By Lemma 4.0(ii),(iv),

$$o(V) = \sigma + \omega^{\theta(DaE)} + \mathrm{ord}(U') \quad (\text{or } \sigma + \omega^{\theta(DaE)} \text{ if } U_{\mathrm{post}} = \varnothing), \qquad o(W(a+1)) = \sigma + \omega^{\theta(D(a+1))}.$$

The blocks of U' are the blocks of V after DaE, so by NF $\mathrm{ord}(U') < \omega^{\theta(DaE)+1}$, and $o(V) < \sigma + \omega^{\theta(DaE)+1}$. So it suffices that $\theta(DaE) < \theta(D(a+1))$, i.e.

$$o\big((D-1)(a-1)(E-1)\big) < o\big((D-1)\,a\big).$$

The worm $(DaE) - 1 = (D-1)(a-1)(E-1)$ is NF, since DaE is a block of the NF worm V, and it has a smaller largest letter than V. The displayed inequality is the lemma for this worm, with $D-1$, $a-1$, $E-1$ in the roles of W, a, U. It holds by the induction hypothesis. The base of the induction (all letters 0) is contained in the case $a = 0$. □

Lemma 6.7 (lexicographic order). For NF worms $U \neq V$: $U <_{\mathrm{lex}} V$ if and only if $o(U) < o(V)$.

Proof. Let $U <_{\mathrm{lex}} V$. If U is a proper prefix of V, then $o(U) < o(V)$ by Lemma 4.1 (appending strictly raises o). Otherwise $U = W\,a\,U'$ and $V = W\,b\,V'$ with $a < b$. Then

$$o(U) < o(W(a+1)) \leq o(Wb) \leq o(WbV') = o(V),$$

by Lemma 6.6, by Lemma 4.2 ($o(Wv)$ is increasing in the last letter v), and by Lemma 4.1. Conversely, if $o(U) < o(V)$, then $U \neq V$ and not $V <_{\mathrm{lex}} U$ (which would give $o(V) < o(U)$), so $U <_{\mathrm{lex}} V$ because the lexicographic order is total. □

6.4 Closure properties, and cofinality of the candidates

Recall (§2.1, Lemma 3.1(e)) that for stripped $Y = (y_1, \ldots, y_p)$ with last letter c, $r = r(Y)$, $B = B(Y)$, the worm $Y[k]$ ($k \geq 0$) is Y^- followed by k copies of B. So $Y[0] = Y^-$, and $Y[k] = C_{kW(Y)-1}(Y)$ for $k \geq 1$. A **truncation** of a worm $U = (u_1, \ldots, u_p)$ is a worm $(u_1, \ldots, u_{i-1}, v)$ with $1 \leq i \leq p$ and $0 \leq v \leq u_i$, or $\varnothing$.

Lemma 6.8 (truncations). Every prefix of an NF worm is NF. If U is NF with last letter $c \geq 1$, then U^- is NF. Hence every truncation of an NF worm is NF.

Proof. Induction on the largest letter. Let U be NF and P a prefix of U. If P ends with 0 (or $P = \varnothing$), then the blocks of P are complete nonempty blocks of U followed by empty blocks, so P is NF. Otherwise $P = A_0\,0 \cdots 0\,A_{k-1}\,0\,A'$, where A' is a nonempty prefix of the block A_k of U, and $A_0, \ldots, A_{k-1}$ are nonempty because they precede the nonempty block A_k. Here $A' - 1$ is a prefix of the NF worm $A_k - 1$, so it is NF by induction, and $\theta(A') = o(A' - 1) \leq o(A_k - 1) = \theta(A_k) \leq \theta(A_{k-1})$ by Lemma 4.1. So P is NF.

Now let U be NF with last letter $c \geq 1$. Then U has no trailing zeros, and its last block is $A_q = A'c$. If $c \geq 2$, then $A_q^- - 1 = (A_q - 1)^-$ is NF by induction (its last letter is $c - 1 \geq 1$), and $\theta(A_q^-) < \theta(A_q)$ by Lemma 4.2, so U^- is NF. If $c = 1$, then $U^- = A_0\,0 \cdots 0\,A_{q-1}\,0\,A'\,0$. If $A' \neq \varnothing$, its blocks are those of U with A_q replaced by A', followed by one empty block. Here $A' - 1$ is a prefix of $A_q - 1 = (A' - 1)0$, hence NF, and $\theta(A') < \theta(A_q)$. If $A' = \varnothing$, the last two blocks of U^- are empty. In both cases U^- is NF. A truncation $(u_1, \ldots, u_{i-1}, v)$ is obtained from the prefix $(u_1, \ldots, u_i)$ by lowering the last letter $u_i - v$ times. □

Lemma 6.9 (expansions, candidates, runs).

- (a) If Y is NF and stripped, then $Y[k]$ is NF for every $k \geq 0$.
- (b) If Y is NF and stripped, then every candidate $C_n(Y)$ is NF.
- (c) If X is NF, then every list $[1,0] + X'$ of the run of R from $[1,0] + X$, up to and including the first all-zero list, has X' NF.
- (d) Every list of the run of R from $S_0 = [1,0]$, before and after every beep, is $[1,0] + X$ with X NF. Also, for all $k \geq 1$ and $z \geq 0$, the worm of $\langle (k), z \rangle$ is $(k)0^z = [\![\tau_{k+1} + z]\!]$, which is NF. So every battle, and every run from a list $\langle (k), z \rangle$, passes through NF worms only.

Proof. (a) Induction on the largest letter of Y. Write $Y = Y_0\,0\,(Dc)$ or $Y = Dc$, where Dc is the last block of Y (zero-free and nonempty, with last letter $c \geq 1$; D possibly empty). Since Y is NF, the blocks of Y_0 are nonempty (they precede the nonempty block Dc), their θ's are non-increasing and $\geq \theta(Dc)$, and each of them minus 1 is NF, as is $(Dc)-1$. By Lemma 3.2 (with $A = Y_0$), $B(Y) = B(Dc)$ and $Y^- = Y_0\,0\,(Dc)^-$, so $Y[k] = Y^-B(Y)^k = Y_0\,0\,(Dc)[k]$ (resp. $Y[k] = (Dc)[k]$), where $(Dc)[k]$ is the expansion of the stripped worm Dc. So it suffices to check that the blocks contributed by $(Dc)[k]$ fit behind those of Y_0.

Case $c = 1$. Every letter of D is $\geq 1 = c$, so $r(D1) = 0$, $B(D1) = D0 = (D1)^-$ and $(D1)[k] = (D0)^{k+1}$. If $D \neq \varnothing$, the blocks of $Y[k]$ are those of Y_0, then $k+1$ copies of D, then one empty block. Here $\theta(D) < \theta(D)+1 = \theta(D1)$ (Lemma 4.0(i)), which is $\leq$ the θ of the last block of Y_0 if Y_0 is present. Also $D-1$ is NF as a prefix of the NF worm $(D1)-1 = (D-1)0$ (Lemma 6.8). If $D = \varnothing$, then $Y[k] = Y_0\,0\,0^{k+1}$ (resp. 0^{k+1}): the blocks of Y_0 followed by empty blocks. Either way $Y[k]$ is NF.

Case $c \geq 2$. By Lemma 3.4 applied to the zero-free worm Dc, $(Dc)-1$ is stripped, $r((Dc)-1) = r(Dc)$ and $B((Dc)-1) = B(Dc)-1$; also $(Dc)^- - 1 = ((Dc)-1)^-$. Every letter of $B(Dc)$ is ≥ 1 (a letter of D or $c-1$), so $(Dc)[k]$ is zero-free and nonempty, and $(Dc)[k]-1 = ((Dc)-1)^-\,(B(Dc)-1)^k = ((Dc)-1)[k]$. The worm $(Dc)-1$ is NF and stripped, and has a smaller largest letter than Y, so $((Dc)-1)[k]$ is NF by the induction hypothesis. Moreover $\theta((Dc)[k]) = o(((Dc)-1)[k]) < o((Dc)-1) = \theta(Dc)$, by Lemma 4.2 for $k = 0$, and by Lemmas 3.1(e) and 4.4 for $k \geq 1$ (then $((Dc)-1)[k]$ is a candidate of $(Dc)-1$). So $Y[k]$ consists of the blocks of Y_0 followed by the single block $(Dc)[k]$, whose θ is below $\theta(Dc)$ and whose decrement is NF. Hence $Y[k]$ is NF.

(b) By §2.1, $C_n(Y) = (e_1, \ldots, e_i, v)$ with $i \geq p = |Y|$ and $v \leq e_{i+1}$, where $e = e(Y) = Y^-BBB\cdots$. For k with $|Y[k]| \geq i+1$, the worm $(e_1, \ldots, e_{i+1})$ is a prefix of $Y[k]$, and $C_n(Y)$ is obtained from it by lowering the last letter to v. So $C_n(Y)$ is a truncation of the NF worm $Y[k]$, hence NF by Lemma 6.8.

(c) Runs. Let $[1,0]+X'$ be a list of the run with X' NF and not all zero. Then $X' = Y'0^z$ with Y' stripped, and by §2.1 the next list is $[1,0]+C_{2z+1}(Y')$, without a beep. Here Y' is NF as a prefix of X' (Lemma 6.8), so $C_{2z+1}(Y')$ is NF by (b). Induction along the run gives (c).

(d) $S_0 = [1,0]+\varnothing$, and $\varnothing = 0^0$ is NF. Let $S_j = [1,0]+X$ with X NF. If $X = 0^m$, step j beeps and $S_{j+1} = [1,0,m+1] = [1,0]+(m+1)$ (§2.1). The one-letter worm $(m+1)$ is $[\![\tau_{m+2}]\!]$ (Lemma 6.5(iv)), hence NF. If X is not all zero, S_{j+1} is NF by the argument of (c). Induction on j gives the first claim. For the second: the CNF of $\tau_{k+1}+z$ is $\omega^{\tau_k}+z$ with $\tau_k \geq 1$, so Lemma 6.4(a) and $[\![\tau_k]\!] = (k-1)$ (Lemma 6.5(iv)) give $[\![\tau_{k+1}+z]\!] = \big((k-1)+1\big)\,0^z = (k)0^z$. The last claim follows from (c). □

Lemma 6.10 (cofinality). For every stripped worm Y and every ordinal $\gamma < o(Y)$ there is k with $o(Y[k]) > \gamma$. Also $o(Y[k]) < o(Y)$ for all k (Lemmas 4.2, 3.1(e), 4.4). So $\sup_k o(Y[k]) = o(Y)$.

Proof. Induction on the largest letter. Let A be the last block of Y. If $Y = Y_0\,0\,A$, then by Lemma 3.2 $Y[k] = Y_0\,0\,A[k]$, and by Lemma 4.0(ii) $o(Y[k]) = o(Y_0) + \mathrm{ord}(A[k])$ and $o(Y) = o(Y_0) + \mathrm{ord}(A)$. If $Y = A = (1)$, then $Y[k] = 0^{k+1}$ and $o(Y[k]) = k+1$, which is cofinal in $o(Y) = \omega$. If $Y = A \neq (1)$, then $o(Y[k]) = \mathrm{ord}(A[k])$ and $o(Y) = \mathrm{ord}(A)$ by Lemma 4.0(iii), since both worms begin with a nonzero letter. (If $|A| \geq 2$, $A[k]$ begins with the first letter of A. If $A = (c)$ with $c \geq 2$, then $r(A) = 0$, $B(A) = (c-1)$ and $A[k] = (c-1)^{k+1}$.) Given $\gamma < o(Y)$: if $\gamma < o(Y_0)$, then already $o(Y[0]) > \gamma$. Otherwise $\gamma = o(Y_0) + \gamma''$ with $\gamma'' < \mathrm{ord}(A)$ (resp. $\gamma'' = \gamma$), and it suffices to find k with $\mathrm{ord}(A[k]) > \gamma''$.

Case $A = D1$. $A[k] = (D0)^{k+1}$, so $\mathrm{ord}(A[k]) = \omega^{\theta(D)}\cdot(k+1)+1$. This is $k+2$ if $D = \varnothing$. Also $\mathrm{ord}(A) = \omega^{\theta(D)+1}$. Every $\gamma'' < \omega^{\theta(D)+1}$ is $< \omega^{\theta(D)}\cdot(k+1)$ for some k.

Case $A = Dc$, $c \geq 2$. By Lemma 3.4, as in the proof of Lemma 6.9(a), $A[k]$ is zero-free and nonempty with $A[k]-1 = (A-1)[k]$, so $\mathrm{ord}(A[k]) = \omega^{o((A-1)[k])}$, and $\mathrm{ord}(A) = \omega^{\theta(A)}$ with $\theta(A) = o(A-1)$. Since $A-1$ is stripped, its last CNF term has a positive exponent, so $\theta(A)$ is a limit. Let $\gamma'' < \omega^{\theta(A)}$ have leading exponent g (or $\gamma'' = 0$). Then $g < \theta(A)$, so $g+1 < \theta(A)$. By the induction hypothesis for $A-1$ (smaller largest letter) there is k with $o((A-1)[k]) > g+1$, and then $\mathrm{ord}(A[k]) \geq \omega^{g+1} > \gamma''$. □

6.5 The greedy property of the candidates

Lemma 6.11 (the expansion is the largest normal continuation). Let Y be NF and stripped, $B = B(Y)$, and let Z be a worm such that Y^-Z is NF. Then $Z \le_{\mathrm{lex}} B^\infty$. That is, either Z is a prefix of the infinite word $B^\infty = BBB\cdots$, or at the first index l with $z_l \neq (B^\infty)_l$ we have $z_l < (B^\infty)_l$.

Proof. Suppose not. Then there is $i \le |Z|$ with $z_l = (B^\infty)_l$ for $l < i$ and $z_i > (B^\infty)_i$. Put $V := Y^- z_1 \cdots z_i$, a prefix of the NF worm Y^-Z, hence NF (Lemma 6.8). For every k with $k|B| \ge i$, the NF worm $Y[k] = Y^-B^k$ (Lemma 6.9(a)) agrees with V before position $|Y^-|+i$ and has the smaller letter $(B^\infty)_i$ there. So $Y[k] <_{\mathrm{lex}} V$, and $o(Y[k]) < o(V)$ by Lemma 6.7. Also $V <_{\mathrm{lex}} Y$, since V agrees with Y on the first $p-1$ positions and has $c-1 < c$ at position $p = |Y|$. So $o(V) < o(Y)$. Thus $o(V) < o(Y)$ bounds $o(Y[k])$ for all large k, and since $o(Y[k])$ increases with k, for all k. This contradicts Lemma 6.10. □

Proposition 6.12 (the candidates are size-greedy). Let Y be NF and stripped, and $n \ge 0$. If X is any worm with $o(X) < o(Y)$ and $\mathrm{size}(X) \le \mathrm{size}(C_n(Y)) = \mathrm{size}(Y) + n$, then $o(X) \le o(C_n(Y))$. Equivalently,

$$o(C_n(Y)) = \max\{\beta < o(Y) : \|\beta\| \le \mathrm{size}(Y) + n\}.$$

Proof. Replacing X by $[\![o(X)]\!]$ does not change $o(X)$ and does not increase the size (Lemma 6.4(c)), so we may assume X is NF. Put $C := C_n(Y)$, which is NF (Lemma 6.9(b)). By Lemma 6.7, $X <_{\mathrm{lex}} Y$, and it suffices to show $X \le_{\mathrm{lex}} C$. Let $p = |Y|$ and $c = y_p$.

- If X is a proper prefix of Y, then $|X| \le p-1$. So X is a prefix of Y^-, hence a proper prefix of C (Lemma 3.1(a): C extends Y^- and $|C| \ge p+1$). Then $X <_{\mathrm{lex}} C$.
- If X first differs from Y at a position $k < p$, or at $k = p$ with $x_p < c-1$, then $X <_{\mathrm{lex}} Y^-$. Since C extends Y^-, $X <_{\mathrm{lex}} C$.
- Otherwise X agrees with Y^- on positions $1, \dots, p$, so $X = Y^-Z$ for a worm Z. By Lemma 6.11, $Z \le_{\mathrm{lex}} B^\infty$. Moreover $\mathrm{size}(Z) = \mathrm{size}(X) - \mathrm{size}(Y^-) \le n+1$. Write $b = B^\infty$ and $C = Y^-T$. By §2.1, $T = (b_1, \dots, b_i, v)$ with $v \le b_{i+1}$ and $\mathrm{size}(T) = n+1$. We show $Z \le_{\mathrm{lex}} T$.
  - If Z is a prefix of b with $|Z| \le i$, then Z is a prefix of T.
  - If Z agrees with b on the positions $1, \dots, i+1$, then $\mathrm{size}(Z) \ge \mathrm{size}(b_1, \dots, b_{i+1}) = \mathrm{size}(T) + b_{i+1} - v \ge \mathrm{size}(T)$. With $\mathrm{size}(Z) \le \mathrm{size}(T)$ this forces $v = b_{i+1}$ and $Z = (b_1, \dots, b_{i+1}) = T$.
  - Otherwise Z first differs from b at some $l \le i+1$ with $z_l < b_l$. If $l \le i$, then $Z <_{\mathrm{lex}} T$. If $l = i+1$, then $\mathrm{size}(b_1, \dots, b_i) + z_{i+1} + 1 \le \mathrm{size}(Z) \le \mathrm{size}(T) = \mathrm{size}(b_1, \dots, b_i) + v + 1$, so $z_{i+1} \le v$. Either $z_{i+1} < v$ and $Z <_{\mathrm{lex}} T$, or $z_{i+1} = v$ and Z extends T with $\mathrm{size}(Z) \le \mathrm{size}(T)$, so $Z = T$.

  Hence $X = Y^-Z \le_{\mathrm{lex}} Y^-T = C$.

For the displayed form: C itself is admissible ($o(C) < o(Y)$ by Lemma 4.4, and $\|o(C)\| \le \mathrm{size}(C)$). An ordinal $\beta < o(Y)$ with $\|\beta\| \le \mathrm{size}(Y) + n$ is $o(X)$ for $X = [\![\beta]\!]$. □

Remark. Lemmas 6.7 and 6.11 describe the candidate list combinatorially. On NF worms, $C_0(Y), C_1(Y), \dots$ are the lexicographically largest NF worms below Y of sizes $\mathrm{size}(Y), \mathrm{size}(Y)+1, \dots$. They are the size truncations of the "largest normal continuation" $Y^-BBB\cdots$ of Y^-.

6.6 Battles are optimal budgeted descents

Definition 6.13 (plays). Let $\alpha < \varepsilon_0$ and $t \in \mathbb{N}$ with $\|\alpha\| \le t$. A **play from** (α, t) is a finite sequence of ordinals

$$\alpha = \delta_0 > \delta_1 > \cdots > \delta_L = 0 \qquad \text{with} \qquad \|\delta_i\| \le t + i \quad (0 \le i \le L).$$

Think of $t+i$ as the time at which δ_i is reached. The play ends at time $t+L$. Let $E(\alpha, t)$ be the largest end time of a play from (α, t). Plays exist: for example $\alpha = \delta_0 > \delta_1 = 0$ if $\alpha > 0$, and $(\delta_0) = (0)$ if $\alpha = 0$. The plays from (α, t) form a tree under extension. It is finitely branching, since δ_i is one of the finitely many ordinals of worm norm $\le t+i$, and it has no infinite branch, since a strictly decreasing sequence of ordinals is finite. So by König's lemma there are finitely many plays, and $E(\alpha, t)$ is well defined. For a real number $s \ge \|\alpha\|$ we put $E(\alpha, s) := E(\alpha, \lfloor s \rfloor)$; this is defined, since $\|\alpha\|$ is an integer.

Theorem 6.14 (no play outlasts the battle). Let Y be NF and either empty or stripped, let $z \geq 0$, and put $S := \mathrm{size}(Y) + z$ (so $\|o(Y) + z\| = S$, since $Y0^z$ is NF). Then every play from $(o(Y) + z,\ S)$ ends at a time $\leq 2\Phi_Y(z)$, and some play ends exactly at $2\Phi_Y(z)$. That is,

$$E\big(o(Y) + z,\ \mathrm{size}(Y) + z\big) = 2\,\Phi_Y(z).$$

Proof. The run. Let $[1,0] + W_0, \ldots, [1,0] + W_T$ be the run from $\langle Y, z\rangle$ up to its first all-zero list (finite by Theorem 1). So $W_0 = Y0^z$, $W_T = 0^m$ with $m = \Phi_Y(z)$ (Definition 4.5), and W_i is not all zero for $i < T$. (If $Y = \varnothing$, then $T = 0$ and $m = z$.) For $i < T$ write $W_i = Y_i 0^{z_i}$ with Y_i stripped. Then $W_{i+1} = C_{2z_i+1}(Y_i)$ (§2.1). By Lemma 6.9(c) every W_i is NF, and so is every Y_i, as a prefix of W_i (Lemma 6.8).

Times. Put $S_i := \mathrm{size}(W_i)$. Then $S_0 = S$, $S_T = m$, and for $i < T$, by Lemma 3.1(c), $S_{i+1} = \mathrm{size}(Y_i) + 2z_i + 1 = (S_i - z_i) + 2z_i + 1 = S_i + z_i + 1$. So the intervals $I_i := [S_i, S_i + z_i] = [S_i, S_{i+1} - 1]$ $(i < T)$ are consecutive and partition the integer interval $[S, m-1]$.

The virtual descent. For $\tau \in I_i$ put $\rho_\tau := o(Y_i) + (S_i + z_i - \tau)$, and for $\tau \in [m, 2m]$ put $\rho_\tau := 2m - \tau$. Then $\rho_{S_i} = o(Y_i) + z_i = o(W_i)$ for $i < T$ (Lemma 4.0(i)), and $\rho_m = m = o(0^m) = o(W_T)$.

Norms. $\|\rho_\tau\| \leq \tau$ for every $\tau \in [S, 2m]$. For $\tau = S_i + j \in I_i$ $(0 \leq j \leq z_i)$: Y_i is NF, so $\|o(Y_i)\| = \mathrm{size}(Y_i) = S_i - z_i$, and Lemma 6.5(i) gives $\|\rho_\tau\| = S_i - z_i + (z_i - j) = S_i - j \leq \tau$. For $\tau \in [m, 2m]$: $\|2m - \tau\| = 2m - \tau \leq m \leq \tau$.

The greedy identity. For every $\tau \in [S, 2m)$:

$$\rho_{\tau+1} = \max\{\beta < \rho_\tau :\ \|\beta\| \leq \tau + 1\}. \tag{6.4}$$

There are three cases. (i) $\tau \in I_i$ with $\tau < S_i + z_i$, or $\tau \in [m, 2m)$. Then $\tau + 1$ lies in the same interval and $\rho_\tau = \rho_{\tau+1} + 1$. The ordinals below ρ_τ are those $\leq \rho_{\tau+1}$, and $\|\rho_{\tau+1}\| \leq \tau + 1$ by *Norms*. So the maximum is $\rho_{\tau+1}$. (ii) $\tau = S_i + z_i$ for some $i < T$. Then $\rho_\tau = o(Y_i)$ and $\tau + 1 = S_{i+1}$, so $\rho_{\tau+1} = o(W_{i+1}) = o(C_{2z_i+1}(Y_i))$ (for $i + 1 = T$ as well, since $\rho_m = o(W_T)$). Since Y_i is NF and stripped, Proposition 6.12 with $n = 2z_i + 1$ says that this is the maximum of $\{\beta < o(Y_i) : \|\beta\| \leq \mathrm{size}(Y_i) + 2z_i + 1\}$, and $\mathrm{size}(Y_i) + 2z_i + 1 = S_i + z_i + 1 = \tau + 1$. These cases cover $[S, 2m)$. In particular ρ is strictly decreasing on $[S, 2m]$, and $\rho_{2m} = 0$.

Domination. Let $\delta_0 > \cdots > \delta_L = 0$ be a play from $(o(Y) + z, S)$, and put $\delta'_\tau := \delta_{\tau - S}$ for $S \leq \tau \leq S + L$. So $\|\delta'_\tau\| \leq \tau$, and $\delta'_S = o(Y) + z = \rho_S$. We show $\delta'_\tau \leq \rho_\tau$ for $S \leq \tau \leq \min(S + L, 2m)$, by induction on τ. If $\delta'_\tau \leq \rho_\tau$ and $\tau < \min(S + L, 2m)$, then $\delta'_{\tau+1} < \delta'_\tau \leq \rho_\tau$ and $\|\delta'_{\tau+1}\| \leq \tau + 1$, so $\delta'_{\tau+1} \leq \rho_{\tau+1}$ by (6.4). Now suppose $S + L > 2m$. Then $\delta'_{2m} \leq \rho_{2m} = 0$, but $\delta'_{2m} > \delta'_{S+L} = 0$ because the play is strictly decreasing and $2m < S + L$. This is a contradiction. So every play ends at a time $S + L \leq 2m$.

Attainment. The sequence $(\rho_S, \rho_{S+1}, \ldots, \rho_{2m})$ is a play from $(o(Y) + z, S)$: it starts at $\rho_S = o(Y) + z$, it is strictly decreasing with last term $\rho_{2m} = 0$, and its i-th term ρ_{S+i} has norm $\leq S + i$ by *Norms*. It ends at time $2m$. Hence $E(o(Y) + z, S) = 2m = 2\Phi_Y(z)$. □

So a battle is the longest strictly decreasing sequence of ordinals that starts at $o(Y) + z$ at time $\mathrm{size}(Y) + z$ and whose term at time τ has a worm of size $\leq \tau$. The battle realizes it greedily: by (6.4), its next ordinal is always the largest one allowed.

Remark (all NF worms). The identity holds for every NF worm X and every $z \geq 0$:

$$\Phi_X(z) = \tfrac{1}{2}\,E\big(o(X) + z,\ \mathrm{size}(X) + z\big).$$

Indeed, write $X = Y0^j$ with Y empty or stripped. Then Y is NF (Lemma 6.8), $\Phi_X(z) = \Phi_Y(z + j)$ (Corollary 4.6), $o(X) + z = o(Y) + (z + j)$ (Lemma 4.0(i)) and $\mathrm{size}(X) + z = \mathrm{size}(Y) + (z + j)$, so Theorem 6.14 for $(Y, z + j)$ applies.

Two consequences of Theorem 6.14:

Corollary 6.15. For all $k \geq 1$ and $z \geq 0$: $\Phi_{(k)}(z) = \frac{1}{2}E(\tau_{k+1} + z,\ k + 1 + z)$.

Proof. Theorem 6.14 with $Y = (k) = [\![\tau_{k+1}]\!]$ (Lemma 6.5(iv)). □

Corollary 6.16 (monotonicity of battles). Let X, X' be NF worms, each empty or stripped, and $z, z' \geq 0$. If $o(X) + z \leq o(X') + z'$ and $\mathrm{size}(X) + z \leq \mathrm{size}(X') + z'$, then $\Phi_X(z) \leq \Phi_{X'}(z')$.

Proof. Let $\alpha = o(X) + z$, $S = \mathrm{size}(X) + z$, $\alpha' = o(X') + z'$, $S' = \mathrm{size}(X') + z'$, and let $\rho_S, \dots, \rho_{2\Phi_X(z)}$ be the play of Theorem 6.14 for X. If $\alpha = \alpha'$, then $(\rho_S, \rho_{S+1}, \dots)$ is also a play from (α', S'), because $\|\rho_{S+i}\| \leq S + i \leq S' + i$. Its end time is $S' + 2\Phi_X(z) - S \geq 2\Phi_X(z)$. If $\alpha < \alpha'$, then $(\alpha', \rho_S, \rho_{S+1}, \dots)$ is a play from (α', S'), because $\|\rho_{S+i}\| \leq S + i \leq S' + 1 + i$, and it ends at time $\geq 2\Phi_X(z) + 1$. In both cases Theorem 6.14 for X' gives $2\Phi_{X'}(z') \geq 2\Phi_X(z)$. □

Corollary 6.16 compares the battles of different worms: on NF worms, Φ depends only on the ordinal and the total size, monotonically. It is not used below, since the lower bounds go through plays directly.

6.7 Building plays

Lemma 6.17 (moves). Let $\|\alpha\| \leq t$.

- (a) *(Move.)* If $\beta < \alpha$ and $\|\beta\| \leq t + 1$, then $E(\alpha, t) \geq E(\beta, t + 1)$.
- (b) *(More time.)* If $t \leq t'$, then $E(\alpha, t) \leq E(\alpha, t')$. This also holds for real $t \leq t'$.
- (c) *(Tail.)* $E(\gamma + K, t) \geq E(\gamma, t + K)$ for all γ and finite K with $\|\gamma + K\| \leq t$.
- (d) *(Finite.)* $E(m, t) \geq t + m$ for finite m.
- (e) *(Sum.)* Let $\rho \geq \omega$ have no finite part, let $\zeta \geq \omega$ have all exponents $\leq$ the last exponent of ρ, and let $\|\rho + \zeta\| \leq t$. Then

$$E(\rho + \zeta, t) \;\geq\; E\big(\rho,\; E(\zeta,\, t - \|\rho\| - 1) + \|\rho\| + 1\big).$$

Proof. (a) Put α (at time t) in front of a play from $(\beta, t + 1)$ that ends at time $E(\beta, t + 1)$. (b) A play from (α, t) is a play from (α, t'), and it ends $t' - t$ later. For real times, use that $\lfloor \cdot \rfloor$ is monotone. (c) The sequence $\gamma + K, \gamma + K - 1, \dots, \gamma$, with $\gamma + K - i$ at time $t + i$, has norms $\|\gamma + K - i\| = \|\gamma + K\| - i \leq t + i$ (Lemma 6.5(i)). Its last term γ is reached at time $t + K$; continue with a play from $(\gamma, t + K)$ that ends at time $E(\gamma, t + K)$. (d) Walk down $m, m - 1, \dots, 0$, with $m - i$ at time $t + i$: the norms are $m - i \leq t + i$, since $\|m\| = m \leq t$. This play ends at time $t + m$.

(e) By Lemma 6.5(ii), $\|\rho + \zeta\| = \|\rho\| + 1 + \|\zeta\|$, so $t_\zeta := t - \|\rho\| - 1 \geq \|\zeta\|$. Take a play $\zeta = \zeta_0 > \dots > \zeta_L = 0$ from (ζ, t_ζ) with $t_\zeta + L = E(\zeta, t_\zeta)$. Consider $\rho + \zeta_0 > \rho + \zeta_1 > \dots > \rho + \zeta_L = \rho$, with $\rho + \zeta_i$ at time $t + i$. For $i < L$ we have $1 \leq \zeta_i \leq \zeta$, so every exponent of ζ_i is $\leq$ its leading exponent $\leq$ the leading exponent of ζ ((F2)) $\leq$ the last exponent of ρ, and Lemma 6.5(ii) applies to $\rho + \zeta_i$. It gives $\|\rho + \zeta_i\| = \|\rho\| + 1 + \|\zeta_i\|$ if $\zeta_i \geq \omega$, and $\|\rho + \zeta_i\| = \|\rho\| + \zeta_i$ if $\zeta_i < \omega$. In both cases $\|\rho + \zeta_i\| \leq \|\rho\| + 1 + \|\zeta_i\| \leq \|\rho\| + 1 + t_\zeta + i = t + i$. The last term ρ is reached at time $t + L = E(\zeta, t_\zeta) + \|\rho\| + 1$, and $\|\rho\| \leq t + L$. Continue with a play from $(\rho, t + L)$ that ends at time $E(\rho, t + L)$. □

Lemma 6.18 (finite exponents). Let $\rho = 0$, or $\rho \geq \omega$ without finite part.

- (a) $E(\omega, t) \geq 2t + 2$ for $t \geq 2$. If $\rho \neq 0$ and $t \geq \|\rho\| + 3$, then $E(\rho + \omega, t) \geq E(\rho,\, 2t + 2 - \|\rho\|)$.
- (b) If $\rho \neq 0$, $k \geq 0$, and $\sigma := t - \|\rho + \omega \cdot k\| \geq 0$, then $E(\rho + \omega \cdot k, t) \geq E\big(\rho,\; \|\rho\| + 2^k(\sigma + 8) - 8\big)$. If $k \geq 1$ and $\sigma := t - \|\omega \cdot k\| \geq 0$, then $E(\omega \cdot k, t) \geq 2^k(\sigma + 8) - 10$.
- (c) $E(\omega^2, t) \geq 2^{(t-3)/6}(t + 2)$ for $t \geq 4$. If $\rho \neq 0$ has last exponent ≥ 2 and $\sigma := t - \|\rho + \omega^2\| = t - \|\rho\| - 5 \geq 0$, then $E(\rho + \omega^2, t) \geq E\big(\rho,\; \|\rho\| + 2^{(\sigma+1)/6}(\sigma + 6)\big)$.

Proof. (a) $t + 1 < \omega$ and $\|t + 1\| = t + 1$, so by Lemma 6.17(a),(d), $E(\omega, t) \geq E(t + 1, t + 1) \geq 2t + 2$. For $\rho \neq 0$ put $K := t + 1 - \|\rho\| \geq 4$. Then $\rho + K < \rho + \omega$ and $\|\rho + K\| = \|\rho\| + K = t + 1$, so by Lemma 6.17(a),(c), $E(\rho + \omega, t) \geq E(\rho + K, t + 1) \geq E(\rho, t + 1 + K) = E(\rho, 2t + 2 - \|\rho\|)$.

(b) First claim, by induction on k. For $k = 0$ it is trivial. For $k \geq 1$, let $\rho' := \rho + \omega \cdot (k - 1)$, which is $\geq \omega$ and has no finite part. By Lemma 6.5(iii), $\|\rho + \omega k\| = \|\rho'\| + 3$, so (a) applies: $E(\rho + \omega k, t) \geq E(\rho', t')$ with $t' = 2t + 2 - \|\rho'\|$. The new slack is $t' - \|\rho'\| = 2(t - \|\rho'\|) + 2 = 2(\sigma + 3) + 2 = 2\sigma + 8$. By induction, $E(\rho', t') \geq E(\rho, \|\rho\| + 2^{k-1}(2\sigma + 16) - 8) = E(\rho, \|\rho\| + 2^k(\sigma + 8) - 8)$. Second claim: for $k = 1$, (a) gives $E(\omega, t) \geq 2t + 2 = 2(\sigma + 2) + 2 \geq 2(\sigma + 8) - 10$. For $k \geq 2$, the first claim with $\rho = \omega$ and $k - 1$ copies ($\|\omega + \omega(k - 1)\| = \|\omega k\|$) gives $E(\omega k, t) \geq E(\omega, 2 + 2^{k-1}(\sigma + 8) - 8) \geq 2^k(\sigma + 8) - 10$ by (a).

(c) First claim. Let $k := \lfloor (t+2)/6 \rfloor \geq 1$ and $K := t+2-3k \geq (t+2)/2$. Then $\beta := \omega \cdot k + K < \omega^2$ and $\|\beta\| = 3k - 1 + K = t+1$. By Lemma 6.17(a),(c), $E(\omega^2, t) \geq E(\beta, t+1) \geq E(\omega k, t+1+K)$. Here the slack is $t+1+K-(3k-1) = 2K$, so by (b) $E(\omega^2, t) \geq 2^k(2K+8) - 10 \geq 2^k(t+2)$. Finally $k \geq (t+2)/6 - 5/6 = (t-3)/6$. Second claim. Let $k := \lfloor (\sigma+6)/6 \rfloor \geq 1$ and $K := \sigma + 6 - 3k \geq (\sigma+6)/2$. Then $\beta := \rho + \omega \cdot k + K < \rho + \omega^2$ and $\|\beta\| = \|\rho\| + 3k + K = \|\rho\| + \sigma + 6 = t+1$. So, by Lemma 6.17(a),(c), $E(\rho + \omega^2, t) \geq E(\beta, t+1) \geq E(\rho + \omega k, t+1+K)$, whose slack is $t+1+K-\|\rho\|-3k = 2K \geq \sigma+6$. By (b) and Lemma 6.17(b), $E(\rho+\omega^2, t) \geq E(\rho, \|\rho\| + 2^k(\sigma+14) - 8) \geq E(\rho, \|\rho\| + 2^{(\sigma+1)/6}(\sigma+6))$, since $k \geq (\sigma+1)/6$. □

6.8 A Hardy bound for plays

Theorem 6.19. For every ordinal $1 \leq \alpha < \varepsilon_0$ and every integer $t \geq 8\|\alpha\|^2 + 1000$:

$$E(\alpha, t) \;\geq\; H_\alpha\big(\lfloor t/(8\|\alpha\|) \rfloor\big).$$

Proof. Transfinite induction on α. Put $n := \|\alpha\| \geq 1$ and $y := \lfloor t/(8n) \rfloor$. Since $t \geq 8n^2 + 1000$, we have $y \geq n \geq 1$ and $n \leq t/(8n) \leq t/8$. Lemma A.6 (H_γ is increasing) is used tacitly throughout.

Case 1: $\alpha = m$ finite. By Lemma 6.17(d), $E(m, t) \geq t + m \geq y + m = H_m(y)$.

Case 2: $\alpha = \gamma + 1$ with $\gamma \geq \omega$. Then $\|\gamma\| = n - 1$ (Lemma 6.5(i)). By Lemma 6.17(a), and by the induction hypothesis at time $t + 1 \geq 8(n-1)^2 + 1000$,

$$E(\alpha, t) \geq E(\gamma, t+1) \geq H_\gamma\big(\lfloor (t+1)/(8(n-1)) \rfloor\big).$$

Now $\frac{t+1}{8(n-1)} - \frac{t}{8n} = \frac{t+n}{8n(n-1)} \geq 1$ because $t \geq 8n^2$. Hence $\lfloor (t+1)/(8(n-1)) \rfloor \geq y+1$, and $E(\alpha, t) \geq H_\gamma(y+1) = H_{\gamma+1}(y)$.

Case 3: $\alpha = \lambda$ a limit. Write $\lambda = \rho + \omega^\delta$ in CNF, with $\delta \geq 1$ the last exponent. Here $\rho = 0$, or $\rho \geq \omega$ has no finite part and all its exponents are $\geq \delta$. Since λ is a limit, $H_\lambda(y) = H_{\lambda[y]}(y)$.

3a: $\delta = 1$. If $\rho = 0$, then $E(\omega, t) \geq 2t + 2 \geq 2y + 1 = H_\omega(y)$ by Lemma 6.18(a) ($H_\omega(y) = H_{y+1}(y) = 2y+1$, as $\omega[y] = y+1$). If $\rho \neq 0$, let $m := \|\rho\| = n - 3$ (Lemma 6.5(iii)) and $t' := 2t + 2 - m \geq t$, an integer with $t' \geq t \geq 8m^2 + 1000$. Lemma 6.18(a) (note $t \geq n = m+3$) and the induction hypothesis for $\rho < \lambda$ at time t' give $E(\lambda, t) \geq E(\rho, t') \geq H_\rho(\lfloor t'/(8m) \rfloor)$. By Lemma A.8(a), since $\rho + \omega$ is a normal sum, $H_\lambda(y) = H_\rho(H_\omega(y)) = H_\rho(2y+1)$. It remains to check $t'/(8m) \geq 2y+1$:

$$\frac{t'}{8m} - (2y+1) \;\geq\; \frac{2t-m}{8m} - \frac{2t}{8(m+3)} - 1 \;=\; \frac{6t}{8m(m+3)} - \frac{9}{8} \;\geq\; 0,$$

since $t \geq 8n^2 = 8(m+3)^2 \geq \frac{3}{2}m(m+3)$.

3b: $\delta = 2$. If $\rho = 0$, then $n = \|\omega^2\| = 4$, $y \leq t/32$, and by Lemma 6.18(c)

$$E(\omega^2, t) \geq 2^{(t-3)/6}(t+2) \;\geq\; 2^{t/32+1}(t/32+1) \;\geq\; 2^{y+1}(y+1) - 1 = F_2(y) = H_{\omega^2}(y),$$

since $(t-3)/6 \geq t/32 + 1$ for $t \geq 1000$; the last equality is Lemma A.8(b). If $\rho \neq 0$, then every exponent of ρ is ≥ 2, ρ has no finite part, and $m := \|\rho\| = n - 5$ (Lemma 6.5(iii)). Here $m \geq 3$. Indeed, by Lemma 6.4(a) the first block of $[\![\rho]\!]$ is $[\![\gamma]\!] + 1$, where $\gamma \geq 2$ is the leading exponent of ρ. If $[\![\gamma]\!]$ is a single letter x, then $x \geq 1$ (since $o((0)) = 1 < \gamma$), and the block $(x+1)$ has size ≥ 3. Otherwise the block has at least two letters, each ≥ 1, and size ≥ 4. (The smallest case is $\|\omega^\omega\| = \mathrm{size}((2)) = 3$.) So $n \geq 8$, $y \leq t/64$, and $\sigma := t - n \geq 7t/8$ (as $n \leq t/8$). Lemma 6.18(c) gives $E(\lambda, t) \geq E(\rho, X')$ with

$$X' := m + 2^{(\sigma+1)/6}(\sigma+6) \;\geq\; 2^{7t/48} \cdot \tfrac{7t}{8}.$$

For $t \geq 1000$ we have $2^{7t/48} \geq t$, so $X' \geq 7t^2/8 \geq t^2/8 + 1000 \geq 8m^2 + 1000$ (as $m \leq t/8$). So the induction hypothesis for $\rho < \lambda$, at the integer time $\lfloor X' \rfloor \geq 8m^2 + 1000$, gives $E(\rho, X') = E(\rho, \lfloor X' \rfloor) \geq$

$H_\rho(\lfloor\lfloor X'\rfloor/(8m)\rfloor) = H_\rho(\lfloor X'/(8m)\rfloor)$. Next, $H_\lambda(y) = H_\rho(H_{\omega^2}(y)) = H_\rho(F_2(y))$ by Lemma A.8(a) ($\rho + \omega^2$ is a normal sum) and Lemma A.8(b), and $F_2(y) = 2^{y+1}(y+1) - 1 \le 2^{t/64+1}(t+1)$. With $m \le t$:

$$\frac{X'}{8m\,F_2(y)} \;\ge\; \frac{2^{7t/48}\cdot 7t/8}{8t\cdot 2^{t/64+1}(t+1)} \;=\; \frac{7\cdot 2^{25t/192}}{128\,(t+1)} \;\ge\; 1 \qquad (t \ge 1000),$$

since $\frac{7}{48} - \frac{1}{64} = \frac{25}{192}$. So $\lfloor X'/(8m)\rfloor \ge F_2(y)$ and $E(\lambda,t) \ge H_\rho(F_2(y)) = H_\lambda(y)$.

3c: $\delta \ge 3$ or δ a limit. Let $\lambda' := \lambda[y] = \rho + (\omega^\delta)[y]$ (Definition A.1). If $\delta = e+1$ is a successor, then $(\omega^\delta)[y] = \omega^e\cdot(y+1)$ with $e \ge 2$. If δ is a limit, then $(\omega^\delta)[y] = \omega^e$ with $e := \delta[y]$, and $e \ge y+1 \ge 2$ by Lemma A.9, applied to the limit δ and the finite ordinal $y < \delta$, whose norm is $N(y) = y$. In both cases $e < \delta$, which is at most every exponent of ρ. So the CNF of λ' is that of ρ followed by copies of ω^e. Thus $\lambda' \ge \omega$ has no finite part and last exponent $e \ge 2$, and the CNF of $\beta := \lambda' + \omega^2$ is that of λ' followed by ω^2. Also $\beta < \lambda$: $(\omega^\delta)[y] < \omega^\delta$ (Lemma A.2(a)) and $\omega^2 < \omega^\delta$, a sum of two ordinals below ω^δ is below ω^δ ((F2), as in Lemma 4.0(iv)), and left addition of ρ is strictly increasing. By Lemma 6.5(v), and by Lemma 6.5(iii) for β,

$$\|\lambda'\| \le (y+1)n \le t/8 + n \le t/4, \qquad \|\beta\| = \|\lambda'\| + 5 \le t+1,$$

using $yn \le t/8$ and $n \le t/8$. By Lemma 6.17(a), and then Lemma 6.17(e) with λ', ω^2 in the roles of ρ, ζ at time $t+1$:

$$E(\lambda,t) \;\ge\; E(\beta,t+1) \;\ge\; E(\lambda',T), \qquad T := E(\omega^2, t_\zeta) + \|\lambda'\| + 1, \quad t_\zeta := t - \|\lambda'\| \ge 3t/4.$$

By Lemma 6.18(c) ($t_\zeta \ge 4$), $T \ge E(\omega^2,t_\zeta) \ge 2^{(t_\zeta-3)/6}(t_\zeta+2) \ge X := 2^{(3t/4-3)/6}(3t/4+2) \ge 2^{t/8-1/2}\cdot 3t/4$. For $t \ge 1000$ we have $2^{t/8-1/2} \ge t$, so $X \ge 3t^2/4 \ge t^2/2 + 1000 \ge 8\|\lambda'\|^2 + 1000$. Since $1 \le \lambda' < \lambda$, the induction hypothesis for λ' at the integer time T gives $E(\lambda',T) \ge H_{\lambda'}(\lfloor T/(8\|\lambda'\|)\rfloor)$. Finally $8\|\lambda'\|\,y \le 8\cdot\frac{t}{4}\cdot\frac{t}{8n} \le \frac{t^2}{4} \le X \le T$, so $\lfloor T/(8\|\lambda'\|)\rfloor \ge y$. Hence $E(\lambda,t) \ge H_{\lambda[y]}(y) = H_\lambda(y)$. □

Remark (where the counter resets went). In 3c the play pays $\|\lambda[y]\| - \|\lambda\| \le y\,\|\lambda\| \le t/8$ units of budget for the index y, a fixed fraction of its time t. This is the analogue of the reset of the counter of R. Right after that move the time is no longer large compared with $\|\lambda[y]\|^2$, so the induction hypothesis would not apply. The play regains time by spending the cheap summand ω^2 first. That summand is worth exponentially more time than its norm 4, so afterwards the play sits at $\lambda[y]$ with time $T \ge 3t^2/4 \ge 8\|\lambda[y]\|^2 + 1000$. In the battle itself the corresponding effect is that short pieces of the worm fight sub-battles that return very large counters (§6.10(c)). Here it is a single explicit move per limit, because Theorem 6.14 lets a lower bound use *any* play, not only the battle itself.

6.9 Single-letter battles, LEMMA G* and LEMMA G

Theorem 6.20 (single letters). For all $k \ge 1$ and $z \ge 0$ with $2z + k + 1 \ge 8(k+1)^2 + 1000$:

$$\Phi_{(k)}(z) \;\ge\; \tfrac{1}{2}\,H_{\tau_{k+1}}\Big(\Big\lfloor \frac{2z+k+1}{8(k+1)}\Big\rfloor\Big).$$

Proof. By Corollary 6.15, Lemma 6.17(c), and Theorem 6.19 with $\|\tau_{k+1}\| = k+1$ (Lemma 6.5(iv)): $2\Phi_{(k)}(z) = E(\tau_{k+1}+z, k+1+z) \ge E(\tau_{k+1}, k+1+2z) \ge H_{\tau_{k+1}}(\lfloor(2z+k+1)/(8(k+1))\rfloor)$. □

Corollary 6.21 (LEMMA G*). For every $m \ge 1$, with $j := \max(m,3)$, we have $\Phi_{(j)}(z) \ge H_{\tau_m}(z)$ for all sufficiently large z.

Proof. First let $m \ge 3$ and $j = m$. Put $u := \lfloor(2z+m+1)/(8(m+1))\rfloor$, and let z be so large that $2z+m+1 \ge 8(m+1)^2+1000$, $u \ge 3$ and $2^u \ge z+1$; the last holds for large z because $u \ge z/(4(m+1)) - 1$. We use Corollary A.11 three times:

- $\tau_m\cdot 3 < \tau_{m+1}$ and $N(\tau_m\cdot 3) = 3 \le u$ give $H_{\tau_{m+1}}(u) \ge H_{\tau_m\cdot 3}(u) = H_{\tau_m}(H_{\tau_m}(H_{\tau_m}(u)))$. The last equality is Lemma A.8(a), applied twice, since $\tau_m\cdot 2 + \tau_m$ and $\tau_m + \tau_m$ are normal sums.
- $\omega^2 < \tau_m$ and $N(\omega^2) = 2 \le u$ give $H_{\tau_m}(u) \ge H_{\omega^2}(u) = F_2(u) \ge 2^u \ge z+1$ (Lemma A.8(b)).
- $\omega < \tau_m$ and $N(\omega) = 1$ give $H_{\tau_m}(w) \ge H_\omega(w) = 2w+1$ for $w \ge 1$.

Hence, by Theorem 6.20 and Lemma A.6 (monotonicity of H_{τ_m}),

$$\Phi_{(m)}(z) \;\geq\; \tfrac{1}{2}H_{\tau_{m+1}}(u) \;\geq\; \tfrac{1}{2}H_{\tau_m}\big(H_{\tau_m}(z+1)\big) \;\geq\; \tfrac{1}{2}\big(2H_{\tau_m}(z+1)+1\big) \;>\; H_{\tau_m}(z+1) \;>\; H_{\tau_m}(z).$$

For $m \in \{1,2\}$ take $j=3$. For large z we get $\Phi_{(3)}(z) \geq H_{\tau_3}(z) \geq H_{\tau_m}(z)$, by the case $m=3$ and Corollary A.11 ($\tau_m < \tau_3$, $N(\tau_m) \leq 1 \leq z$). $\square$

Corollary 6.22 (LEMMA G). For every $\beta < \varepsilon_0$ there is K with $\Psi(k) \geq H_\beta(k+1)$ for all $k \geq K$. Moreover, $\Psi(k) \geq H_{\tau_{\lfloor k/2 \rfloor}}(k)$ for all sufficiently large k.

Proof. The first statement follows from Corollary 6.21 and Proposition 6.1. For the second, let k be large and $j := \lfloor k/2 \rfloor \geq 3$. By Corollary 5.3(c), $\Psi(k) \geq \Phi_{(j)}(Z)$ with $Z := 2^{k-j+1} - 2 \geq 2^{k/2}$. For large k, $2Z + j + 1 \geq 8(j+1)^2 + 1000$, and $u := \lfloor (2Z+j+1)/(8(j+1)) \rfloor \geq k+1$. Theorem 6.20 then gives $\Psi(k) \geq \frac{1}{2}H_{\tau_{j+1}}(u) \geq \frac{1}{2}H_{\tau_{j+1}}(k+1)$. Next, $H_{\tau_{j+1}}(k+1) \geq H_{\tau_j \cdot 2}(k+1) = H_{\tau_j}(H_{\tau_j}(k+1)) \geq 2H_{\tau_j}(k+1)+1$. Here the first step is Corollary A.11 ($\tau_j \cdot 2 < \tau_{j+1}$, and $N(\tau_j \cdot 2) = 2 \leq k+1$), the equality is Lemma A.8(a) for the normal sum $\tau_j + \tau_j$, and the last step is $H_{\tau_j}(w) \geq H_\omega(w) = 2w+1$ for $w \geq 1$ (Corollary A.11, as in the proof of Corollary 6.21). So $\Psi(k) \geq \frac{1}{2}\big(2H_{\tau_j}(k+1)+1\big) > H_{\tau_j}(k+1) \geq H_{\tau_j}(k)$. $\square$

The second statement is the hypothesis of Proposition 6.2 with $f(k) = \lfloor k/2 \rfloor$, so it also gives LEMMA G directly. Since $\mathrm{ord}((k)) = \tau_{k+1}$, a battle from (k) keeps, at counter 0, about half of its tower of omegas. By Theorem 6.20, at a large counter it keeps all of it.

6.10 Remarks: why a direct simulation fails, and how this section avoids it

Three features of R block a step-by-step comparison of a battle with a Buchholz–Wainer descent, of the kind that works for the Kirby–Paris hydra battle.

- *(a) The step is not a Buchholz–Wainer step.* For stripped Y, the expansion $Y[j]$ (Lemma 3.1(e)) is candidate number $jW(Y) - 1$. Here $W(Y) = \mathrm{size}(B(Y))$ is the *width* of the periodic part. So with counter z the process reaches only about $2z/W(Y)$ copies of $B(Y)$, and when $W(Y) > 2z+2$ it takes a partial copy. That copy can lie below $\mathrm{ord}(Y)[1]$. For example, from $\langle (7), 2\rangle$ (ordinal $\tau_8 + 2$) the step goes to $C_5((7)) = (6,5)$, and $\tau_8[0] < \mathrm{ord}((6,5)) < \tau_8[1]$ (computed with `ordw.py`, §9).
- *(b) The counter is reset.* The next counter is the number of trailing zeros of the chosen candidate, unrelated to z. So the process offers no invariant of the form "counter $\geq$ (a function of the Buchholz–Wainer argument)".
- *(c) The counter is rebuilt by sub-battles.* After a wide step, the partial copy fights a sub-battle of its own, which returns a counter far larger than the width. For instance $\Phi_{P1}(z) = \Phi_P\big(1 + \Phi_{T_{2z+1}(P)}(0)\big)$ (Lemma 3.3 and Corollary 4.6).

This section never follows the battle step by step.

- Proposition 6.12 shows that the candidate $C_{2z+1}(Y)$ is the best ordinal the size budget allows. Theorem 6.14 then compares the battle with *every* budgeted descent at once.
- The lower bound follows a Buchholz–Wainer descent with index $y = \lfloor t/(8\|\lambda\|) \rfloor$ (Theorem 6.19). The width shows up only as the division by $\|\lambda\|$, through Lemma 6.5(v).
- The budget is the total size t, which is never reset; it grows by $z+1 \geq 1$ per step (Lemma 3.1(c)).
- The rebuilding by sub-battles corresponds to the regeneration move in case 3c of Theorem 6.19, which spends a cheap summand ω^2 first.

The one-step comparison $\Phi_Y(z) \geq H_{\mathrm{ord}(Y)[0]}(z)$, for all stripped Y and all z, holds in all small cases computed (§9), but it is neither proved nor needed. Theorem 6.19 needs a large time $t \geq 8\|\alpha\|^2 + 1000$, and Corollary 6.22 shows that from counter 0 a single-letter battle keeps at least $\lfloor k/2 \rfloor$ of its $k+1$ tower levels.

Racheline's informal description of the hierarchy computed by M_{51} ($h_{\lim}(n) = h_{\lim[n]}(c_2)$, "comparable to f_{ε_0}") is not used anywhere in this document. It cannot be taken literally, since the process does not apply $\lim[n]$ with n the counter (remark (a)). Its conclusion, growth at every level below ε_0, is confirmed by Theorem 6.20.

6.11 Numerical checks of this section (not part of any proof)

Scripts and outputs are in the folder `lemmaG_work/` of `checks51.tar.gz` (see also §9). They use `ordw.py`, `r51.py` and `hardy.py` from `u51_lemmaG/`. "Tight" means a worm of minimal size for its ordinal.

- `nfchecks.py` (`nfchecks.out`). Definition 6.3 agrees with tightness on all 16384 worms of size ≤ 14. Lemma 6.6 holds: 7641 checks. Lemma 6.11 holds: 4645 checks. No failures.
- `lexorder.py` (`lexorder.out`), all worms of size ≤ 16. Each ordinal has exactly one tight worm. Tight worms are closed under prefixes. On tight worms the ordinal order is the lexicographic order: 3415 adjacent pairs checked. No failures.
- `claimA2.py` (`claimA2_S19.out`) and `claimAj.py` (`claimAj.out`). Proposition 6.12 holds on every tight stripped worm of size ≤ 18 and every candidate of size ≤ 19: 11359 checks. All candidates are tight. The analogous greedy statement with the norms ν_j ($j \le 3$) in place of size also holds: 6070 checks. It is not used in the proofs. No failures.
- `normcheck.py` (`normcheck.out`). Lemma 6.4(a),(b) and the fundamental-sequence norm bound Lemma 6.5(v) (for $y \le 4$, $j \le 2$) hold on 1762 ordinals. No failures.
- `battle_nf.py` (`battle_nf.out`). Lemma 6.9(c),(d) along actual runs of `ref51.step`: the run of R from $[1,0]$ (until the list is longer than 300), the runs from $\langle (k), z\rangle$ for $k \le 6$, $z \le 7$, and 600 runs from random NF worms with random counters. In total 39358 steps are checked, including 63 beeps with their restarts. Every list is $[1,0]+X$ with X NF. At each non-beep step the size grows as in Lemma 3.1(c) and the ordinal drops. No failures.
- `brute_E2.py` (`brute_E2.out`). Theorem 6.14 by exhaustive search, for 26 small NF states, among them $\langle (1), z\rangle$ for $z \le 11$, $\langle (1,1), z\rangle$ for $z \le 4$, $\langle (2), 0\rangle$ and $\langle (1,1,1), 0\rangle$. Put $N := 2\Phi_Y(z)+1$. The script computes exactly, by dynamic programming over *all* ordinals below $o(Y)+z$ of norm $\le N$ (up to 74994 of them), the largest end time E_N of a play that uses only such ordinals and ends by time N. If $E_N \le N-1$, then $E = E_N$. (A play ending at $T \le N$ is counted; a play ending later can be cut at time $N-1$ and sent to 0, which would give $E_N \ge N$.) In every case $E_N = 2\Phi_Y(z)$, with Φ from `phis.py`. So the identity $E(o(Y)+z, \mathrm{size}(Y)+z) = 2\Phi_Y(z)$ is verified exactly for these states. The older `brute_E.py` (`brute_E.out`) enumerated only ordinals of norm ≤ 24, so for its larger states it checked only the attainment half.
- `monocheck.py` (`monocheck.out`). Corollary 6.16 holds on all 649 comparable pairs of computable states.

The values of Φ are far too large for Theorems 6.19–6.20 to be tested directly. The checks above test the combinatorial lemmas on which they rest.

---

7. Theorem 2: unprovability for R

The classification theorem enters as follows. Theorem A.13 (Wainer 1970; Buchholz–Wainer 1987, Theorem 5) says that every function provably computable in PA, in the sense of Definition A.12, is majorised by $F_{\alpha+1} = H_{\omega^{\alpha+1}}$ for some $\alpha < \varepsilon_0$ (Lemma A.8(b)). Lemma A.14 turns a Π_2 theorem $\forall x\, \exists y\, \theta(x,y)$ of first-order PA with $\theta \in \Delta_0$ into such a function, $x \mapsto$ least y with $\theta(x,y)$. In Lemma A.14, Buchholz–Wainer's axiom group (f) is read as the defining equations of each elementary *definition*, so bounded sums and products come with their recursion equations (see the proof of Lemma A.14).

Theorem 2.

- (a) For every function f provably computable in PA: $f(j) < k_{j+1}$ for all large j, and $f(i) < \tau_q(i)$ for all large i and $q \in \{30, 31\}$ (the beep times of M_{51}, §2.2).
- (b) PA $\nvdash \mathsf{Inf}_R$.

Proof. (a) Take $\alpha < \varepsilon_0$ with $f(x) \le H_{\omega^{\alpha+1}}(x)$ for all large x, and put $\beta = \omega^{\alpha+1}$. LEMMA G (Corollary 6.22) for $\beta+1$ gives K with $\Psi(k) \ge H_{\beta+1}(k+1) = H_\beta(k+2)$ for $k \ge K$. Since $k_j \ge j$ (Corollary 5.3(d)), for $j \ge K$:

$$k_{j+1} = 1 + \Psi(k_j) > H_\beta(k_j+2) \ge H_\beta(j+2) > H_\beta(j+1) > H_\beta(j) \ge f(j),$$

using Lemma A.6. And by Corollary 2.3, $\tau_q(j+2) > k_{j+1} > H_\beta(j+2) \geq f(j+2)$ for large j.

(b) Suppose PA $\vdash \mathsf{Inf}_R$, i.e. PA $\vdash \forall j\, \exists w\, \theta_R(j,w)$ with $\theta_R \in \Delta_0$ (§2.4). By Lemma A.14, $g(j) :=$ least w with $\theta_R(j,w)$ is provably computable in PA. By (R1), $g(j) \geq k_j + 1 > k_j$ for $j \geq 2$. So $g(j+1) > k_{j+1}$ for all $j \geq 1$. This contradicts (a) applied to $f(j) := g(j+1)$, which is provably computable, being g composed with the successor function. □

Remarks.

1. For each fixed j, PA proves "R has a j-th beep", since it is a true Σ_1 sentence.
2. For each fixed height h, the battles from worms with letters $\leq h$ involve only ordinals $< \tau_{h+2}$ (Lemma 4.0(v)). So PA proves that they terminate, since PA proves transfinite induction below each fixed $\alpha < \varepsilon_0$ (a classical theorem of Gentzen; this remark is not used in any proof). The unprovability comes entirely from the unbounded starting letters k_j.
3. Theorem 1 does not depend on §6.
4. From §6, the proof of Theorem 2 uses only Corollary 6.22 (LEMMA G).

---

8. Theorem 3: the machine M_{51}

Theorem 3.

1. $\neg\mathsf{QH}$ is true: M_{51} (started in state 30 on the blank tape) enters state 50 infinitely often.
2. PA $\nvdash \mathsf{QH}$.
3. PA $\nvdash \neg\mathsf{QH}$. So whether M_{51} quasihalts is independent of PA.

The only input about M_{51} is the machine-checked statement `M51_run30` of §2.3 (with `Steps_run`), through (Sim1), (Sim2) and Corollary 2.3. Only its truth is used. The same holds for the start-31 run, with `M51_run_fun` in place of `M51_run30`.

Proof. (1) Theorem 1 and (Sim2). (2) PA is sound and QH is false by (1); only soundness for Σ_2 sentences is used. (3) Suppose PA $\vdash \neg\mathsf{QH}$. By Lemma 2.4, PA $\vdash \forall k\, \exists w\, \theta_M(k,w)$ with the Δ_0 formula $\theta_M(k,w) :\equiv (k = 0 \wedge w = 0) \vee \mathrm{Run}_M(k,w)$. By Lemma A.14, $g_M(k) :=$ least w with $\theta_M(k,w)$ is provably computable in PA. By (R2), $g_M(k) \geq \tau_{30}(k)$ for $k \geq 1$. By Theorem 2(a) applied to g_M, $g_M(i) < \tau_{30}(i)$ for all large i, a contradiction. □

Corollary 8.1 (Beeping Busy Beaver). Formalize "BBB$(n) \leq N$" following Aaronson ([Aa] §5.10): every n-state, 2-symbol machine that beeps only finitely often does not beep after step N. Then for every $n \geq 51$ and every $N \in \mathbb{N}$, PA $\nvdash$ "BBB$(n) \leq N$".

Proof. Let M_n be M_{51} with $n-51$ unreachable states added. PA proves, by induction on t, that a run code of M_n of length t exists iff one of M_{51} of length t exists, and that the two then describe the same configurations (the added states are never entered). This needs no assumption that either machine runs forever. Hence PA proves $\mathsf{QH}_{M_n} \leftrightarrow \mathsf{QH}$. Suppose PA $\vdash$ "BBB$(n) \leq N$". Then PA $\vdash \mathsf{QH}_{M_n} \rightarrow$ "M_n does not beep after step N". But "M_n beeps at some step $> N$" is a true Σ_1 sentence, by Theorem 3(1), so PA proves it (Σ_1 completeness of PA: every true Σ_1 sentence is provable). Hence PA $\vdash \neg\mathsf{QH}_{M_n}$, so PA $\vdash \neg\mathsf{QH}$, contradicting Theorem 3(3).

Aaronson counts a beep on *exiting* the beep state; §2.4 counts entries. This makes no difference: PA proves that a machine has finitely many exits from state 50 iff it has finitely many entries. Each exit is preceded by an entry, and each entry is followed by an exit unless the machine halts there, which happens at most once (and M_{51} never halts). □

Remark (halting versus quasihalting). Theorem 3 is an independence result for a Σ_2 property, quasihalting. It does **not** give a machine whose ordinary *halting* is independent of PA. Halting is Σ_1, and PA proves every true Σ_1 sentence, so an independent halting question must concern a machine that in fact runs forever, with a true but unprovable Π_1 sentence "it never halts". The Π_2 sentence Inf_R is not of that form.

---

9. Numerical checks (not part of any proof)

All scripts are in `checks51.tar.gz`. Those for §§3–5 are in its folder `u51_lemmaG/`. Those for §6 are in `lemmaG_work/` and are listed in §6.11. They were run with `nice` and memory limits. Outputs are saved next to them (`*.out`).

- `r51.py`. The arithmetic candidate formula agrees with `ref51.cand` on 3000 random cases. The recursive evaluation of Φ (composition law, last-block expansion) agrees with direct simulation of `ref51.step` on 214 random small cases. Also $\Phi_{(1)}(z) = 2z + 3$ for $z \le 5$, $\Psi(1) = 3$ and $\Psi(2) = 11$.
- `check_basic.py`, on worms with letters ≤ 5 and length ≤ 7 and candidates $n \le 24$. Lemma 4.4 (both ord and o decrease): 14600 checks. Lemma 4.2: 584. Lemma 3.1(b),(c): 14600 each. Lemma 3.4: 6840 lifted candidates. Lemma 5.1: 151. (M): 122. (D): 126. (I): 969. No failures.
- `mono.py`. Further monotonicity checks, including two that are **not** used or proved here: letterwise monotonicity (raising any letter does not decrease Φ; 291 checks) and prefix monotonicity (195 checks). No failures.
- `ordw.py`, a direct implementation of §4.1. It gives $\mathrm{ord}((k)) = \tau_{k+1}$ for $k \le 4$, $\mathrm{ord}((k-1)^{j+1}) = \tau_{k+1}[j]$ for $2 \le k \le 6$ and $j \le 3$, and $\tau_k[0] < \mathrm{ord}((k-2, v)) < \tau_k[1]$ for $1 \le v \le k-3$ and $k \le 8$. The last two illustrate remark (a) of §6.10.
- `bek_vs_bw.py`. It compares the Beklemishev expansion with the Buchholz–Wainer fundamental sequence: $\mathrm{ord}(Y[j]) \ge \mathrm{ord}(Y)[j]$ for $j \le 3$ on 1255 stripped worms (letters ≤ 5, length ≤ 7). That is 5020 checks, 1339 with equality, and no failure. This is not proved here and not used.
- Small values. $\Phi_{(1,1)}(z) = 11, 19, 35, 59, 91, 155$ for $z = 0, \dots, 5$, and $\Phi_{(1,1,1)}(0) = 91$, $\Phi_{(1,1,1)}(1) = 4091$. Compare $H_{\omega^2}(z) = 2^{z+1}(z+1) - 1 = 1, 7, 23, 63, 159, 383$. The naive bound $\Phi_Y(z) \ge H_{\mathrm{ord}(Y)}(z)$ fails at $Y = (1,1)$, $z = 3, 4, 5$, because of the width effect of §6.10(a). This is consistent with Theorem 6.19, which needs a time $t \ge 8\|\alpha\|^2 + 1000$ and divides the argument by $8\|\alpha\|$.
- `conj_c0.py`. `hardy.py` implements the capped Buchholz–Wainer H and F, sanity-checked against the definition. On all 198 pairs (Y, z) from a test family where both sides are below 10^{40}, the one-step comparison $\Phi_Y(z) \ge H_{\mathrm{ord}(Y)[0]}(z)$ holds, and so does the stronger $\Phi_Y(z) \ge H_{\mathrm{ord}(Y)[1]}(z)$. The fixed-argument bounds $\Phi_Y(z) \ge H_{\mathrm{ord}(Y)}(1)$ and $\Phi_Y(z) \ge H_{\mathrm{ord}(Y)}(2)$ fail (e.g. $Y = (1,1,1)$, $z = 0$: $91 < 2047$). None of these comparisons is used.
- `check_height1.py`. It checks exact iteration laws for the worms 1^M ($M \le 6$) from an earlier approach to §6. They are not used here.
- The checks of §6 (`lemmaG_work/`, details in §6.11):
  - normal form equals minimal size on all 16384 worms of size ≤ 14;
  - Lemmas 6.6, 6.7 and 6.11;
  - the greedy property, Proposition 6.12 (11359 checks);
  - the norm formulas and the fundamental-sequence norm bound, Lemma 6.5(v);
  - normal forms along 39358 steps of actual runs, including 63 restarts after beeps (Lemma 6.9);
  - the identity $E = 2\Phi$ of Theorem 6.14, verified exactly by exhaustive search on 26 states;
  - Corollary 6.16.

  No failures.
- Rocq: `rocqchk` was re-run (§2.3), and `Machine51.v` was regenerated byte-identically from `beeper51.txt`.

---

10. Gaps and assumptions

10.1 Dependencies

- **Appendix A**, all proved there except the cited classification theorem:
  - the standard ordinal facts (F1)–(F6);
  - the Buchholz–Wainer fundamental sequences (Definition A.1, Lemma A.2);
  - the Hardy and fast-growing functions, with Lemma A.6, Corollary A.7 and Lemma A.8;
  - the gap lemma A.9, the norm lemma A.10 and Corollary A.11;
  - the classification theorem A.13, which is Buchholz–Wainer 1987, Theorem 5 (cited, and checked

against the source in §A.8);
  - Lemma A.14, reading Buchholz–Wainer's axiom group (f) as the defining equations of each elementary *definition*.
- **Standard ordinal arithmetic and logic, in the metatheory.**
  - The identity $\omega^a + \omega^b = \omega^b$ for $a < b$ (§6.2).
  - Transfinite induction and recursion below ε_0 (§§4 and 6).
  - König's lemma (Definition 6.13).
- **The Rocq development** `rocq51.tar.gz` (machine-checked, `Axioms: <none>`, re-checked with `rocqchk`). Only the *truth* of `M51_run30` and `M51_run_fun` is used, with `Steps_run`. The trusted base is the Rocq kernel plus the definitions listed in §2.3:
  - `TM.v`: `step`, `Steps`, `run`, `bp`;
  - `Machine51.v`: `delta`, `isbeep`, generated from `beeper51.txt`;
  - `Proc.v`: `tz`, `cand`, `step`, `Sseq`, `nbeep_before`, `enc`, `Dconf`, `C31`;
  - `Main.v`: `nbeeps_run`, `C30`, `tail30`;
  - `Cycle.v`: `DconfJ`.

  These definitions were read, and the process was compared with `ref51.py` line by line.

10.2 Status

- **No open gap.** Theorem 1, Theorem 2 (both parts), Theorem 3 (all three parts) and Corollary 8.1 are proved. They depend on nothing beyond 10.1. The growth input is LEMMA G, proved as Corollary 6.22, by way of Theorem 6.20 and LEMMA G* (Corollary 6.21).
- The growth facts proved:
  - $\Psi(k) \geq 2^{k+1} - 1$ and $k_{j+1} \geq 2^{k_j+1}$ (Corollary 5.3(d));
  - the time bound $\tau_q(j+2) \geq 2k_{j+1} + 2$ (Corollary 2.3);
  - $\Phi_{(k)}(z) \geq \frac{1}{2} H_{\tau_{k+1}}(\lfloor (2z+k+1)/(8(k+1)) \rfloor)$ for $2z + k + 1 \geq 8(k+1)^2 + 1000$ (Theorem 6.20);
  - $\Psi(k) \geq H_{\tau_{\lfloor k/2 \rfloor}}(k)$ for large k (Corollary 6.22).

  The constants 8 and 1000 of Theorem 6.19 are not optimized. No upper bounds on Ψ are claimed.
- Statements in §9 marked "not used" are not proved here. These are letterwise and prefix monotonicity, the comparison $\mathrm{ord}(Y[j]) \geq \mathrm{ord}(Y)[j]$, and the one-step comparison $\Phi_Y(z) \geq H_{\mathrm{ord}(Y)[0]}(z)$. Nothing depends on them.

10.3 Steps given at the usual level of detail

- The arithmetization of §2.4 and the proof of Lemma 2.4 in PA are given in outline, as is usual.
  - Run_R and Run_M are Δ_0 formulas, with running counts and the auxiliary data of each step stored in the witness.
  - The proof of Lemma 2.4 uses agreement of run codes by induction, truncation by the β-function lemma, and the bounded least-number principle.

  Only the direction ($\Rightarrow$) of the quasihalting equivalence is used. It needs neither the existence of long runs nor non-halting inside PA. (Non-halting is in any case trivial: all 102 transitions are defined.)
- The ones-count Lemma 2.1 is argued from the quoted definition of `step` in `TM.v`, not machine-checked.
- The proofs of §6 are complete. They are ordinary mathematics in the metatheory: transfinite inductions below ε_0, and König's lemma for the finiteness of E. Their correctness is not formalized in PA, and it need not be.

10.4 Conventions

- A *beep* of M_{51} is an entry into state 50; the Rocq count `nbeeps_run` counts steps taken from state 50. For a run that starts in state 30 or 31 these count the same visits (§2.3).
- The start-30 run (the actual machine) and the start-31 run (used inside the Rocq proof) have the same beep counts at every displayed list `Sseq j` (`M51_run30`, `M51_run_fun`). Theorems 1 and 3 are stated for start 30. They hold for start 31 as well.
- ord is the ordinal of §4.1. Several worms share an ordinal (for example $\mathrm{ord}((2,3)) = \mathrm{ord}((3)) = \tau_4$), and Φ is not a function of the ordinal alone. On normal-form worms, $\Phi_X(z)$ is determined by the pair

$(o(X) + z,\ \mathrm{size}(X) + z)$ (Theorem 6.14), monotonically (Corollary 6.16).

10.5 Claims made during the work, checked

- The original specification of the process:
  - the composition law, the atomic case and the worm ordinal are proved (Corollary 4.6, Lemma 4.4);
  - "the battle from $[1, 0, k]$ has ordinal ω_k" is $\mathrm{ord}((k)) = \tau_{k+1}$ (Lemma 4.0(v)), and every step lowers it (Theorem 1);
  - the identities $\Phi_{(1)}(z) = 2z + 3$ and $\Phi_{(2)}(0) = 11$ hold.
- Working conjectures from the start of the work:
  - The composition law is proved.
  - The case-A/case-B candidate structure and the lifts are Lemmas 3.1, 3.3 and 3.4.
  - The failure of the naive bound at ω^2 is confirmed (§9).
  - Battles restart with counter 0. This is true at every step, not only at beeps; it is the reset discussed in §6.10.
  - "The right statement loses bounded tower levels." At a large counter no tower level is lost: $\Phi_{(k)}$ dominates $H_{\tau_{k+1}}$ at a linearly slowed argument (Theorem 6.20). From counter 0, Corollary 6.22 keeps at least $\lfloor k/2 \rfloor$ of the $k + 1$ levels. Whether levels are really lost at counter 0 is not decided here, and it does not matter: LEMMA G* allows any loss.
  - *Reduction to Beklemishev's worm principle or to Kirby–Paris.* The expansions $Y[j]$ of Beklemishev's worm calculus are among the candidates (Lemma 3.1(e)). But R takes candidate $2z + 1$, which usually lies between two expansions, and its counter is reset. Beklemishev's unprovability proof of the worm principle (via reflection) is about the battle whose copy parameter is the step number, and it gives no growth bound for R. No reduction was found in either direction. §6 uses a different mechanism: on normal forms, R is a greedy descent under a size budget (Proposition 6.12, Theorem 6.14).
- Racheline's informal analysis is not used (§6.10).

10.6 Flaws

No flaw was found in the process or the machine. The specification's claims hold as stated.

Appendix A. Background: ordinals, Hardy functions, and the classification theorem

This appendix collects the standard material that §§4–8 use. Everything is proved here except the classification theorem (Theorem A.13), which is quoted from Buchholz–Wainer 1987 ([BW], Theorem 5); §A.8 records what that source and the corroborating source [Ca] actually say. The labels (F1)–(F6) and the numbers "A.n" refer to this appendix.

A.1 Ordinals below ε_0: notation and standard facts

All ordinals below are $< \varepsilon_0$. The facts in this section are standard (ordinal arithmetic, as in any set theory or proof theory text) and are used without proof except where a short proof is given.

(F1) Cantor normal form. Every $\alpha < \varepsilon_0$ has a unique representation

$$\alpha = \omega^{\alpha_1} + \omega^{\alpha_2} + \cdots + \omega^{\alpha_r}, \qquad \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r, \quad r \geq 0,$$

with $r = 0$ iff $\alpha = 0$. Each exponent satisfies $\alpha_i < \alpha$, because $\alpha_1 < \omega^{\alpha_1}$ below ε_0. So definitions and proofs may recurse on exponents. We call ω^{α_r} the **last term** and α_r the **last exponent**. The ordinal α is a **successor** if $r \geq 1$ and $\alpha_r = 0$, and a **limit** if $r \geq 1$ and $\alpha_r > 0$.

(F2) Comparison. $\alpha < \beta$ iff the exponent sequence of α is lexicographically smaller than that of β: exponents are compared recursively, and a proper initial segment counts as smaller. Also $x \mapsto \omega^x$ is strictly increasing, and $\omega^\gamma \cdot k < \omega^{\gamma+1}$ for every finite k.

(F3) Normal sums. Call $\rho + \sigma$ **normal** if $\rho = 0$, or $\sigma = 0$, or the last exponent of ρ is $\geq$ the first exponent of σ. The CNF of a normal sum is the concatenation of the two CNFs.

(F4) Natural sum. $\alpha \oplus \beta$ is the ordinal whose CNF exponents are the multiset union of those of α and β, sorted in non-increasing order. It is commutative and associative, 0 is neutral, and it is **strictly increasing in each argument**.

Proof of the monotonicity. For an ordinal γ and an exponent value v, let $\mu_\gamma(v)$ be the number of CNF terms of γ equal to ω^v. By (F2), $\beta < \beta'$ iff, at the largest v with $\mu_\beta(v) \neq \mu_{\beta'}(v)$, we have $\mu_\beta(v) < \mu_{\beta'}(v)$. Since $\mu_{\alpha\oplus\beta} = \mu_\alpha + \mu_\beta$, the multiplicities of $\alpha \oplus \beta$ and $\alpha \oplus \beta'$ differ at exactly the same values v and in the same direction. $\square$

If $\alpha_1 \geq \cdots \geq \alpha_m$, then $\omega^{\alpha_1} \oplus \cdots \oplus \omega^{\alpha_m} = \omega^{\alpha_1} + \cdots + \omega^{\alpha_m}$, the ordinary sum.

(F5) Norm. Group equal exponents: $\alpha = \omega^{\beta_1}k_1 + \cdots + \omega^{\beta_s}k_s$ with $\beta_1 > \cdots > \beta_s$ and $k_i \geq 1$. Put $N(0) = 0$ and $N(\alpha) = \max(k_1, \ldots, k_s, N(\beta_1), \ldots, N(\beta_s))$, "the largest integer mentioned in the Cantor normal form" (Caicedo 2007, Def. 2.3). Then $N(\beta + 1) \leq N(\beta) + 1$, and every exponent and every CNF tail of α has norm $\leq N(\alpha)$.

(F6) Towers and cofinality. Put $\tau_0 = 0$ and $\tau_{s+1} = \omega^{\tau_s}$, so $\tau_1 = 1$, $\tau_2 = \omega$, $\tau_3 = \omega^\omega$, and so on. Then $\tau_s < \tau_{s+1}$ and $\varepsilon_0 = \sup_s \tau_s$. Define the height $\mathrm{ht}(0) = 0$ and $\mathrm{ht}(\alpha) = 1 + \mathrm{ht}(\alpha_1)$, where α_1 is the leading exponent. Then $\alpha < \tau_{\mathrm{ht}(\alpha)+1}$.

Proof. By induction on height. The case $0 < \tau_1$ is clear. For $\alpha > 0$, $\mathrm{ht}(\alpha_1) = \mathrm{ht}(\alpha) - 1$, so $\alpha_1 < \tau_{\mathrm{ht}(\alpha)}$ by induction. Hence $\alpha < \omega^{\alpha_1+1} \leq \omega^{\tau_{\mathrm{ht}(\alpha)}} = \tau_{\mathrm{ht}(\alpha)+1}$. $\square$

So **every $\beta < \varepsilon_0$ lies below τ_s for all sufficiently large s.**

A.2 Fundamental sequences, in the Buchholz–Wainer convention

Definition A.1. For $\alpha < \varepsilon_0$ and $n \in \mathbb{N}$, define $\alpha[n]$ from the CNF of α. Write $\alpha = \rho + \omega^\delta$, where ω^δ is the last CNF term and ρ the sum of the earlier terms.

$$\alpha[n] = \begin{cases} 0 & \alpha = 0, \\ \rho & \delta = 0 \ \ (\alpha \text{ a successor}), \\ \rho + \omega^\gamma \cdot (n+1) & \delta = \gamma + 1, \\ \rho + \omega^{\delta[n]} & \delta \text{ a limit}. \end{cases}$$

This is the convention of Buchholz–Wainer (1987, p. 180). They write $\alpha = \omega^{\beta_k} + \cdots + \omega^{\beta_1} + \omega^{\beta_0}$ and, for limit α, set $\alpha_n = \omega^{\beta_k} + \cdots + \omega^{\beta_1} + \omega^\gamma(n+1)$ if $\beta_0 = \gamma + 1$, and $\alpha_n = \omega^{\beta_k} + \cdots + \omega^{\beta_1} + \omega^{(\beta_0)_n}$ if β_0 is a limit. Their example (p. 181): *"ω is assigned the fundamental sequence $1 < 2 < 3 < \ldots$"*, so $\omega[n] = n + 1$. BW define α_n only for limits. Extending it by $(\rho + 1)[n] = \rho$ and $0[n] = 0$ is how BW's descent relation (Definition 1, "$\gamma_{i+1} = \gamma_i - 1$ if γ_i is a successor") and their Hardy recursion treat successors. **This is the only convention used in this document.**

Lemma A.2. Let $\alpha, \sigma, \mu < \varepsilon_0$ and $n \in \mathbb{N}$.

- (a) If $\alpha > 0$ then $\alpha[n] < \alpha$.
- (b) If $\sigma > 0$ and $\sigma[n] > 0$, the leading exponent of $\sigma[n]$ is at most the leading exponent of σ.
- (c) If $\mu + \sigma$ is normal (F3) and $\sigma > 0$, then $(\mu + \sigma)[n] = \mu + \sigma[n]$, and $\mu + \sigma[n]$ is again normal.

Proof. (a) Induction on α. For a successor, $\rho < \rho + 1$. If $\delta = \gamma + 1$, then $\omega^\gamma \cdot (n+1) < \omega^{\gamma+1}$ (F2), and left addition by ρ is strictly increasing. If δ is a limit, then $\delta < \omega^\delta \leq \alpha$, so the induction hypothesis gives $\delta[n] < \delta$, hence $\omega^{\delta[n]} < \omega^\delta$. (b) If σ has at least two CNF terms, $\sigma[n]$ changes only the last one, so the leading term is kept. If $\sigma = \omega^\delta$ is a single term, $\sigma[n]$ is 0, or $\omega^\gamma \cdot (n+1)$ with $\gamma < \delta$, or $\omega^{\delta[n]}$ with $\delta[n] < \delta$ by (a). (c) Write $\sigma = \sigma' + \omega^\delta$ (last term). By (F3) the CNF of $\mu + \sigma$ is that of μ followed by that of σ. So its last term is ω^δ and the rest is $\mu + \sigma'$. Apply the definition to $\mu + \sigma$ and to σ, in the same one of the three nonzero cases: the result is $\mu + \sigma' + (\text{same replacement}) = \mu + \sigma[n]$. Normality of $\mu + \sigma[n]$ follows from (b). $\square$

A.3 Hardy and fast-growing functions (Buchholz–Wainer 1987, p. 181)

$$H_0(x) = x, \qquad H_{\alpha+1}(x) = H_\alpha(x+1), \qquad H_\lambda(x) = H_{\lambda[x]}(x) \ \ (\lambda \text{ a limit}).$$

$$F_0(x) = x+1, \qquad F_{\alpha+1}(x) = F_\alpha^{x+1}(x), \qquad F_\lambda(x) = F_{\lambda[x]}(x) \ \ (\lambda \text{ a limit}).$$

Here F_α^{x+1} is the $(x+1)$-fold iterate. These are verbatim BW's definitions with their fundamental sequences $\lambda_n = \lambda[n]$. They are well defined by transfinite recursion, since $\lambda[x] < \lambda$ (Lemma A.2(a)). Examples: $H_k(x) = x + k$ for finite k, $H_\omega(x) = H_{x+1}(x) = 2x+1 = F_1(x)$, and $F_2(x) = 2^{x+1}(x+1) - 1$.

Reachability. For $y \in \mathbb{N}$ write $\alpha \Rightarrow_y \beta$ if there is a chain $\alpha = \gamma_0, \gamma_1, \ldots, \gamma_r = \beta$ ($r \geq 0$) with $\gamma_i > 0$ and $\gamma_{i+1} = \gamma_i[y]$ for all $i < r$. The chain strictly decreases (Lemma A.2(a)), and $\Rightarrow_y$ is reflexive and transitive. For $\beta < \alpha$ this is BW's relation $\beta <_y \alpha$ (BW Definition 1: $\gamma_{i+1} = \gamma_i - 1$ at successors and $(\gamma_i)_y$ at limits). Note that $\alpha \Rightarrow_y 0$ for every α and y.

A.4 Two ordinal lemmas

Lemma A.3 (passing through a prefix). If $\mu + \sigma$ is a normal sum, then $\mu + \sigma \Rightarrow_q \mu$ for every q.

Proof. Let $\sigma_0 = \sigma$ and $\sigma_{i+1} = \sigma_i[q]$ while $\sigma_i > 0$. This sequence strictly decreases, so some $\sigma_r = 0$. By Lemma A.2(c), applied repeatedly (each $\mu + \sigma_i$ is again normal), the chain $\mu + \sigma_0,\ \mu + \sigma_1, \ldots, \mu + \sigma_r = \mu$ is a $\Rightarrow_q$-chain. □

Lemma A.4 (exponent lifting). If $\rho + \omega^\beta$ is normal and $\beta \Rightarrow_q \beta'$, then $\rho + \omega^\beta \Rightarrow_q \rho + \omega^{\beta'}$.

Proof. It suffices to treat one step, $\beta' = \beta[q]$ with $\beta > 0$. Every ordinal in the chain is $\leq \beta$, so each $\rho + \omega^{\beta_i}$ stays normal; then compose the steps. If β is a limit, Definition A.1 gives $(\rho + \omega^\beta)[q] = \rho + \omega^{\beta[q]}$ in one step. If $\beta = \gamma + 1$, then $\beta[q] = \gamma$, and the first step gives $(\rho + \omega^{\gamma+1})[q] = \rho + \omega^\gamma \cdot (q+1)$. For $j = q, q-1, \ldots, 1$ the sum $\rho + \omega^\gamma \cdot (j+1) = (\rho + \omega^\gamma \cdot j) + \omega^\gamma$ is normal, so Lemma A.3 gives $\rho + \omega^\gamma \cdot (j+1) \Rightarrow_q \rho + \omega^\gamma \cdot j$. Chaining these ends at $\rho + \omega^\gamma = \rho + \omega^{\beta[q]}$. □

Lemma A.5 (consecutive terms of a fundamental sequence). For every limit λ and all $n, q \in \mathbb{N}$: $\lambda[n+1] \Rightarrow_q \lambda[n]$.

Proof. Induction on λ. Write $\lambda = \rho + \omega^\delta$ with $\delta > 0$ (last term).

- $\delta = \gamma + 1$. Then $\lambda[n+1] = \rho + \omega^\gamma \cdot (n+2) = \big(\rho + \omega^\gamma \cdot (n+1)\big) + \omega^\gamma$, a normal sum. By Lemma A.3 it reaches $\rho + \omega^\gamma \cdot (n+1) = \lambda[n]$.
- δ a limit. Then $\lambda[n+1] = \rho + \omega^{\delta[n+1]}$ and $\lambda[n] = \rho + \omega^{\delta[n]}$, and both sums are normal, since $\delta[n], \delta[n+1] < \delta$. As $\delta < \lambda$, the induction hypothesis gives $\delta[n+1] \Rightarrow_q \delta[n]$; apply Lemma A.4. □

A.5 Monotonicity of Hardy functions

Lemma A.6 (strict monotonicity). For every $\alpha < \varepsilon_0$ and every x: $H_\alpha(x) < H_\alpha(x+1)$. (This is BW Theorem 1(i), *"H_α is strictly increasing"*; we prove it here in the convention used.)

Proof. Induction on α. Assume the claim for all $\gamma < \alpha$. We first record:

($*$) If $\gamma_0, \gamma_1, \ldots, \gamma_r$ is a $\Rightarrow_y$-chain with $\gamma_0 < \alpha$, then $H_{\gamma_r}(y) \leq H_{\gamma_0}(y)$.

Each step satisfies $H_{\gamma_{i+1}}(y) \leq H_{\gamma_i}(y)$. If $\gamma_i = \delta + 1$ is a successor, then $\gamma_{i+1} = \delta < \alpha$, so $H_{\gamma_{i+1}}(y) = H_\delta(y) < H_\delta(y+1) = H_{\gamma_i}(y)$ by the induction hypothesis. If γ_i is a limit, then $H_{\gamma_{i+1}}(y) = H_{\gamma_i[y]}(y) = H_{\gamma_i}(y)$ by definition.

Now the three cases. For $\alpha = 0$: $x < x+1$. For $\alpha = \beta + 1$: $H_\alpha(x) = H_\beta(x+1) < H_\beta(x+2) = H_\alpha(x+1)$. For a limit $\alpha = \lambda$:

$$H_\lambda(x) = H_{\lambda[x]}(x) < H_{\lambda[x]}(x+1) \leq H_{\lambda[x+1]}(x+1) = H_\lambda(x+1).$$

The strict inequality is the induction hypothesis for $\lambda[x] < \lambda$. The weak inequality is ($*$) with $y = x+1$, applied to the chain $\lambda[x+1] \Rightarrow_{x+1} \lambda[x]$ from Lemma A.5; it starts at $\lambda[x+1] < \lambda$. □

Corollary A.7 (Hardy functions decrease along descents). If $\alpha \Rightarrow_y \beta$, then $H_\beta(y) \le H_\alpha(y)$.

Proof. The argument for $(*)$, now using Lemma A.6 for all ordinals. □

A.6 Relation to the fast-growing hierarchy

Lemma A.8. (a) If $\alpha + \gamma$ is a normal sum, then $H_{\alpha+\gamma}(x) = H_\alpha(H_\gamma(x))$. (b) $H_{\omega^\alpha} = F_\alpha$ for every $\alpha < \varepsilon_0$.

BW state both on p. 181: *"H_{α+γ} = H_α ∘ H_γ"*, for $\gamma < \omega^{\beta_0+1}$ where β_0 is the last exponent of α (exactly normality), *"and hence for every α < ε₀, H_{ω^α} = F_α"*.

Proof. (a) Induction on γ. The case $\gamma = 0$ is trivial. Successor, $\gamma = \delta + 1$: the sum $\alpha + \delta$ is again normal, since δ and $\delta + 1$ have the same leading exponent when $\delta > 0$. So $H_{\alpha+\delta+1}(x) = H_{\alpha+\delta}(x+1) = H_\alpha(H_\delta(x+1)) = H_\alpha(H_{\delta+1}(x))$. Limit: by Lemma A.2(c), $(\alpha+\gamma)[x] = \alpha + \gamma[x]$, again normal, so $H_{\alpha+\gamma}(x) = H_{\alpha+\gamma[x]}(x) = H_\alpha(H_{\gamma[x]}(x)) = H_\alpha(H_\gamma(x))$. (b) Induction on α. $H_{\omega^0}(x) = H_1(x) = x + 1 = F_0(x)$. For $\alpha = \beta + 1$: $\omega^{\beta+1}[x] = \omega^\beta \cdot (x+1)$, which is a normal sum of $x+1$ copies of ω^β, so by (a) $H_{\omega^{\beta+1}}(x) = H_{\omega^\beta}^{x+1}(x) = F_\beta^{x+1}(x) = F_{\beta+1}(x)$. For a limit α: $\omega^\alpha[x] = \omega^{\alpha[x]}$, so $H_{\omega^\alpha}(x) = H_{\omega^{\alpha[x]}}(x) = F_{\alpha[x]}(x) = F_\alpha(x)$. □

A.7 The norm lemma, the classification theorem, and the bridge from first-order PA

Recall $N(\cdot)$ from (F5).

Lemma A.9 (gap lemma). If λ is a limit, $\beta < \lambda$ and $N(\beta) \le n$, then $\beta < \lambda[n]$.

Proof. Induction on λ. Write $\lambda = \rho + \omega^\delta$, $\delta > 0$ (last term), so $\lambda[n] = \rho + (\omega^\delta)[n]$ by Lemma A.2(c), with $(\omega^\delta)[n]$ defined by Definition A.1 for the single term ω^δ. If $\beta < \rho$, then $\beta < \rho \le \lambda[n]$. Otherwise $\rho \le \beta < \rho + \omega^\delta$. Compare CNFs (F2): the CNF of β must begin with the CNF of ρ. At a first difference inside ρ's terms, β would be either $< \rho$ or $> \lambda$. So $\beta = \rho + \beta'$, where β' is the remaining CNF tail. Then $\beta' < \omega^\delta$, so every exponent of β' is $< \delta$, and $N(\beta') \le N(\beta) \le n$. Left addition is strictly increasing, so it suffices to show $\beta' < (\omega^\delta)[n]$.

- $\delta = \gamma + 1$: $(\omega^\delta)[n] = \omega^\gamma \cdot (n+1)$. All exponents of β' are $\le \gamma$, and the coefficient of ω^γ in β' is some $j \le N(\beta') \le n$ (possibly $j = 0$). So $\beta' = \omega^\gamma \cdot j + \beta''$ with $\beta'' < \omega^\gamma$, and $\beta' < \omega^\gamma \cdot (j+1) \le \omega^\gamma \cdot (n+1)$.
- δ a limit: $(\omega^\delta)[n] = \omega^{\delta[n]}$. If $\beta' = 0$ we are done. Otherwise its leading exponent β_1' satisfies $\beta_1' < \delta$ and $N(\beta_1') \le n$. The induction hypothesis for the limit $\delta < \lambda$ gives $\beta_1' < \delta[n]$. Hence $\beta' < \omega^{\beta_1'+1} \le \omega^{\delta[n]}$. □

(This is where BW's convention $\omega[n] = n + 1$ matters: $\beta = n$ has $N(\beta) = n$ and $n < \omega[n]$. In the convention $\{\omega\}(n) = n$ the inequality would only be $\le$.)

Lemma A.10 (norm lemma). If $\beta < \alpha < \varepsilon_0$ and $N(\beta) \le n$, then $\alpha \Rightarrow_n \beta$. **The threshold $N(\beta)$ does not depend on α.**

Proof. Induction on α. If $\alpha = \gamma + 1$, then $\alpha[n] = \gamma \ge \beta$. Either $\gamma = \beta$ (one step), or $\beta < \gamma$ and the induction hypothesis gives $\gamma \Rightarrow_n \beta$. If α is a limit, Lemma A.9 gives $\beta < \alpha[n] < \alpha$; the induction hypothesis gives $\alpha[n] \Rightarrow_n \beta$, so $\alpha \Rightarrow_n \beta$. ($\alpha = 0$ cannot occur.) □

Corollary A.11. If $\beta < \alpha$ and $y \ge N(\beta)$, then $H_\beta(y) \le H_\alpha(y)$.

Proof. Lemma A.10 with $n = y$, then Corollary A.7. □

Ketonen–Solovay, Kirby–Paris and Caicedo use similar "norm"/"$\to_n$" lemmas (Caicedo Lemma 2.7 and §3; BW Lemma 1, second part). BW's version gives, for each fixed λ, *some* threshold k with $\beta <_n \lambda$ for all $n \ge k$, and that threshold may depend on λ. In §6 the ordinals being compared change from battle to battle, so we need the uniform threshold $N(\beta)$ of Lemma A.10.

Definition A.12 (BW 1987, Definition 3). A function $F : \mathbb{N}^k \to \mathbb{N}$ is **provably computable in PA** if there are elementary functions V, T such that (i) for all arguments *"F(n₁,…,n_k) = V(least m such that T(n₁,…,n_k,m) = 0)"*, and (ii) the sentence $\forall x_1 \cdots \forall x_k \, \exists y \, (T(x_1, \ldots, x_k, y) = 0)$ is a theorem of PA. (The elementary functions are the Kalmár-elementary ones, BW Definition 2.)

Theorem A.13 (Wainer 1970; proof in Buchholz–Wainer 1987, Theorem 5). *"Every function provably computable in PA is elementary in some F_α for α < ε₀, and is therefore majorised by F_{α+1}."*

We use only the weakest reading of "majorised": there is x_0 with $F(x) \le F_{\alpha+1}(x)$ for all $x \ge x_0$. By Lemma A.8(b), $F_{\alpha+1} = H_{\omega^{\alpha+1}}$. So **every function provably computable in PA is eventually bounded by some H_β with $\beta < \varepsilon_0$.**

Independent confirmation in another convention. Caicedo (2007, Theorem 1.9, credited to Wainer 1970) states: *"any recursive f provably total in PA is eventually dominated by some f_α, α < ε₀"*. Caicedo's f_α satisfies $H_{\omega^\alpha}(n) = f_\alpha(n+1) - 1$ (his Theorem 1.14(1)). His H_α coincides with BW's: his Def. 1.13 sets $H_\lambda(n) = H_{d(\lambda,n+1)}(n)$, where d is the Kirby–Paris operation, which at argument $n+1$ replaces a last term $\omega^{\gamma+1}$ by $n+1$ copies of ω^γ ([KP] p. 289); so $d(\lambda, n+1) = \lambda[n]$, by induction on λ through the cases of Definition A.1. This confirmation is not used in any proof. Hence $f_\alpha(m) = H_{\omega^\alpha}(m-1) + 1 \le H_{\omega^\alpha}(m)$ for $m \ge 1$ by Lemma A.6, and the conclusion is the same.

On the formal system. BW formalize PA in a Tait-style calculus whose language has a function symbol for every elementary function (prime formulas $f(t_1, \dots, t_n) = t_{n+1}$). Lemma A.14 shows that a Π_2 theorem of ordinary first-order PA (language $0, S, +, \times$) yields a function that is provably computable in BW's sense, and that is all the proofs of Theorems 2 and 3 need.

Lemma A.14 (from first-order PA to Buchholz–Wainer's formalization). Let $\theta(x, y)$ be a Δ_0 formula of the language $\{0, S, +, \times\}$, and let $\chi_\theta(x, y) = 0$ if $\theta(x, y)$ holds and $\chi_\theta(x, y) = 1$ otherwise. If PA $\vdash \forall x \exists y\, \theta(x, y)$, then for every elementary function V, the function $x \mapsto V(\text{least } y \text{ with } \theta(x, y))$ is provably computable in PA in the sense of Definition A.12, with $T = \chi_\theta$.

Proof. Write $\mathrm{PA}_{\mathrm{BW}}$ for BW's formal system ([BW] §2). It is a Tait-style calculus with cut. Its prime formulas are $f(t_1, \dots, t_n) = t_{n+1}$ and their negations, for every elementary f, with terms built from variables, 0 and S. Its axioms are the equality axioms (=), the successor axioms (S), the axioms (f) (*"defining equations for each elementary function f"*), and induction axioms for every formula.

Reading of (f). BW do not list the equations in (f) explicitly. We read (f) the natural way: there is one function symbol for each elementary *definition* in the sense of BW Definition 2 (a term built from the initial functions by composition, bounded sums and bounded products), and (f) contains the recursion equations of that definition. In particular the symbol for χ_θ below comes with the recursion equations of the bounded sums and products from which it is built. Under a purely extensional reading (one symbol per function, with some unspecified defining equations) step 3 could fail, since two definitions of the same function may agree only by a true Π_1 fact that PA does not prove. We adopt the natural reading. It is harmless for Theorem 5, because in BW's proof the (f)-axioms enter only through their numerical instances: in the proof of their Embedding Lemma 2, part (ii), BW write *"The result follows immediately when Γ is an elementary axiom because, on substitution of numerals for the free variables, we obtain a true cepf"* ([BW] p. 188; "cepf" is BW's abbreviation for *closed elementary prime formula*, p. 187). So any true defining equations may be used.

1. *χ_θ is elementary.* This goes by induction on θ, writing $\dot{-}$ for modified subtraction:
   - an atomic formula $t_1 = t_2$ (the t_i are polynomials, hence elementary) has $\chi = 1 \dot{-} \big(1 \dot{-} ((t_1 \dot{-} t_2) + (t_2 \dot{-} t_1))\big)$;
   - $\chi_{\neg\varphi} = 1 \dot{-} \chi_\varphi$, $\chi_{\varphi\wedge\psi} = 1 \dot{-} (1 \dot{-} (\chi_\varphi + \chi_\psi))$, and $\chi_{\varphi\vee\psi} = \chi_\varphi \cdot \chi_\psi$;
   - $\chi_{\exists z \le t\, \varphi} = \prod_{z \le t} \chi_\varphi$ and $\chi_{\forall z \le t\, \varphi} = 1 \dot{-} \big(1 \dot{-} \sum_{z \le t} \chi_\varphi\big)$.

   Only addition, modified subtraction, bounded sums and products, and composition are used. These are the closure operations of BW Definition 2.
2. *$\mathrm{PA}_{\mathrm{BW}}$ proves every theorem of PA, after unnesting.* Both + and × are elementary. For × take the definition $x \cdot y = \sum_{i<y} x$, whose defining equations are $x \cdot 0 = 0$ and $x \cdot Sy = x \cdot y + x$. A formula of PA becomes a formula of $\mathrm{PA}_{\mathrm{BW}}$ once nested terms are eliminated in the usual way; for example, $(x + y) \cdot z = w$ becomes $\exists u\, (x + y = u \wedge u \cdot z = w)$. Under this translation every axiom of PA is a theorem of $\mathrm{PA}_{\mathrm{BW}}$:
   - the successor axioms are the axioms (S);
   - the recursion equations of + and × are defining equations (f) (BW, p. 183, list those for addition, including functionality);

- each induction axiom of PA becomes an instance of BW's induction scheme 3), which covers every formula.

The logical part is covered because the Tait calculus with cut derives every valid first-order sequent (completeness of the Tait calculus; BW base their treatment on Tait 1968), and modus ponens is simulated by cut. Hence the translation of every theorem of PA is a theorem of $\mathrm{PA}_{\mathrm{BW}}$.

3. $PA_{\mathrm{BW}} \vdash \forall x\,\forall y\,\big(\theta^{\circ}(x,y) \leftrightarrow \chi_{\theta}(x,y) = 0\big)$, *where* θ° *is the unnested* θ. The proof is by induction on the construction of θ. The atomic and propositional cases follow from the defining equations of $+$, $\cdot$ and $\dot{-}$ by inductions inside $\mathrm{PA}_{\mathrm{BW}}$. For a bounded existential quantifier, prove $\forall b\,\big((\exists z \le b\,\varphi) \leftrightarrow \prod_{z\le b}\chi_{\varphi} = 0\big)$ by induction on b, using two facts:
   - the defining equations $\prod_{z\le 0}\chi_{\varphi} = \chi_{\varphi}(0)$ and $\prod_{z\le b+1}\chi_{\varphi} = \big(\prod_{z\le b}\chi_{\varphi}\big)\cdot\chi_{\varphi}(b+1)$;
   - $u\cdot v = 0 \leftrightarrow (u = 0 \vee v = 0)$, which is provable by induction.

   The bounded universal case is the same, with sums.

Now suppose $\mathrm{PA} \vdash \forall x\,\exists y\,\theta$. Step 2 gives $\mathrm{PA}_{\mathrm{BW}} \vdash \forall x\,\exists y\,\theta^{\circ}$. Step 3 then gives $\mathrm{PA}_{\mathrm{BW}} \vdash \forall x\,\exists y\,(\chi_{\theta}(x,y) = 0)$, which is condition (ii) of Definition A.12 with $T = \chi_{\theta}$. Condition (i) holds by the definition of the function. □

Steps 2 and 3 are routine formalization. They are given at the usual textbook level of detail, not symbol by symbol (§10.3).

A.8 Sources: what they say, and how their conventions match

Each source below was opened and checked, as stated. Quotations are short excerpts.

[BW] W. Buchholz and S. Wainer, "Provably computable functions and the fast growing hierarchy", in: S. G. Simpson (ed.), *Logic and Combinatorics*, Contemporary Mathematics 65, AMS, 1987, pp. 179–198. Read from a scan on LMU's open-access server (epub.ub.uni-muenchen.de/3843); pp. 179–182 and 193–197 were checked against the page images.

- Fundamental sequences, p. 180: for $\alpha = \omega^{\beta_k} + \cdots + \omega^{\beta_0}$ a limit, α_n replaces the last term by $\omega^{\gamma}(n+1)$ if $\beta_0 = \gamma + 1$, and by $\omega^{(\beta_0)_n}$ if β_0 is a limit. Example, p. 181: *"ω is assigned the fundamental sequence $1 < 2 < 3 < \ldots$"*. This is Definition A.1.
- Hierarchies, p. 181: $F_0(n) = n+1$, $F_{\alpha+1}(n) = F_{\alpha}^{n+1}(n)$, $F_{\alpha}(n) = F_{\alpha_n}(n)$; $H_0(n) = n$, *"H_{α+1}(n) = H_α(n + 1)"*, $H_{\alpha}(n) = H_{\alpha_n}(n)$ for limits. Also *"H_{α+γ} = H_α ∘ H_γ"* (for normal sums) and *"H_{ω^α} = F_α"*. These are the definitions of §A.3 and Lemma A.8.
- Theorem 1(i), p. 182: *"H_α is strictly increasing"*. Lemma A.6 reproves it.
- Definition 3, pp. 183–184 (provably computable): $F(\vec{n}) = V(\text{least } m \text{ with } T(\vec{n},m) = 0)$ with V, T elementary and $\forall\vec{x}\,\exists y\,(T(\vec{x},y) = 0)$ a theorem of PA. This is Definition A.12.
- Theorem 5, p. 193: *"Every function provably computable in PA is elementary in some F_α"* (for $\alpha < \varepsilon_0$), *"and is therefore majorised by F_{α+1}"*. This is Theorem A.13, the key external input.
- Theorem 6 and its proof, pp. 196–197: Cichon's proof of Goodstein independence. It argues exactly as §7 does: the Goodstein function *"would be provably computable in PA"*, which contradicts Theorem 5.
- History, p. 179: result I *"goes back to Kreisel [1952]"*. BW credit Schwichtenberg 1971 and Wainer 1970/72 with the characterization by the fast-growing hierarchy below ε_0.
- **Convention match.** Exact. [BW] is the only convention used in this document.

[Ca] A. E. Caicedo, "Goodstein's function", Revista Colombiana de Matemáticas 41(2) (2007) 381–391. Read from the journal's PDF (revistas.unal.edu.co).

- Theorem 1.9, credited to Wainer [16] = Wainer 1970: *"any recursive f provably total in PA is eventually dominated by some f_α, α < ε₀"*.
- Definition 1.13: $H_{\alpha}(n) = H_{d(\alpha,n+1)}(n)$ for limit α, where d is the KP/KS fundamental sequence. So his H_{α} equals the [BW] H_{α}. Theorem 1.14(1): $H_{\omega^{\alpha}}(n) = f_{\alpha}(n+1) - 1$.
- Definition 2.3: $N(\alpha)$ is *"the largest integer mentioned in the Cantor normal form"*. This is (F5).
- **Use.** An independent confirmation of Theorem A.13 in a second convention (§A.7), and the source of the norm N.

[Aa] S. Aaronson, "The Busy Beaver Frontier", ECCC TR20-115 (2020), §5.10. Read from the PDF on the author's site.

- A beeping machine *"emits a 'beep' as it exits a certain designated state"*. Deciding whether M beeps finitely often *"is complete for $\Sigma 2$"*. The bbchallenge wiki (BBB page): *"A TM quasihalts if it beeps only a finite number of times."*
- **Use.** Definitions in §§1–2, and Corollary 8.1.

Appendix B. The machine M_{51}

51 states, 2 symbols (0 is blank), start state 30, beep state 50. Each entry is the symbol written, the head move (L or R) and the next state. This is Racheline's 51-state table with one change: state 50 reading 1, a halt in the original, is `0L48`. Every one of the 102 entries is defined, so the machine never halts. (File `beeper51.txt`; the Rocq file `Machine51.v` is generated from it, §2.2.)

state	read 0	read 1
0	1L0	0L1
1	1L0	1R2
2	1L2	0L4
3	0L10	1L4
4	0L3	0L5
5	1L4	1R6
6	1L7	0L15
7	0L8	1L7
8	0L9	1L7
9	1R11	1L9
10	0L20	0L17
11	0R12	1R11
12	0R13	1R13
13	1L14	1R12
14	0L6	1L10
15	0L16	1L16
16	1L20	1L15
17	0L18	1L18
18	1R19	1L17
19	1L19	1L20
20	0L50	0L21
21	0R43	0L9
22	0R22	1R23
23	1R26	0L24
24	1L25	0L24
25	1R25	1R22
26	0R50	0R27
27	1R26	1L35
28	0R29	0R30
29	0L4	1R28
30	1L31	1L0
31	0L32	1L33
32	0L33	1L31
33	0L34	0L35
34	0L37	1L33
35	1R36	1R38
36	0R37	1R28
37	0L29	1R36
38	1R38	1R39

state	read 0	read 1
39	0R38	1R40
40	0R41	1L42
41	1L49	1R40
42	1R36	0L42
43	0R44	1R44
44	1L45	1R43
45	0L46	1L46
46	0R22	1L45
47	1L41	1L47
48	0L47	0L49
49	1L22	1L48
50	0L48	0L48

References

- [Aa] S. Aaronson, *The Busy Beaver Frontier*, ECCC Technical Report TR20-115, 2020. https://www.scottaaronson.com/papers/bb.pdf
- [BW] W. Buchholz, S. Wainer, *Provably computable functions and the fast growing hierarchy*, in: S. G. Simpson (ed.), Logic and Combinatorics, Contemp. Math. 65, AMS, 1987, 179–198. https://epub.ub.uni-muenchen.de/3843/1/3843.pdf
- [Ca] A. E. Caicedo, *Goodstein's function*, Rev. Colombiana Mat. 41(2) (2007) 381–391. https://revistas.unal.edu.co/index.php/recolma/article/view/94926
- [Ci] E. A. Cichon, *A short proof of two recently discovered independence results using recursion theoretic methods*, Proc. AMS 87 (1983) 704–706. https://doi.org/10.1090/S0002-9939-1983-0687646-0
- [KP] L. Kirby, J. Paris, *Accessible independence results for Peano arithmetic*, Bull. London Math. Soc. 14 (1982) 285–293. https://www.cs.tau.ac.il/~nachumd/term/Kirbyparis.pdf
- [KS] J. Ketonen, R. Solovay, *Rapidly growing Ramsey functions*, Ann. of Math. 113 (1981) 267–314 (not consulted directly).
- [Ta] W. W. Tait, *Normal derivability in classical logic*, in: J. Barwise (ed.), The Syntax and Semantics of Infinitary Languages, LNM 72, Springer, 1968, 204–236 (not consulted; cited via [BW]).
- [W70] S. S. Wainer, *A classification of the ordinal recursive functions*, Arch. Math. Logik Grundlagenforsch. 13 (1970) 136–153 (not consulted directly; used only through [BW] Theorem 5).
- bbchallenge wiki, *Beeping Busy Beaver*. https://wiki.bbchallenge.org/wiki/BBB