Collapsing numbers in bases 2, 3, and beyond

Steve Butler* Ron Graham† Richard Stong‡

Gathering for Gardner 10

Abstract

Starting with a number n and a base b you can collapse it to a smaller number by writing the number in base b and then writing plus signs between the digits (i.e., splitting the expansion of n into pieces and adding them up). By strategically placing the plus signs then in base 2 we can always collapse n to a single digit using at most 2 steps, in base 3 this can also be done in at most two steps with the exception of 11 numbers which require 3 steps, and in base 4 and above 3 steps always suffice.

1 Introduction

Starting with a number n and a base b we will repeatedly apply the following procedure which collapses n to a smaller number until we get down to a single digit.

1. Write n in base b and insert, as desired, plus signs in between the digits. (This couples digits into groups which become new numbers in base b.)
2. Do the indicated addition operations in base b and add up the result to get n'.
3. If n' is a single digit then we are done; otherwise return to step 1 replacing n with n'.

Given any number and base this procedure can always be done in finitely many steps, for instance we can break the string up into substrings of length one and add up the resulting entries. In this case the new value is smaller and on average we would expect $n' \approx c \log n$ so we would expect to take approximately $\log^* n$ steps (where $\log^* n$ is the number of logs that need to be applied to n to get to 1). For example if we start with the number 10211914 (Martin Gardner's birthdate) in base 10 we would have the following:

$$\begin{aligned} 10211914 &\longrightarrow 1+0+2+1+1+9+1+4=19 \\ &\longrightarrow 1+9=10 \\ &\longrightarrow 1+0=1 \end{aligned}$$

A number can only be done in one step if the sum of the digits of the number written in base b is less than b, so 10211914 cannot be done in one step in base 10. There is a unique way to get this number to a single digit in two steps; we leave this as a challenge to the reader.

*Iowa State University, Ames, IA 50011 butler@istate.edu
†UC San Diego, La Jolla, CA 92093 graham@ucsd.edu
‡Center for Communications Research, La Jolla, CA 92121 stong@ccrwest.org

In general we are interested in the fewest number of steps that it takes to get down to a single digit. The case for base 2 was a problem of Gregory Galperin featured in the Puzzles Column of *Emissary* [1].

The amazing thing is that no matter the size of n of the choice of our base, it can always be done very efficiently. In this note we will outline what is known about this problem for bases 2, 3, and 4 and greater.

Theorem 1. *Any number n using base 2 can be collapsed to a single digit in at most* two *steps. Equivalently, any number can be collapsed to a power of 2 in just one step.*

Theorem 2. *Any number n using base 3 can be collapsed to a single digit in at most* two *steps except for the eleven numbers listed below.*

$$\begin{array}{rlrl} 1781 = 2102222_{(3)} & & 41065 = 2002022221_{(3)} \\ 3239 = 11102222_{(3)} & & 43981 = 2020022221_{(3)} \\ 3887 = 12022222_{(3)} & & 98657 = 12000022222_{(3)} \\ 11177 = 120022222_{(3)} & & 131461 = 20200022221_{(3)} \\ 14821 = 202022221_{(3)} & & 393901 = 202000022221_{(3)} \\ 33047 = 1200022222_{(3)} & & \end{array}$$

Theorem 3. *Any number n using base 4 or higher can be collapsed to a single digit in at most* three *steps. Further, there are infinitely many numbers which require three steps.*

2 Collapsing numbers in base 2

We start with the simpler case of base 2 to establish Theorem 1. The main approach is to split the work into two parts. First we will show that if the the number of ones is large then we can easily find a way to couple our digits to get to the nearest larger power of 2 (we initially break our number into singletons and then rejoin, or couple, consecutive digits back together to increase the sum to the desired value). Second we show that the result still holds if the number of ones is small.

We let m denote the number of ones in the base 2 representation of n (i.e., what we would get from breaking the number up into singletons and adding as we did above). If m is a power of 2 we are done; otherwise suppose that 2^k is the smallest power of 2 greater than m. Then our goal will be to find a way to couple numbers so that we can make up the difference $D = 2^k - m$.

To start we will look at what happens when we couple some small number of singletons together. In the table below the "$*$" represents a digit which can be either 0 or 1 (the digit in the last position of a coupled number does not contribute to the change in the difference caused by the coupling).

replace	with	change
$1 + *$	$1*$	+1
$1 + 0 + *$	$10*$	+3
$1 + 1 + *$	$11*$	+4
$1 + 0 + 0 + *$	$100*$	+7
$1 + 0 + 1 + *$	$101*$	+8
$1 + 1 + 0 + *$	$110*$	+10
$1 + 1 + 1 + *$	$111*$	+11

replace	with	change
$1 + 0 + 0 + 0 + *$	$1000*$	+15
$1 + 0 + 0 + 1 + *$	$1001*$	+16
$1 + 0 + 1 + 0 + *$	$1010*$	+18
$1 + 0 + 1 + 1 + *$	$1011*$	+19
$1 + 1 + 0 + 0 + *$	$1100*$	+22
$1 + 1 + 0 + 1 + *$	$1101*$	+23
$1 + 1 + 1 + 0 + *$	$1110*$	+25
$1 + 1 + 1 + 1 + *$	$1111*$	+26

We can of course continue filling in this table, but this will suffice for our needs and the moral here is that by coupling numbers together we can reduce the difference D to be made up. Our basic strategy then, at least for large m, will be to start at the left hand side of a number and couple together the largest number we can without making the net change go over D, then repeat the process with our new difference with the remaining singletons as often as needed. This is an efficient strategy as the following theorem shows.

Theorem 4. *Given a number n written in base 2 containing q digits and a difference $1 \le D < 2^q$ that we are trying to make up by coupling digits together, take the largest initial string of the expansion of n for which the total change does not exceed D. With the remaining uncoupled digits we now have to make up a difference of at most $\frac{2}{3}D$.*

Proof. Given that $D \ge 1$ then we can couple at least the first two terms together making a difference of 1 and not exceeding D, so coupling can be used to reduce D. On the other hand since $D < 2^q$ then we cannot couple all the digits together without going over $\frac{1}{2}D$, so we must either stop at some point or take the whole string and in the latter case the result still holds.

Now suppose that we have coupled the first $r+1$ digits, i.e., $a_r \dots a_0$. The contribution Δ that this will have to the difference is

$$\Delta = (2^r a_r + \cdots + 2a_1 + a_0) - (a_r + \cdots + a_1 + a_0).$$

On the other hand by assumption if we had coupled the first $r+2$ digits, i.e., $a_r \dots a_0 x$ then we would exceed D, i.e.,

$$D + 1 \le \Delta' = (2^{r+1}a_r + \cdots + 2^2 a_1 + 2a_0 + x) - (a_r + \cdots + a_1 + a_0 + x).$$

Combining these we can conclude that $\Delta' - 2\Delta = a_r + \cdots + a_1 + a_0$. On the other hand we also have $a_r + \cdots + a_1 + a_0 \le 2^r a_r + \cdots + 2a_1 + a_0 = \Delta$. Using that $D \le \Delta' - 1$ we have

$$D - \Delta \le \Delta' - 1 - \Delta = 2\Delta + a_r + \cdots + a_0 - 1 - \Delta < 2\Delta.$$

So we can conclude that $D < 3\Delta$ or $\Delta > \frac{1}{3}D$. Therefore we have the remaining difference that needs to be made up is $D - \Delta < \frac{2}{3}D$. ☐

In particular, at every stage we can quickly collapse the difference that needs to be made up. Also note that $(\frac{2}{3})^2 = \frac{4}{9} < \frac{1}{2}$, so we will never couple ℓ digits more than twice for any ℓ.

Returning to our original goal, we want to couple numbers together to make up a difference of $D = 2^k - m$. Since $D < 2^k$ we initially couple at most k terms together, and further, in such a couple we will use at most k ones. We then continue to couple numbers together and as noted we quickly reduce the size of the coupling that we need.

Therefore we need to use at most $2k + 2(k-1) + 2(k-2) + \cdots + 2 = k(k+1)$ ones in reducing our number (though in practice we will need far fewer). This gives us a simple bound on the number of ones for which the trivial coupling off the largest initial number works, namely, if the number of ones in the base 2 expansion of n is at least $k(k+1)$. On the other hand we have $2^{k-1} < m$. So if $k(k+1) < 2^{k-1}$ then we have enough ones to do this strategy. This relationship holds for $k \ge 7$, in particular it holds when $m \ge 2^6$. So that leaves us with the cases when the number of ones in our base 2 expansion is below 64.

We now come to the second half of our process which is to look at what happens when the sum of digits is small, in particular below 64. The coupling strategy is still very effective and we can handle most cases using this technique but being more careful in counting. Also we can employ some simple tactics for many cases, i.e., if $m \geq 2D$ then we can couple D pairs of the form $1*$ in our number. This handles most of the cases leaving us with the following:

m	5	9	10	17	18	19	20	21	33	34	35	36	37	38	39	40	41	42
D	3	7	6	15	14	13	12	11	31	30	29	28	27	26	25	24	23	22

- $m = 5$: If there is a $10*$ in the expansion of n then we can couple these terms together giving a difference of 3 and we are done. If there is no $10*$ then our number must be of the form 11111 followed by some number of zeros. If there is at least one zero we can form three pairs of $1*$ and get a difference of three and we are done. Otherwise there is no zero and we use the following strategy $11111_{(2)} \to 1111_{(2)} + 1_{(2)} = 10000_{(2)}$.

- $m = 9$: If the second digit in the expansion of n is a 0 then we can pair up the first four digits losing at most three ones and have at most a difference of two to make up, which we can easily do since there will be at least six ones remaining that we can use to make $1*$.

  If the second digit is a 1 then we can pair up the first three numbers reducing our difference by 4 and then couple our next available three digit number. We have then used at most five ones and reduced it by 7 or at most six ones and reduced it by 8. In either case we can still finish by forming some $1*$s.

- $m = 10$: We couple the first three terms together. If we coupled $10*$ (or $11*$) then we have at least eight (or seven) ones left to make up a difference of 3 (or 2) which can be done by forming some $1*$s.

- $m = 17$: We couple the first four terms together. There are several sub-cases:

  - If we coupled $100*$ (or $101*$) then we have at least 15 (or 14) ones remaining to make up a difference of 10 (or 9). If the digit immediately following the next available 1 is a 0 then we can again pair up the next four consecutive terms and then find pairs $1*$ to finish the case. Otherwise we pair up the next block of size three and reduce the remaining difference by 4 using at most three ones. We now have at least 12 (or 11) ones to make up a difference of 6 (or 5) which we can easily do by coupling pairs $1*$.
  - If we coupled $110*$ (or $111*$) then we have at least 14 (or 13) ones remaining to make up a difference of 7 (or 6), and this can easily be accomplished by coupling pairs $1*$.

- $m = 18, 19, 20, 21$: We split the expansion into two parts, one involving the first 9 ones and the rest which will have $9, 10, 11, 12$ ones respectively. From the first part we can use the strategy above for $m = 9$ to achieve a difference of 7, and we use the strategies for $m = 9, 10, 11, 12$ respectively on the second parts to achieve differences of $7, 6, 5, 4$ giving us precisely what we need.

- $m = 33$: We couple the first five terms together. There are several sub-cases:

  - If we coupled $1000*$ (or $1001*$) then we have at least 31 (or 30) ones remaining to make up a difference of 18 (or 17). We then take the next block of 9 ones and use the strategy

for $m = 9$ to get a difference of 7 leaving us with at least 22 (or 21) ones to make up a difference of 11 (or 10), and this can easily be accomplished by coupling pairs $1*$.
  - If we coupled $1010*$ (or $1011*$, $1100*$, $1101*$, $1110*$, $1111*$) then we have at least 30 (or 29, 30, 29, 29, 28) ones remaining to make up a difference of 15 (or 14, 11, 10, 8, 7), and this can easily be accomplished by coupling pairs $1*$.

- $m = 34, 35, 36, 37, 38, 39, 40, 41, 42$: We split the expansion into two parts, one involving the first 17 ones and the rest which will have $17, 18, 19, 20, 21, 22, 23, 24, 25$ ones respectively. From the first part we can use the strategy above for $m = 17$ to achieve a difference of 15, and we use the strategies for $m = 17, 18, 19, 20, 21, 22, 23, 24, 25$ respectively on the second parts to achieve differences of $15, 14, 13, 12, 11, 10, 9, 8, 7$ giving us precisely what we need.

This finishes the small cases and we have shown that every number in base 2 can be done in two steps, and in fact, we have shown something stronger.

Theorem 5. *Given a number n let m be the number of ones in the base 2 expansion of n and let 2^k be the smallest power of 2 greater than or equal to m. Then except for $n = 31 = 11111_{(2)}$, in one step we can take n to 2^k.*

There are other ways to establish the above result. In particular, if we choose to generalize what was done for the cases 18-21 and 34-42 then it suffices to show that it holds for $m = 10$ and $m = 2^k + 1$ for all k. Our choice of method is to introduce techniques that can be used for base 3.

3 Collapsing numbers in base 3

We now turn to the more interesting base 3 case and outline a proof of Theorem 2. While in the base 2 case we always would finish the process with a single 1, now there are two possibilities, either we finish with 1 or 2. We can always tell at the very beginning which one of the two will occur. This is because when we couple the digits together we do *not* change the value modulo $b-1$. (This works by the same principle which states a number is divisible by 9 if and only if the sum of its digits is divisible by 9.)

In particular, if m is the sum of the digits in our base 3 expansion, then if m is odd we will finish with 1 and if m is even we will finish with 2. This tells us if m is odd then our first step is to get to a power of 3 while if m is even then our first step is to get to a sum of two powers of 3. This makes the even case much easier in general because there are more targets for us to shoot for.

We want to take the same basic coupling of the largest initial run we can. We have the following result which tells us that again coupling is efficient. The proof is nearly identical to the base 2 case and so we omit the proof here.

Theorem 6. *Given a number n written in base 3 containing q digits and a difference $4 \le D < 3^q$ that we are trying to make up by coupling digits together, take the largest initial string of the expansion of n for which the total change does not exceed D. With the remaining uncoupled digits we now have to make up a difference of at most $\frac{3}{4}D$.*

In particular, at every stage we can quickly collapse the difference that needs to be made up. Also note that $(\frac{3}{4})^4 = \frac{81}{256} < \frac{1}{3}$, so we will never couple ℓ digits more than four times for any ℓ.

Let m be the sum of digits of the base 3 expansion of n. We now look at some cases depending on the parity of m and the nature of the expansion.

m is even

In this case we want to get to a sum of two powers of three. So choose k as small as possible so that $m \leq 3^k + 3$. Our goal is to make up the difference between these, i.e., $D = 3^k + 3 - m$. Since $D < 3^k$ then we initially couple at most k terms together, further in such a couple we will lose at most k non-zero terms. We then continue to couple numbers together until we either reach our goal, or we can no longer couple terms. If we reached our goal we are done. On the other hand if we stopped coupling because $D < 4$ then we are also done because that means we have made up the difference to get us to $3^k + 1$ which can be done in one step.

So we only need to make sure that we do not run out of non-zero terms before we stop coupling. Counting we will have used at most $4k + 4(k-1) + \cdots + 4 = 2k(k+1)$ non-zero terms in coupling. On the other hand we have that $3^{k-1} + 3 < m$ and the number of non-zero digits is at least $m/2 > (3^{k-1} + 3)/2$. Therefore if

$$2k(k+1) \leq \frac{3^{k-1} + 3}{2}$$

then we have enough non-zero terms to finish the coupling. This inequality holds for $k = 6$ and therefore we have that if $m > 3^5 + 3 = 246$ then we can use the coupling strategy to get the result. We point out that we have trivial strategies available for $m = 244$ and $m = 246$.

m is odd with a 1 not at the last entry

Part of the advantage of the even case is we had two numbers to shoot for that were close to one another so failure to reach one target corresponded to successfully hitting the other. For the odd case our targets are so far apart that failure to hit our goal is catastrophic. So we need to have some insurance so that if coupling stops and we are 2 short then we can make up the difference. To help do that we will look for a $1*$ and "protect" this small block. We do this by finding the first 1 in the expansion and by our assumption this 1 is not at the end and we protect this number and the digit immediately following it.

Now we couple as before, but if our coupling were to ever contain part of the protected block then we skip past the block and continue, but at the *soonest* opportunity we will finish coupling what comes before the block. In particular this means that keeping the $1*$ block safe will cost us at most 4 non-zero terms (i.e., the possible two non-zero digits of $1*$ and since we used up everything before the block at the earliest opportunity at most two more non-zero digits went unused). So we continue until we can no longer couple. If we are at our goal we are finished, otherwise we have a difference of 2 to make up and we use the $1*$ to bridge the gap.

So let k be the smallest value so $m \leq 3^k$ and that we are trying to make up the difference $D = 3^k - m$. Then as in the above case we will need at most $2k(k+1) + 4$ non-zero terms to use in coupling. On the other hand we have that $3^{k-1} < m$ and the number of non-zero digits is at least $m/2 > 3^{k-1}/2$. Therefore if

$$2k(k+1) + 4 \leq \frac{3^{k-1}}{2}$$

then we have enough non-zero terms to finish the coupling. Again this inequality holds for $k = 6$ and therefore we have that if $m > 3^5 = 243$ then we can implement the coupling strategy to get the result. We point out that we have a trivial strategy available for $m = 243$.

m is odd with a single 1 at the last entry

This is a worst case scenario. The problem is we must go up by multiples of 4 when we couple and so we have to shoot for a power of 3 that differs from m by a multiple of 4. Let k be the smallest power of 3 so that $m \leq 3^k$ and 4 divides $3^k - m$. Then we have that $3^{k-2} < m$ (i.e., powers of 3 alternate between ± 1 modulo 4).

This is one of the reasons why $m = 13$ is an unusually hard case (as indicated by having 9 of the 11 exceptional numbers for base 3). When we have a single 1 at the last entry then we have to get to 81 or 729 and with relatively few non-zero digits this is difficult.

As before we will never couple ℓ digits more than four times for any ℓ. Therefore in our coupling we will never use more than $2k(k+1)$ non-zero digits. Therefore if

$$2k(k+1) \leq \frac{3^{k-2}+1}{2}$$

then we have enough non-zero terms to finish the coupling. This inequality holds for $k = 7$, and therefore we have that if $m > 3^5 = 243$ then we can implement the coupling strategy to get the result.

Medium sized m

We now have established the result for sufficiently large m, in particular we know the result holds if the sum of digits is at least 243. We can do better if we keep track of how much coupling will be involved as we go through the process. Note that at each stage we go down by at least $\frac{1}{4}$ and the remaining difference to make up must always be even. Suppose for example we have $m = 90$, in the worst case scenario as given by Theorem 6 we have the following situation where the row for D and non-zeroes corresponds to upper bounds for these quantities:

step	1	2	3	4	5	6	7	8	9	10	11	12
D	156	116	86	64	48	36	26	18	12	8	6	4
non-zeroes to use	5	5	5	4	4	4	4	3	3	3	2	2

In particular the coupling will use at most 44 non-zero terms, but we have at least 45 and so we can handle this case. The only even case with $m > 81$ that we cannot easily handle by this moderately improved bookkeeping is $m = 86$. But this is a trivial case because $90 = 3^5 + 3^3$ so we only have to make up a difference of 4 which we can do by either a single $2*$ or two $1*$s and one of these two must occur.

Similarly we can keep track of how much coupling will be involved in the case when m is odd and we have a $1*$ protected block. Modestly improved record keeping allows us to handle everything except for $m = 83, 85, 87, 89, 91, 93$.

For these cases we can simply observe that to this point we have been *extremely* generous when we looked at the number of non zeroes used and how much our difference declined. We can be a little bit more careful about the range of possibilities. In the chart below we have listed the possible differences that can occur as we pull off a string.

length of string	5	5	5	5	5	4	4	4	4	3	3	3	2	2
number of non zeroes	5	4	3	2	1	4	3	2	1	3	2	1	2	1
maximum Δ	232	232	228	212	160	72	72	68	52	20	20	16	4	4
minimum Δ	116	114	106	80	80	36	28	26	26	10	8	8	2	2

In addition we have the following facts which give us great freedom when we have at least 18 nonzero terms left.

Fact 1. *Given any arrangement of 18 non-zero terms and any arbitrary number of zeroes we have the following.*

- *If there is at least a single* $1*$ *in the arrangement then any even difference over the sum of the singletons from* 2 *to* 44 *can be achieved by coupling.*

- *If there is no* $1*$ *in the arrangement then any difference over the sum of the singletons of the form a multiple of four from* 4 *to* 176 *can be achieved by coupling.*

Branching on the various possibilities of how the string starts allows us to quickly reduce our difference to a manageable range. This also happens for $m = 85, 87, 89, 91, 93$ when there is some $1*$ in the expansion.

Now let us turn to the case when $m > 81$ is odd and we have a single 1 at the end. Our modestly improved bookkeeping does not work for $m = 85, 89, 93, \ldots, 133$. Let us consider these cases. If we repeat the above table for $n = 85$ we have the following situation (because of parity we need to get at least to $3^7 = 729$):

step	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
D	644	480	360	268	200	148	108	80	60	44	32	24	16	12	8	4
non-zeroes to use	6	6	6	6	5	5	5	5	4	4	4	3	3	3	3	2

This would seem to indicate colossal failure because there is no way we have enough zeroes. As in the last case we can easily handle this with some simple case analysis.

- Suppose the first two digits are 20. Then we can group the first six digits together and we will reduce the difference by *at least* 480 which will get us into a range for which we can use the above fact to finish off the case.

- Suppose the first two digits are 22. Then we can pull off the first five digits together, and then pull off the next group of five starting with the next available non-zero, at this point the difference has gone down by at least 372 so have at most a difference of 272 to make up and we still have at least 31 non-zeroes to do this. We now branch on several possibilities (depending on the remaining difference and the leading terms) and in each case we can get to a difference below 176 to make up while we still have at least 18 nonzero digits to work with.

The cases for $m = 89, 93, \ldots, 133$ can be handled in a similar fashion. (In general the hardest cases are the ones just past a power of three because we have fewer nonzero terms and a larger difference to make up.)

This finishes off the medium cases.

Small sized m

We now turn to the remaining cases when $m \leq 81$. Most of the upper range can easily be handled using similar analysis as for the medium sized m. As m gets smaller the approach gets more complicated and we have to look at many more cases. We opted to use brute force computation. In particular, suppose we put a limit to our coupling size so that we do not couple anything larger than size 5. Then we can reduce the number of remaining cases to a finite collection of possibilities because we can take any long run of zeroes and reduce it to a run of size 4. We now run over this large finite collection and we either find a way to couple to make up our difference in which case we are done or we cannot find a way to couple to make up our difference and we examine the number to see if it is either in the case where we need to take three steps or there is a better coupling strategy available to us that simply requires a larger coupling.

The above computations for $m \leq 81$ was carried out on a `sage` server (where we also used some efficient pruning to reduce the number of cases) and we were able to determine when $m \leq 81$ the only cases that take three steps are the ones given in Theorem 2. A complete output of the program plus a copy of the `sage` code is available online[1].

A similar computation was used to establish the preceding fact about the differences that can be achieved with some arbitrary patterns.

It is interesting to note that things get much more difficult in the base 3 case. For example we know that there are many cases where we must skip over the next power of 3, whereas in base 2 there was only one number for which this had to happen. The most extreme case of this in base 3 is $387420487 = 2222222222222222221_{(3)}$, the sum of the digits is 35 so in the best case scenario we would want to try to get to $3^5 = 243$, but in fact there are only two ways to collapse this number in two steps and both of them start by going to 3^{17}.

Many of these small cases are easily handled, for example we invite the reader to establish the fact for $m = 7$ by hand. For the reader who is more adventurous they can also try to do $m = 11$ and $m = 13$ by hand.

4 Collapsing numbers in base 4 and higher

Most of the work in the base 3 case was ensuring that we could collapse most of the numbers in two steps. When we relax and allow ourselves a third step that gives us a lot more breathing room and a much simpler proof. For example using the consequences of Theorem 4 we know if the sum of digits $m \geq 243$ that two steps suffice. On the other hand if the sum of digits is $m < 243$ then in our first step we simply add up the digits (i.e., take n to m) and only need to verify that we can then finish whatever is left in two steps. This is easy to do since in our second step we will be left with a number whose sum of digits is at most 10 and these are easy cases when the base is 3.

For bases 4 and above we will use this a similar approach. Namely we will show if the sum of digits is small then we can easily finish in three steps; and if the sum of digits is large we have a lot of flexibility in coupling to again allow us to finish in three steps.

Lemma 1. *Let $b \geq 4$ be our base. Then any $n < 3b^2 - b - 1$ can be collapsed to a single digit in at most two steps.*

Proof. First we observe that for $n = 1, 2, \ldots, 2b - 2$ that we can always apply the trivial strategy of adding up the digits and collapse to a single digit. The first nontrivial number for which the simple sum of digits strategy does not get us down into this range is $1(b-1)(b-1)_{(b)}$, but for this number we use the strategy

$$1(b-1)(b-1)_{(b)} \rightarrow 1_{(b)} + (b-1)(b-1)_{(b)} = 100_{(b)} \rightarrow 1_{(b)} + 0_{(b)} + 0_{(b)} = 1_{(b)}.$$

The next smallest number is $2(b-2)(b-1)_{(b)} = 3b^2 - b - 1$, establishing the lemma. □

[1]Available at `http://www.math.iastate.edu/butler/base_3.zip`.

This lemma is tight, the number $2(b-2)(b-1)_{(b)}$ takes three steps as is easy to check by a few cases. In fact more is true.

Observation 7. *Let $b \geq 4$ be our base. Then $20\ldots0(b-2)(b-1)_{(b)}$ takes two steps for* any *number of zeroes. In particular, there are infinitely many numbers that take three steps.*

To see this recall, as noted in the base 3 case, the final digit is completely determined by the initial sum of digits modulo $b-1$. In this case we see that the final digit at the end of this process will be 1. So if it could be done in two steps we would have to be able to get it to a number of the form $10\ldots0_{(b)}$, i.e., a power of b. In particular the last digit of any coupling strategy would have to be zero, but for a number of this form the last digit in coupling would come from $b-1$, $(b-1)+(b-2)$, $(b-1)+2$ or $(b-1)+(b-2)+2$ and none of these are 0 modulo b when $b \geq 4$. (For $b=3$ we can get a 0 in the last digit and so this number can be done in two steps.)

If the sum of digits is small then we can apply Lemma 1 after doing a simple step of adding all the digits. We now show that if the sum of digits is large then we can take one step to get us to a number whose form can be finished in at most two more steps.

Lemma 2. *Let $b \geq 4$ be our base and let n a number with the sum of its digits $m \geq b^2$. Then in one step n can be collapsed to a number of the form $c0\ldots0de$ where $c \leq 2$ and $de_{(b)} \leq b^2-2b$.*

Proof. Let $n = (\ldots a_4a_3a_2a_1a_0)_{(b)}$ and let $A = \max\{a_1+a_3+\cdots, a_2+a_4+\cdots\}$. We do not use the last digit for A and so we have $A \geq (m-(b-1))/2$. We now consider what happens if we couple in pairs so that the leading digits in the pairs sum to A (i.e., we pair so that all the digits are even or odd depending on which gave us A). The sum total of this coupling strategy will be

$$(b-1)A + m \geq \frac{(b-1)(m-b+1)}{2} + m = \frac{mb+m-(b-1)^2}{2} > \frac{mb}{2}.$$

(The last step is by our assumption that $m \geq b^2$.) If we now break these pairs one at a time, say from left to right, then the difference in the total would be at most $(b-1)^2$ at each pair. Therefore we have a sequence of coupling strategies which go from m to $(b-1)A+m$ where the difference between two consecutive strategies is at most $(b-1)^2$.

Now m is in an interval of the form $[b^t, 2b^t)$ or $[2b^t, b^{t+1})$ for some $t \geq 2$. However we have that $(b-1)A+m > bm/2$ can not be in the same interval (here we use that $b \geq 4$). Therefore there will be some smallest coupling strategy which will exceed the top of the given range containing m. Let M be the resulting total using this coupling. Then we either have $2b^t \leq M < 2b^t + (b-1)^2$ or $b^{t+1} \leq M < b^{t+1} + (b-1)^2$, depending on which case we are in, which gives exactly the sort of base b representation as given in the statement of the lemma. □

So to establish Theorem 3 if the sum of digits $m < b^2$ then sum the singletons and then apply Lemma 1 to the result and we take at most three steps. On the other hand, if the sum of digits $m \geq b^2$ then by Lemma 1 in one step we will collapse to a number of the form $c0\ldots0de$ with $c \leq 2$ and $db + e \leq b(b-2)$. In particular we have that $d+e \leq (b-3)+(b-1) = 2b-4$, and so adding all the singletons gives a number of size at most $2b-2$, which can be finished in one more step.

References

[1] Elwyn Berlekamp and Joe Buhler, Puzzles Column, *Emissary*, Fall 2011, 9.