

Number Representations and Registers*)

By *Ulrike Brandt*

Abstract: We investigate the following question: Are the usual geometric bases the best possible choice with respect to additive number representation in registers? The answer depends obviously on an appropriate measure. We choose a product measure taking into account the necessary register length as well as the energy performance caused by certain standard operations. Using special results from additive number theory and upper bounds for continuants we can show that the Fibonacci numbers with even respectively with odd index are optimal infinite bases with respect to our measure. This result is achieved under the additional assumption that only number representations are used, which can be created by Euclid's algorithm.

1. Introduction

Nearly all current architectures of computers are based on von Neumann's architecture, that means we meet an arithmetical-logical unit, memories of different kind, peripheral units and so on. The arithmetical-logical unit performs the arithmetical and boolean operations. Usually it has a memory of registers with rapid access time. Mostly, these registers contain numbers in binary representation or in a related representation to another basis than two. Our question is, if this kind of "geometric" number representation is the best possible choice. Clearly, with respect to human purposes it seems to be natural to use such systems, especially the decimal system. However, even if these systems may be unavoidable for the man-machine communication, faced with modern technologies, like VLSI, it could be much better to use exotic basis systems for number representation, for example looking at machine-machine communication.

Asking this question, we are faced at least with three different approaches. First, leave the architecture of registers at all, secondly cancel the additive representation of numbers, thirdly accept standard architectures and use non-geometric representations. It is well known, that the second approach has been used at least in theory [8]. We shall attempt the third approach, which is the modest one of these three. This means, we extend the usual way to get the represented number multiplying the content of the register cells with the powers of two (or another basis number), allowing squares, cubes, primes, faculty and so on. To be more precise, we introduce the notion of a *minimal additive* representation of a number with respect to a *basis* A .

A *basis* is simply a subset $A = \{a_1, a_2, \dots\} \subseteq \mathbb{N}$ with $1 \in A$. This set can be finite or infinite. For convenience we assume $1 = a_1 < a_2 < a_3 < \dots$. Given a basis, every $x \in \mathbb{N}$ has a representation $x = \sum_{a_i \in A} n_i a_i$. In the set of these representations we distinguish so called

*) Keywords: Registers, number representations, bases, section bases, reach of a base, regular reach of a base, continuant.

minimal representations containing the least possible number of terms, more precisely, where $\sum_{a_i \in A} n_i$ is minimal. Accordingly, x is represented in a given register R , if the i -th cell contains the “digit” n_i , where $x = \sum_{a_i \in A} n_i a_i$ is a minimal representation. Note that in general n_i cannot be interpreted as a digit in the usual way, because the n_i ’s may be unbounded.

Our attempt is now to study the effect for standard operations, if the basis is changed for example from $\{2^i \mid i \geq 0\}$ to some exotic subset of \mathbb{N} . Every effort in this direction depends obviously on the choice of an appropriate measure. Let us take a brief look on the question how to fix it.

The first criterion should be the length of a register, if you want to represent *all* numbers of an interval $[1..x] = \{y \in \mathbb{N} \mid y \leq x\}$. Representing a number x in our way, elements a_i of A are useless if $a_i > x$. Hence, the register needs at most as many cells as there are $a_i \in A$ with $a_i \leq x$. This number is denoted by $A(x)$. Now, looking at $A = \{1\}$, a possible base, we get obviously $A(x) = 1$ for every x . With respect to the length of the register we can’t do better. But this choice of A is absurd, because the one necessary register cell replaces a whole register alone, if we look at the internal representation of the “digit”. Thus, $A(x)$ doesn’t suffice as a single criterion, i.e. we need at least one more different measure.

To this purpose we choose an energetic measure concerning standard operations. To make things easier, we skip at this time the quite complicated arithmetical operations and restrict ourselves to load operations and comparisons. For these operations we can assume that the amount of energy is proportional to the amount of destroying, conserving and building up the necessary energy profile of *one* register. Typical operations are: load zero (reset), compare with zero, load a number into an empty register.

Considering again an interval $[1..x]$ we set this amount proportional to $T(x, A)$, that is the number of terms from A appearing in worst case in a minimal representation of some $y \in [1..x]$. Thereby we assume that preserving zero costs no energy at all. Now, looking at $A = \mathbb{N}$, we get $T(x, A) = 1$ because every $y \in [1..x]$ is representable by the single term $y \in A$. But now $A(x) = x$ for all x . Hence, $T(x, A)$ alone is an unfair measure again. Looking back to the choice $A = \{1\}$ we observe that both choices are characterised by the fact, that the whole information is contained in *one* cell. This parallel is captured, if we consider the product measure $A(x) \cdot T(x, A)$, since $T(x, \mathbb{N}) \cdot \mathbb{N}(x) = T(x, \{1\}) \cdot \{1\}(x) = x$. Using $A = \{2^i \mid i \geq 0\}$, we get $T(x, A) \cdot A(x) = \log_2^2(x + 1)$ (provided $x + 1$ is a power of 2), which is better than the value for the extremal choices. In conclusion, the examples show that there is a trade-off between $T(x, A)$ and $A(x)$.

We are now in the position to establish our problem in a precise way: determine to any x $M(x) = \text{Min}\{T(x, A) \cdot A(x) \mid A \subseteq \mathbb{N}, 1 \in A\}$ and all A with $M(x) = T(x, A) \cdot A(x)$. In the following we shall refer to this problem as the $M(x)$ -problem. Clearly, it is possible to use further criterions, for example other energetic measures or complexity measures for arithmetical operations. But by solving the $M(x)$ -problem we shall get at least an overview over all reasonable choices of bases A .

2. Rohrbach’s reach-problem

It seems that the $M(x)$ -problem is currently too hard to be solved. The reason is that it is completely equivalent to an unsolved problem in additive number theory. Consider a

finite base $A \subseteq \mathbb{N}$ and $h \in \mathbb{N}$. Let $n(h, A) = \text{Max}\{x \in \mathbb{N} / T(y, A) \leq h \text{ for all } 1 \leq y \leq x\}$. This number is called the h -reach of A . Choosing $A = \{1, 4, 9, 31, 51\}$, every $x \in [1..126]$ is representable by less than 6 terms. For example $125 = 4 \cdot 31 + 1 \cdot 1$ and $126 = 1 \cdot 51 + 2 \cdot 31 + 1 \cdot 9 + 1 \cdot 4$. By testing we get that 127 needs at least 6 terms. Hence $n(5, A) = 126$.

The reach-problem of Rohrbach [7] is now to compute to a given $h, k \in \mathbb{N}$ the number $n(h, k)$, which is the maximal h -reach of all bases A with cardinality k , and moreover to determine all those A with $n(h, k) = n(h, A)$. These bases A are called *extremal section bases*. The finite set $A = \{1, 4, 9, 31, 51\}$ from the above example is an extremal section base for $h = k = 5$, because there is no base of cardinality 5, such that every $x \leq 127$ is representable by at most 5 terms. Thus $n(5, 5) = 126$. See [5] for further details.

The connection to the $M(x)$ -problem is easily seen observing that the two problems, to minimize $T(x, A) \cdot A(x)$ for $A \subseteq \mathbb{N}$ and to determine all pairs (h, k) with $n(h, k) \geq x$, such that $h \cdot k$ is minimal, are equivalent. Hence, $M(x) = \text{Min}\{hk / n(h, k) \geq x\}$. Unfortunately, only few values of $n(h, k)$ are known.

To overcome this difficulty, one can proceed in two different ways. The first way is to restrict the set of admissible bases to certain subclasses, for example to geometric bases of type $\{p^i \mid i \geq 0\}$ with $p \geq 2$, while the second way is to allow not all representations but only some canonical ones. We take the second approach restricting the set of admissible representations to those, which we get by applying Euklid's algorithm, this means we use every element $a_i \in A$ in descending order as often as possible. This algorithm leads for every $x \in \mathbb{N}$ to a unique representation of the form

$$x = \sum_{a_i \in A} n_i a_i \quad \text{with} \quad \sum_{1 \leq i \leq j} n_i a_i < a_{j+1} \quad \text{for all } j.$$

Following Hofmeister [4] we call these representations *regular*. If we look at our running example, we get $2 \cdot 51 + 2 \cdot 9 + 1 \cdot 4 + 2 \cdot 1$ as regular representation of 126. Observe that the number of terms is 7, which is greater than the best possible number 5.

In general, the minimal representation is not unique so that even testing equality is a harmful task. Moreover, the coefficients can by no means be interpreted as digits in the usual way. Looking at applications, the useful representations should possess at least these two properties, so that the restriction to regular representations is reasonable.

Now, we rephrase our problem. Define $S(x, A)$ in the same way as $T(x, A)$, but use only regular representations. Then the $m(x)$ -problem is to minimise $S(x, A) \cdot A(x)$ and to determine all $m(x)$ -optimal bases A , that means all A with $m(x) = S(x, A) \cdot A(x)$. Again, the $m(x)$ -problem is equivalent to a problem of additive number theory, introduced by Hofmeister [4] as an analog of Rohrbach's problem. This problem is called the *regular reach-problem*, which differs from the general version by using exclusively regular representations. Analogously, we are speaking of *regular h-reach* and *extremal regular section bases*. In this case the numbers $n(h, A)$ and $n(h, k)$ are replaced by $g(h, A)$ and $g(h, k)$. As before, we get $m(x) = \text{Min}\{hk / g(h, k) \geq x\}$.

We are now in a better position, because Mrose has solved the regular reach-problem in 1969 [6]. Using Mrose's results we can determine $m(x)$ and the $m(x)$ -optimal bases A .

3. Mrose's results

A first step to determine $m(x)$ is to find those pairs (h, k) with $m(x) = hk$ for some x . Hofmeister [4] has shown that $g(h, k) = g(k, h)$ holds, hence, it is enough to consider all pairs (h, k) with $h \geq k$.

So, our discussion starts in solving the following problem: determine all $(h, k) \in \mathbb{N}^2$ with $h \geq k$, fulfilling the following property: if $(h', k') \in \mathbb{N}^2$ with $g(h', k') \geq g(h, k)$ then $h'k' \geq hk$. Such pairs are called *r-optimal*. The stated problem can be solved with the help of an algorithm due to Mrose, which computes $g(h, k)$.

To do this, we need the notion of a continuant. A finite sequence μ of real numbers x_1, \dots, x_n will be written in the form $\mu = \langle x_1, \dots, x_n \rangle$. If $\sigma = \langle y_1, \dots, y_m \rangle$ is another sequence, we denote the concatenation of sequences by $\mu \circ \sigma = \langle x_1, \dots, x_n, y_1, \dots, y_m \rangle$. As usual, we define exponentiation of sequences by $\mu^0 = \langle \rangle$ and $\mu^{i+1} = \mu^i \circ \mu$ ($i \geq 0$). A *continuant* is a special determinant, usually used in developing continued fractions. For $k \geq 1$ and $x_1, \dots, x_k \in \mathbb{R}$ continuants can be defined recursively by the functional equations

$$C(\langle x_1, \dots, x_k \rangle) = x_1 C(\langle x_2, \dots, x_k \rangle) - C(\langle x_3, \dots, x_k \rangle)$$

or equivalently

$$C(\langle x_1, \dots, x_k \rangle) = x_k C(\langle x_1, \dots, x_{k-1} \rangle) - C(\langle x_1, \dots, x_{k-2} \rangle)$$

with the initial conditions

$$C(\langle x_r, \dots, x_s \rangle) = \begin{cases} 1, & \text{if } s = r - 1 \\ 0, & \text{if } s < r - 1. \end{cases}$$

Mrose has shown in his dissertation [6] the following two results:

(1) $g(h, k) = \text{Max}\{C(\langle x_1, \dots, x_k \rangle) \mid x_i \in \mathbb{N}, x_i \geq 2(1 \leq i \leq k) \text{ and}$

$$\sum_{1 \leq i \leq k} x_i = h + 2k - 1\} - 1$$

(2) (i) $g(h, k) = C(\langle x_1, \dots, x_k \rangle) - 1 = C(\langle y_1, \dots, y_k \rangle) - 1$, where

$$x_i = \begin{cases} 2 + \left\lfloor \frac{ih}{k} \right\rfloor - \left\lfloor \frac{(i-1)h}{k} \right\rfloor, & \text{if } 1 \leq i < k \\ h + 1 - \left\lfloor \frac{(k-1)h}{k} \right\rfloor, & \text{if } i = k \end{cases}$$

$$y_i = \begin{cases} 1 + \left\lfloor \frac{h}{k} \right\rfloor, & \text{if } i = 1 \\ 2 + \left\lfloor \frac{ih}{k} \right\rfloor - \left\lfloor \frac{(i-1)h}{k} \right\rfloor, & \text{if } 1 < i \leq k. \end{cases}$$

(ii) If the sequences $\langle x_1, \dots, x_k \rangle$ and $\langle y_1, \dots, y_k \rangle$ are determined with respect to (2) (i), they are reversed to each other, moreover,

$$A = \{C(\langle x_1, \dots, x_i \rangle) \mid -1 \leq i < k\} \quad \text{and} \quad B = \{C(\langle y_1, \dots, y_i \rangle) \mid -1 \leq i < k\}$$

are the two uniquely determined extremal section bases with $g(h, k) = g(h, A) = g(h, B)$.

4. *r*-optimal pairs

To determine all *r*-optimal pairs, we need the values $g(z, z - 1)$ and $g(z, z)$. By *Mrose's* result we get

$$g(z, z - 1) = C(\langle 3 \rangle^{z-1}) - 1 \quad \text{and} \quad g(z, z) = C(\langle 3 \rangle^{z-1} \circ \langle 2 \rangle) - 1.$$

Using the well known approach to solve linear difference equations with initial conditions we obtain

Lemma 1: *Let $\alpha = \sqrt{5}^{-1}$, $\lambda_1 = \frac{1}{2}(3 + \sqrt{5})$ and $\lambda_2 = \frac{1}{2}(3 - \sqrt{5})$, then $g(z, z - 1) = \alpha(\lambda_1^z - \lambda_2^z) - 1$ and $g(z, z) = \alpha(\lambda_1^{z+1} - \lambda_2^{z+1}) - \alpha(\lambda_1^z - \lambda_2^z) - 1$.*

We combine this lemma and *Mrose's* results with the following theorem.

Theorem 1: *If $\langle x_1, \dots, x_k \rangle \in \mathbb{N}^k$ with $x_i \geq 2$ for all $1 \leq i \leq k$ and $s = \sum_{1 \leq i \leq k} x_i$, the following holds:*

- (i) $C(\langle x_1, \dots, x_k \rangle) \leq C\left(\left\langle \frac{s}{k} \right\rangle^k\right)$, if $s \geq 3k$,
- (ii) $C(\langle x_1, \dots, x_k \rangle) \leq \beta(1 + \beta)^{s \bmod k} \lambda^{k+1}$, if $s \bmod k < \left\lfloor \frac{k}{2} \right\rfloor$,

where the constants are given by

$$\beta = (\sqrt{b^2 - 4})^{-1}, \quad \lambda = \frac{1}{2}(b + \sqrt{b^2 - 4}) \quad \text{and} \quad b = \left\lfloor \frac{s}{k} \right\rfloor.$$

The proof of Theorem 1 is difficult, lengthy and rather technical. A detailed version is given in [1].

Combining *Mrose's* results with Theorem 1 we get

Theorem 2:

- (i) *If $k < h$ then $g(h, k) \leq \alpha(\lambda_1^{k+1} - \lambda_2^{k+1}) - 1$, where $\alpha = (\sqrt{b^2 - 4})^{-1}$, $\lambda_1 = \frac{1}{2}(b + \sqrt{b^2 - 4})$, $\lambda_2 = \frac{1}{2}(b - \sqrt{b^2 - 4})$ and $b = 2 + \frac{h}{k}$.*
- (ii) *If $k < h < \frac{3}{2}k$ then $g(h, k) \leq \beta(1 + \beta)^{h-k-1} \lambda^{k+1} - 1$, where $\beta = \sqrt{5}^{-1}$ and $\lambda = \frac{1}{2}(3 + \sqrt{5})$.*

Using this theorem we determine all *r*-optimal pairs.

Lemma 2: *If (h, k) is *r*-optimal and $h \geq 90$, then $h < \frac{3}{2}k$.*

Proof: Let (h, k) *r*-optimal with $h \geq 90$. Suppose in contradiction to the statement, that $h \geq \frac{3}{2}k$ holds. Discussing functions in the well-known way yields

- (1) $F(c) = 2.576^c - \frac{1}{2}(2 + c^2 + \sqrt{(2 + c^2)^2 - 4})$ is a monotonous function for $c \geq 1$,
- (2) $G(c) = 2.618^{90c/(90+c^2)} - \frac{1}{2}(2 + c^2 + \sqrt{(2 + c^2)^2 - 4}) \geq 0$, if $c \in [\sqrt{1.5}, \sqrt{90}]$.

Determine $\alpha, \lambda_1, \lambda_2$ as in Lemma 1. Let $\gamma = (\sqrt{b^2 - 4})^{-1}$, $\lambda = \frac{1}{2}(b + \sqrt{b^2 - 4})$ with $b = 2 + \frac{h}{k}$. If $k \geq 60$ then $\lambda_1^{k/(k+1)} \geq 2.576$. Hence we get by (1)

$$\lambda_1^{\sqrt{hk}/(k+1)} - \lambda = \lambda_1^{k/(k+1)\sqrt{h/k}} - \lambda \geq F\left(\sqrt{\frac{h}{k}}\right) \geq F(\sqrt{1.5}) \geq 0.$$

If $k < 60$ we use (2) to obtain

$$\lambda_1^{\sqrt{hk}/(k+1)} - \lambda = \lambda_1^{h\sqrt{h/k}/(h+h/k)} - \lambda \geq G\left(\sqrt{\frac{h}{k}}\right) \geq 0.$$

Observe that in this case $\frac{h}{k} \in [1.5, 90]$. In total we conclude

(3) $\lambda_1^{\sqrt{hk}} - \lambda^{k+1} \geq 0.$

Now, consider $z = \text{Max}\{m \in \mathbb{N} \mid hk > m(m - 1)\}$. By definition $hk \leq (z + 1)z$.

First case: $z(z - 1) < hk \leq z^2$.

By Lemma 1 and Theorem 2 we obtain

$$\begin{aligned} g(z, z - 1) - g(h, k) &\geq \alpha(\lambda_1^z - \lambda_2^z) - \gamma\lambda^{k+1} \\ &= \gamma\left[\frac{\alpha}{\gamma}\left(1 - \left(\frac{\lambda_2}{\lambda_1}\right)^z\right)\lambda_1^z - \lambda^{k+1}\right] \\ &\geq \gamma[\lambda_1^z - \lambda^{k+1}] \geq \gamma[\lambda_1^{\sqrt{hk}} - \lambda^{k+1}]. \end{aligned}$$

By this, together with (3): $g(z, z - 1) \geq g(h, k)$. Since (h, k) is r -optimal and $hk > z(z - 1)$, this is a contradiction.

Second case: $z^2 < hk \leq (z + 1)z$.

Now, $hk \leq \left(z + \frac{1}{2}\right)^2$, and by Lemma 1 and Theorem 2 we get

$$\begin{aligned} g(z, z) - g(h, k) &\geq \alpha(\lambda_1^{z+1} - \lambda_2^{z+1}) - \alpha(\lambda_1^z - \lambda_2^z) - \gamma\lambda^{k+1} \\ &\geq \alpha\lambda_1^z(\lambda_1 - 1) - \gamma\lambda^{k+1} \\ &\geq \gamma\left[\frac{\alpha}{\gamma}(\lambda_1 - 1)\lambda_1^{-0.5+\sqrt{hk}} - \lambda^{k+1}\right] \\ &\geq \gamma[\lambda_1^{\sqrt{hk}} - \lambda^{k+1}]. \end{aligned}$$

Hence, $g(z, z) \geq g(h, k)$ by (3). Since $z^2 \leq hk$, we reach again a contradiction to the r -optimality of (h, k) .

Lemma 3: For any r -optimal (h, k) with $h \leq \frac{3}{2}k$ the following holds:

- (i) $h \leq k + 5,$
- (ii) $(h, k) \in \begin{cases} \{(z, z), (z + 1, z - 1), (z + 2, z - 2)\}, & \text{if } z \in \mathbb{N} \text{ and } z(z - 1) < hk < z^2 \\ \{(z + 1, z), (z + 2, z - 1), (z + 3, z - 2)\}, & \text{if } z \in \mathbb{N} \text{ and } z^2 < hk < (z + 1)z. \end{cases}$

Proof: If (h, k) is r -optimal with $h \leq \frac{3}{2}k$, consider $z = \text{Max}\{m \in \mathbb{N} \mid hk > m(m - 1)\}$. Obviously, $hk \leq (z + 1)z$. Let $h = z + x$ and $k = z - y$. Consider again the constants $\alpha, \lambda_1, \lambda_2$ of lemma 1. Then we get

$$(1) \quad g(z, z - 1) \geq \alpha \lambda_1^z \left(1 - \left(\frac{\lambda_2}{\lambda_1} \right)^2 \right) - 1 \text{ and}$$

$$(2) \quad g(z, z) \geq \alpha \lambda_1^z (\lambda_1 - 1) - 1.$$

Furthermore Theorem 2 implies

$$(3) \quad g(z + x, z - y) \leq \alpha \lambda_1^z (1 + \alpha)^x \left(\frac{1 + \alpha}{\lambda_1} \right)^{y-1} - 1.$$

First case: $z(z - 1) < hk \leq z^2$.

If $z = 1$, we get $h = k = z$. Thus, without loss of generality, $z \geq 2$. It is easy to show

$$(4) \quad 0 \leq y \leq x < \sqrt{1.5}y.$$

Now, on the right hand side of (3), we can use $x < \sqrt{1.5}y$ and compare the resulting expression with the right hand side of (1). Doing this we obtain $g(z, z - 1) \geq g(z + x, z - y)$ for $y \geq 5$.

If $y \leq 4$, (4) implies $0 \leq x - y \leq (\sqrt{1.5} - 1)y \leq 0.9$ and therefore $x = y$. Replacing x on the right hand side of (3) by (4) and comparing again the right hand sides of (1) and (3) results in

$$g(z, z - 1) \geq g(z + x, z - y) \quad \text{for } y \in \{3, 4\}.$$

Since (h, k) is r -optimal and $z(z - 1) < hk$, we obtain $g(z, z - 1) < g(z + x, z - y)$. But this can only be the case for $x = y \in \{0, 1, 2\}$.

Second case: $z^2 < hk \leq (z + 1)z$.

Analogously to the first case, we get

$$(5) \quad 0 \leq y < x \leq \sqrt{1.5}y + 0.5(1 + \sqrt{1.5}).$$

Now, compare the right hand sides of (2) and (3). If we replace x by the upper bound $\sqrt{1.5}y + 0.5(1 + \sqrt{1.5})$, we obtain $g(z, z) \geq g(z + x, x - y)$ for $y \geq 7$.

If $y \leq 6$, (5) implies

$$1 \leq x - y \leq \begin{cases} 1, & \text{if } 0 \leq y \leq 3 \\ 2, & \text{if } 4 \leq y \leq 6. \end{cases}$$

Replacing x in the right hand side of (3) by $y + 1$ respectively $y + 2$ and comparing again with the right hand side of (2), we get $g(z, z) \geq g(z + x, z - y)$, if $y = 3$ respectively if $4 \leq y \leq 6$.

Since $z^2 < hk$, the only remaining possibilities are $(x, y) \in \{(1, 0), (2, 1), (3, 2)\}$ to fetch an r -optimal pair (h, k) . ■

Lemma 4:

- (i) $g(k + 2, k + 2) < g(k + 5, k)$, if $k \geq 8$,
- (ii) $g(k + 2, k + 1) < g(k + 4, k)$, if $k \geq 3$,
- (iii) $g(k + 1, k + 1) < g(k + 3, k)$, if $k \geq 3$,
- (iv) $g(h, k) < g(h - 1, k + 1)$, if $h > k + 1$.

Proof: A simple induction proof (see Theorem 1.2 of [6]) shows for all $0 \leq i < k$ and $\langle x_1, \dots, x_k \rangle \in \mathbb{R}^k$ that

$$(1) \quad C(\langle x_1, \dots, x_k \rangle) = C(\langle x_1, \dots, x_i \rangle) \cdot C(\langle x_{i+1}, \dots, x_k \rangle) \\ - C(\langle x_1, \dots, x_{i-1} \rangle) \cdot C(\langle x_{i+2}, \dots, x_k \rangle).$$

By this we get

$$(2) \quad C(\langle x_1, \dots, x_k \rangle) = C(\langle x_1, \dots, x_{i-1}, x_i - 1, x_{i+1}, \dots, x_k \rangle) \\ + C(\langle x_1, \dots, x_{i-1} \rangle) \cdot C(\langle x_{i+1}, \dots, x_k \rangle).$$

If $k \geq 8$, $a_1 = C(\langle 3 \rangle^{k-8})$ and $a_0 = C(\langle 3 \rangle^{k-9})$, we obtain from (1) and by *Mrose's* algorithm to determine $g(h, k)$ that

$$\begin{aligned} g(k+2, k+2) &= C(\langle 3 \rangle^{k+1} \circ \langle 2 \rangle) - 1 \\ &= a_1 C(\langle 3 \rangle^9 \circ \langle 2 \rangle) - a_0 C(\langle 3 \rangle^8 \circ \langle 2 \rangle) - 1 \\ &= 10946a_1 - 4181a_0 - 1 < 10981a_1 - 4030a_0 - 1 \\ &= a_1 C(\langle 3, 4 \rangle^2 \circ \langle 4, 3 \rangle^2) - a_0 C(\langle 4, 3, 4 \rangle \circ \langle 4, 3 \rangle^2) - 1 \\ &= C(\langle 3 \rangle^{k-8} \circ \langle 3, 4 \rangle^2 \circ \langle 4, 3 \rangle^2) - 1 \\ &\leq g(k+5, k) \end{aligned}$$

proving (1).

Turning to (ii) and (iii) we can argue analogously. We show

$$g(k+2, k+1) = C(\langle 3 \rangle^{k+1}) - 1 \leq C(\langle 3 \rangle^{k-3} \circ \langle 4 \rangle^3) - 1 \leq g(k+4, k),$$

respectively,

$$g(k+1, k+1) = C(\langle 3 \rangle^k \circ \langle 2 \rangle) - 1 < C(\langle 3 \rangle^{k-2} \circ \langle 4 \rangle^2) - 1 \leq g(k+3, k).$$

It remains to show (iv). If $h > k+1$ and $g(h, k) = C(\langle x_1, \dots, x_k \rangle) - 1$, where x_1, \dots, x_k are determined by *Mrose's* algorithm, there exists a $b \geq 3$, such that $x_i \in \{b, b+1\}$ for $1 \leq i \leq k$. Let $j = \text{Max}\{i/1 \leq i \leq k \text{ and } x_i \geq 4\}$, $\mu = \langle x_1, \dots, x_{j-1} \rangle$ and $\mu' = \langle x_1, \dots, x_{j-2} \rangle$.

By (1) we obtain first

$$\begin{aligned} &C(\mu \circ \langle b \rangle^{k-j+1}) - C(\mu \circ \langle b \rangle^{k-j}) \\ &= C(\mu) C(\langle b \rangle^{k-j+1}) - C(\mu') \cdot C(\langle b \rangle^{k-j}) \\ &\quad - C(\mu) C(\langle b \rangle^{k-j}) + C(\mu') C(\langle b \rangle^{k-j-1}) \\ &= C(\mu) [(b-1) C(\langle b \rangle^{k-j}) - C(\langle b \rangle^{k-j-1})] \\ &\quad - C(\mu') [C(\langle b \rangle^{k-j}) - C(\langle b \rangle^{k-j-1})] \\ &> C(\mu) C(\langle b \rangle^{k-j}) \end{aligned}$$

and from this together with (2)

$$\begin{aligned} g(h-1, k+1) &\geq C(\langle x_1, \dots, x_{j-1}, x_j - 1, x_{j+1}, \dots, x_k, 2 \rangle) - 1 \\ &= 2C(\mu \circ \langle b \rangle^{k-j+1}) - C(\mu \circ \langle b \rangle^{k-j}) - 1 \\ &> C(\mu \circ \langle b \rangle^{k-j+1}) + C(\mu) C(\langle b \rangle^{k-j}) - 1 \\ &= C(\mu \circ \langle b+1 \rangle \circ \langle b \rangle^{k-j}) - 1 = C(\langle x_1, \dots, x_k \rangle) - 1 \\ &= g(h, k). \quad \blacksquare \end{aligned}$$

Lemma 5: If $(h, k), (r, s) \in \mathbb{N}^2$ with $k \geq 9, |h - k| \leq 5, |r - s| \leq 5$ and $rs < hk$, then $g(r, s) < g(h, k)$.

Proof: Consider such $(h, k), (r, s) \in \mathbb{N}^2$.

First case: $s = k$

Then $r < h$ and our statement follows immediately, since $g(h, k)$ is monotonously increasing in both arguments.

Second case: $s = k - j$ with $j \geq 1$.

Let $r = h + j + i$, then $h - k + 2j + i = r - s \leq 5$. Since $rs < hk$, we get

$$0 < hk - rs = -ik + (h - k)j + j(2j + i) - j^2 = -ik + j(r - s) - j^2 \leq -ik + 5j - j^2 \leq -ik + 6$$

resulting in $i < k$, because $k \geq 9$. Hence, $r \leq h + j$ and therefore $g(r, s) \leq g(k - j, h + j) < g(k, h)$ by Lemma 4.

Third case: $s = k + j$ with $j \geq 1$.

Let $r = k + i$. Since $r \geq s$ and $rs < hk$, we get $i \geq j \geq 1$ and $2 \leq i + j < h - k \leq 5$. In this case, only three possibilities remain:

- $(i, j) = (1, 1)$ and $h = k + 3$,
- $(i, j) \in \{(1, 1), (2, 1)\}$ and $h = k + 4$,
- $(i, j) \in \{(1, 1), (2, 1), (3, 1), (2, 2)\}$ and $h = k + 5$.

By the monotonicity of g and Lemma 4 we get in all cases $g(r, s) < g(h, k)$. ■

Now, we can summarize all the results.

Theorem 3: (h, k) is r -optimal if and only if $h - k \leq 5$ with $(h > 18$ or $k > 10)$ or (h, k) is marked in the following table with “x”.

k/h	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	x	x	x	x	x	x	.	x
2		x	x	x	x	x	x	.	x
3			x	x	x	x	x	x	x
4				x	x	x	x	x	x	.	x
5					x	x	x	x	x	.	.	x	.	x
6						x	x	x	x	.	x
7							x	x	x	x	.	.	x
8								x	x	x	x	x	.	x	x
9									x	x	x	x	x	.	.	x
10										x	x	x	x	x	x	.	.	x
11											x	x	x	x	x	x
12												x	x	x	x	x	x
13													x	x	x	x	x	x
14														x	x	x	x	x	x
15															x	x	x	x	x
16																x	x	x	x	x	.	.	.

Proof:

“ \Rightarrow ”: Let (h, k) be r -optimal. If $h \geq 90$ or $h \leq \frac{3}{2}k$, we get $h - k \leq 5$ by Lemma 3 and 4. The remaining finitely many cases are checked with a computer program.

“ \Leftarrow ”: We check the table with a computer program. It remains to show the r -optimality of (h, k) under the conditions $|h - k| \leq 5$ and $h > 18$ or $k > 10$. Suppose, such an (h, k) is not r -optimal, i.e. there is some r -optimal pair (r, s) with $rs < hk$ and $g(r, s) \geq g(h, k)$. By Lemma 5 $r - s \geq 6$, moreover, the first part of the proof asserts $r \leq 18$ and $s \leq 10$ contradicting $g(r, s) \geq g(h, k)$. ■

5. $m(x)$ -optimal bases

Once the r -optimal pairs have been found, it is not difficult to determine $m(x)$ and the $m(x)$ -optimal bases. Consider $A, B \subseteq \mathbb{N}$. We call B a *strict monotonous extension* of A if $A \subsetneq B$ and $\text{Max}(A) < \text{Min}(B - A)$. B is a *monotonous extension* of A , if either $A = B$ or B is a strict monotonous extension of A .

Theorem 4: *Let $x > 215268$, then the following holds:*

- (i) $m(x) = \text{Min}\{hk/g(h, k) \geq x \text{ and } |h - k| \leq 5\}$;
- (ii) *the $m(x)$ -optimal bases can be computed with the help of the following algorithm:*

1. $z := \text{Max}\{m \in \mathbb{N}/g(m, m - 1) < x\}$;
 if $x \leq g(z, z)$ then $t := 0$ else $t := 1$;
 $i := \text{Max}\{j \in \{0, 1, 2\}/g(z + t + j, z - j) \geq x\}$;
 $h := z + t + i$;
 $k := z - i$;
2. *Use Mrose’s algorithm to determine the extremal regular section bases $A_i, B_i \subseteq \mathbb{N}_0$ for $i = 1, 2$ with $|A_i| = k, |B_i| = h$ and $g(h, k) = g(h, A_i) = g(k, B_i)$. Then the $m(x)$ -optimal bases are exactly the monotonous extensions of A_i and B_i ($i = 1, 2$).*

Proof: Consider $x > 215268 = g(18, 10)$ and $(h, k) \in \mathbb{N}^2$ with $m(x) = hk$ and $g(h, k) \geq x$. By symmetry of g we can assume $h \geq k$, i.e. (h, k) is r -optimal. Since g is monotonous, we get $h > 18$ or $k > 10$. Now, Theorem 3 yields $h - k \leq 5$ and (i) is proven.

Let $z = \text{Max}\{m \in \mathbb{N}/g(m, m - 1) < x\}$.

First case: $x \leq g(z, z)$. Since $(z, z - 1)$ and (h, k) are r -optimal, we get $z(z - 1) < hk \leq z^2$ and Lemma 3 implies $(h, k) \in \{(z + 1 + j, z - j)/j = 0, 1, 2\}$.

Second case: $g(z, z) < x$. In this case $x \leq g(z + 1, z)$. r -optimality of (z, z) and (h, k) implies $z^2 < hk \leq (z + 1)z$. Hence $(h, k) \in \{(z + 1 + j, z - j)/j = 0, 1, 2\}$ by Lemma 3.

Since the values $(z + j)(z - j)$ and $(z + 1 + j)(z - j)$ are strictly decreasing with increasing j , we obtain in the first case $(h, k) = (z + j, z + j)$ for the maximal $j \in \{0, 1, 2\}$ with $g(z + j, z - j) \geq x$ and in the second case $(h, k) = (z + 1 + j, z - j)$ for the maximal $j \in \{0, 1, 2\}$ with $g(z + 1 + j, z - j) \geq x$.

Consequently, (h, k) is uniquely determined and the algorithm computes this (h, k) . Now, the rest of the proof follows immediately from the definition of $m(x)$ -optimal sets and the symmetry of g . ■

Should the range $[1 \dots x]$ be represented optimally with respect to our measure, the above theorem and Mrose’s algorithm tell us, that the optimal base has to be changed under complete new recomputations, if x increases. This seems to be not adequate for practical purposes. Therefore we are interested into *infinite* bases A_∞ , such that the product $S(x, A_\infty) \cdot A_\infty(x)$ is asymptotically equal to $m(x)$.

We can find such bases in the following way. Let $FIBG = \{a_{2i}/i \geq 1\} = \{1, 3, 8, 21 \dots\}$ and $FIBU = \{a_{2i+1} \mid i \geq 0\} = \{1, 2, 5, 18 \dots\}$ be the sets of Fibonacci numbers with even respectively with odd index.

Theorem 5:

$$m(x) \sim S(x, FIBG) \cdot FIBG(x) \sim S(x, FIBU) \cdot FIBU(x) \sim c \cdot \ln^2(x),$$

where $c = [\ln(\frac{1}{2}(3 + \sqrt{5}))]^{-2} = 1.07961 \dots$. Moreover $FIBG$ and $FIBU$ are the only bases, such that for infinitely many x

$$m(x) = S(x, FIBG) \cdot FIBG(x) \quad \text{respectively} \quad m(x) = S(x, FIBU) \cdot FIBU(x).$$

Proof: It is easy to show that

$$FIBG = \{C(\langle 3 \rangle^z) \mid z \geq 0\} \text{ and } FIBU = \{C(\langle 2 \rangle \circ \langle 3 \rangle^z) \mid z \geq 0\}.$$

By *Mrose's* algorithm we get $g(z, z - 1) = C(\langle 3 \rangle^{z-1}) - 1$ and $g(z, z) = C(\langle 2 \rangle \circ \langle 3 \rangle^{z-1}) - 1$.

Let $x = g(z, z - 1)$ and $y = g(z, z)$, then $m(x) = z(z - 1) = S(x, FIBG) \cdot FIBG(x)$ and $m(y) = z^2 = S(y, FIBU) \cdot FIBU(y)$. Hence, $FIBG$ and $FIBU$ are $m(x)$ -optimal bases for infinitely many x .

Consider α and λ_1 from Lemma 1, then:

$$\begin{aligned} m(x) &\leq S(x, FIBU) \cdot FIBU(x) \leq \text{Min}\{z^2/g(z, z) \geq x\} \\ &\leq (\text{Min}\{z/\alpha\lambda_1^z(\lambda_1 - 1) - 1 \geq x\})^2 \leq \lceil c \cdot \ln(x + 1) + \tilde{c} \rceil^2, \end{aligned}$$

where $c = \left[\ln \left(\frac{3 + \sqrt{5}}{2} \right) \right]^{-2}$ and $\tilde{c} = c \cdot \ln \left(\frac{2 + \sqrt{5}}{1 + \sqrt{5}} \right)$.

On the other hand by Theorem 4

$$\begin{aligned} m(x) &= \text{Min}\{(z + t + j)(z - j)/t \in \{0, 1\}, j \in \{0, 1, 2\} \text{ and } g(z + t + j, z - j) \geq x\} \\ &\geq \text{Min}\{z^2 - 4/g(z + 1, z) \geq x\} \geq \lceil c \cdot \ln(x + 1) + \tilde{c} - 1 \rceil^2 - 4. \end{aligned}$$

Since lower and upper bound of $m(x)$ are asymptotically equal to $c \cdot \ln^2(x)$, we get

$$m(x) \sim S(x, FIBU) \cdot FIBU(x) \sim c \cdot \ln^2(x).$$

Analogously (again by Lemma 1),

$$S(x, FIBG) \cdot FIBG(x) \leq \text{Min}\{z(z - 1) \mid g(z, z - 1) \geq x\},$$

implies

$$S(x, FIBG) \cdot FIBG(x) \sim c \cdot \ln^2(x).$$

Now, consider a base A with $m(x) = S(x, A) \cdot A(x)$ for infinitely many x .

By Theorem 4, there is an $m \in [-5 \dots 5]$ and an infinite sequence of extremal regular section bases $A_1, A_2 \dots$ with $|A_i - \{0\}| = k_i$ ($i \geq 1$), such that for all $i \geq 1$

- $\bigcup_{i \in \mathbb{N}} A_i - \{0\} = A$,
- A_{i+1} is a strictly monotonous extension of A_i and
- $g(k_i + m, k_i) = g(k_i + m, A_i)$.

By *Mrose's* algorithm to compute $g(h, k)$, there must exist a sequence x_1, x_2, \dots of natural numbers, such that for all $i \geq 1$ the following holds:

- (i) $\{C(\langle x_1, \dots, x_\lambda \rangle) / -1 \leq \lambda < k_i\} = A_i$ and
- (ii) either $x_\lambda = 2 + \left\lfloor \frac{\lambda(k_i + m)}{k_i} \right\rfloor - \left\lfloor \frac{(\lambda - 1)(k_i + m)}{k_i} \right\rfloor$ for $1 \leq \lambda < k_i$
 or $x_\lambda = \begin{cases} 1 + \left\lfloor \frac{k_i + m}{k_i} \right\rfloor, & \text{if } \lambda = 1 \\ 2 + \left\lfloor \frac{\lambda(k_i + m)}{k_i} \right\rfloor - \left\lfloor \frac{(\lambda - 1)(k_i + m)}{k_i} \right\rfloor, & \text{if } 1 < \lambda \leq k_i. \end{cases}$

By (ii) $x_1 \in \{2, 3\}$ and $x_\lambda = 3$ for $\lambda \geq 2$, since for any i there exists a $j > i$ such that $\frac{\lambda m}{k_j} < 1$ for $1 \leq \lambda \leq k_i$.

Therefore

$$A = \bigcup_{i \in \mathbb{N}} A_i - \{0\} \in \{FIBG, FIBU\},$$

which shows, that there is no other infinite base A_∞ than *FIBG* and *FIBU* such that $m(x) = S(x, A_\infty) \cdot A_\infty(x)$ for infinitely many x . ■

We compare this result with the geometric bases $GEO_a = \{a^i \mid i \in \mathbb{N}_0\}$ for $a \geq 2$. One easily checks that

$$S(x, GEO_a) \cdot GEO_a(x) \sim \frac{a - 1}{\ln^2 a} \cdot \ln^2 x,$$

where the constant $c_a = \frac{a - 1}{\ln^2 a}$ adopts the following values for $a \in [2 \dots 10]$:

a	2	3	4	5	6	7	8	9	10
c_a	2.08 ...	1.65 ...	1.56 ...	1.54 ...	1.55 ...	1.58 ...	1.61 ...	1.65 ...	1.69 ...

One can show, that c_a is strictly increasing for $c_a \geq 10$, so that c_a is minimal for $a = 5$.

6 Concluding remarks

We discuss shortly arithmetical and logical operations. The main property in using $m(x)$ optimal bases, is the fact, that indeed the coefficients in the representation of a number are digits. Using for example *FIBG* of *FIBU* these digits are 0, 1, 2. Consequently, the logical operations $<$, \leq , $>$, \geq , $=$, \neq can be performed in the usual way.

Things are different with arithmetical operations. Consider *FIBG* and addition. In contrast to GEO_a the carry will be transferred both to the left and right. The digit addition of 2 and 1 respectively 2 results in the digit 0 respectively 1, where the carry 1 wents in both cases to the left as well as to the right.

Table of the first values of $g(h, k)$

$k/h:$	1	2	3	4	5	6	7	8	9	10	11	12	13
1:	1	2	3	4	5	6	7	8	9	10	11	12	13
2:		4	7	10	14	18	23	28	34	40	47	54	62
3:			12	20	29	40	55	71	90	114	139	168	203
4:				33	54	78	111	152	208	286	344	435	550
5:					88	143	207	296	417	570	779	1004	1291
6:						232	376	544	781	1108	1559	2130	2910
7:							609	986	1427	2058	2939	4140	5821
8:								1596	2583	3738	5399	7740	10980
9:									4180	6764	9789	14148	20381
10:										10945	17710	25630	37072
11:											28656	46367	67103
12:												75024	121392
13:													196417
14:													

The values of r -optimal pairs (h, k) are bold.

Table of the first values of $g(h, k)$ (continued)

k/h :	14	15	16	17	18	19	20	21	22	23
1:	14	15	16	17	18	19	20	21	22	23
2:	70	79	88	98	108	119	130	142	154	167
3:	239	280	328	377	432	495	559	630	710	791
4:	670	815	984	1188	1398	1644	1925	2254	2590	2975
5:	1652	2088	2639	3215	3914	4755	5740	6929	8154	9593
6:	3750	4826	6190	7919	10008	12648	15408	18763	22818	27719
7:	7952	10863	13999	18033	23183	29665	37946	47955	60604	73829
8:	15454	21727	29680	40544	52248	67319	86616	111096	142138	181814
9:	29102	41002	57680	81089	110770	151315	194996	251260	323606	416019
10:	53454	76627	108700	153104	215268	302631	413402	564718	727738	937790
11:	97082	140191	201760	288089	405999	571427	803396	1129437	1542840	2107559
12:	175680	254196	367328	529154	758538	1076031	1516001	2132677	2998318	4215119
13:	317810	459939	665575	962469	1387778	1997227	2851796	4019015	5660743	7959559
14:	514228	832039	1204138	1742564	2520418	3639591	5238094	7508761	10651618	15011018
15:		1346268	2178308	3152477	4562173	6600215	9536404	13737726	19770518	28229879
16:			3524577	5702886	8253294	11944134	17281656	24987146	36028798	51851799
17:				9227464	14930351	21607407	31270319	45249377	65433823	94489199
18:					24157816	39088168	56568928	81866968	118468788	171351439
19:						63245985	102334154	148099379	214331052	310166779
20:							165580140	267914295	387729210	561126422
21:								433494436	701408732	1015088253
22:									1134903169	1836311902
23:										2971215072

The values of r -optimal pairs (h, k) are bold.

A more serious problem as the behaviour of the carry stems from the fact, that in regular representations there must be always a 0 between two 2's. It is not difficult to get results containing only the digits 0, 1, 2 when adding in *FIBG* in a serial manner, but these results may not be regular. Therefore we are unable to present a real-time serial algorithm for addition. *C. Frougny* [3] investigates addition and conversion into regular representation by finite state transducers when numbers are represented to bases related to the Fibonacci numbers. In particular she gives a linear-time serial algorithm for *FIBG* (see Example 6.9 in [3]).

But with respect to computer design parallel algorithms are of more interest. Look at a *von Neumann* design, where digits and carries are treated simultaneously as long as there is still a carry. If we assume the usual binary system, it is well-known, that the expected time (counted in steps) is less or equal to $\log_2 n + 1$, where n is the operand length, if all data cases are of equal probability (see [2]). It is possible to do the same in the case *FIBG*, but the resulting representation may not be regular again, i.e. the condition that between two 2's there is always a 0 is not fulfilled necessarily.

Things are different looking at subtraction. Performing $x - y$, we get a regular representation even if we start with a non-regular representation of y . Many computers use complementation to reduce subtraction to addition. The complement of a number z represented in a register of length n to a geometric base A is obtained by subtracting z from the $(n + 1)$ -th element of A . Instead of subtracting z all can be done by addition of the complement. In the case of *FIBG* it seems reasonable to use the converse way reducing addition to subtraction. This procedure will work, however, providing the complement still needs, as far as it is known, a full subtraction.

Appendix:

Table of the first values of $g(h, k)$. Values of r -optimal pairs (h, k) are bold.

Acknowledgement:

The results of this paper are part of the authors Habilitationsschrift, accepted by the Fachbereich Informatik of the Technische Hochschule Darmstadt.

References

- [1] *Brandt, U.*: Abschätzungen für Kontinuanten. To appear in Mainzer Seminarberichte, edited by M. Djawadi, E. Härtter, G. Hofmeister, Johannes Gutenberg-Universität Mainz, Fachbereich Mathematik, or as technical report: TI-2/89, Technische Hochschule Darmstadt, Fachbereich Informatik.
- [2] *Claus, V.*: Die mittlere Additionsdauer eines Parallelladdierwerks. *Acta Inform.* **2** (1973), 283 – 291.
- [3] *Frougny, C.*: Systèmes de numération linéaires et automates finis. Thèse d'Etat, Université Paris 6, L.I.T.P., 1989.
- [4] *Hofmeister, G.*: Über eine Menge von Abschnittsbasen. *Reine Angew. Math.* **213** (1963), 43 – 57.

- [5] *Hofmeister, G.*: Einige gelöste und ungelöste Probleme aus der Zahlentheorie. In: Zahlen, Codes und Computer. Hrsg.: Arbeitsgruppe für Lehrerfortbildung anlässlich einer Tagung zur Lehrerfortbildung vom 04.–05. 10. 1984, Johannes Gutenberg-Universität Mainz, Fachbereich Mathematik, 1985, 1–29.
- [6] *Mrose, A.*: Die Bestimmung der extremalen regulären Abschnittsbasen mit Hilfe einer Klasse von Kettenbruchdeterminanten. Dissertation, Freie Universität Berlin, 1969.
- [7] *Rohrbach, H.*: Ein Beitrag zur additiven Zahlentheorie. *Math. Z.* **42** (1937), 1–30.
- [8] *Spaniol, O.*: Arithmetik in Rechenanlagen. Teubner Studienbücher Informatik 1976.

Zusammenfassung

Es stellt sich die Frage, ob die in der Regel verwendeten geometrischen Zahlensysteme am besten geeignet sind, um Zahlen additiv in Registern darzustellen. Die Antwort hängt selbstverständlich vom Beurteilungskriterium ab. In dieser Arbeit wird ein Produktmaß gewählt, das sowohl die vorzusehende Registerlänge als auch den Energieumsatz bei bestimmten Standardoperationen berücksichtigt. Man erhält dann die Fibonacci-Zahlen mit geradem bzw. ungeradem Index als optimale unendliche Basen. Dabei werden spezielle Ergebnisse aus der additiven Zahlentheorie sowie Majoranten-theoreme für Kontinuanten verwendet.

(Received: September 12, 1991)

Author's address:

U. Brandt
Institut für Theoretische Informatik
Fachbereich Informatik
Technische Hochschule Darmstadt
W-6100 Darmstadt
Frankfurter Str. 69A
Germany