Approximating Addition by XOR: how to go all the way

Didier Alquié
DGA, CELAR
didier.alquie@laposte.net

March 12, 2010

Abstract

In this paper, we study approximation of addition by XOR, taking P. Sarkar's publication [1] as the reference work and starting point. In this work, among various results, it was claimed that explicit formulas seemed difficult to obtain when the number n of summands is more than 5. In the first part of our work, we show a systematic way to find explicit formulas: the complexity to compute them is $O(n^3)$, which allows large values of n. We present some numerical computation and point out a - conjectural - observation on the coefficients.

In the second part, we study a generalization of P. Sarkar's work to q-ary addition, instead of binary. We show that the mechanics of the addition is essentially the same as in the binary case. In particular, sequence of carries behaves very similarly: it is a Markov chain whose transition matrix can be computed. Running some experiments on small values of n leads us to a conjecture, the first part of which is intuitive and the second part of which reveals an amazing coincidence (and is probably not!).

Finally, in a section titled "very last news", we refer to a paper published by Holte in 1997, that was brought to us after our first post and that we had missed before. It happens that this paper studies the topic and solves a major part of our open problems. Henceforth, the present post is an updated version of our previous "Approximating Addition by XOR: how to go (a little) further than P. Sarkar", taking into account this previous Holte's reference.

1 Cryptographic motivation

As recalled in [1], cryptographic algorithm designers are particularly interested in non linear functions that are easy to compute. Arithmetic addition is a good candidate for them. For example its algebraic degree grows rapidly with respect to the bits of inputs. Famous algorithms, such as SHA-1 or SHA-2 involve arithmetic addition of more than 2 terms (up to 7 terms actually).
General analysis of hash functions use approximation of addition by XOR, setting that the probability that both results are equal is 1/2. P. Sarkar's work, for which we give some further result, essentially says that this approximation is asymptotically the right one. More precisely:

- when the number n of summands is fixed, the probability that the i-th bit of addition and XOR are equal has a limit when $i \to +\infty$, equal to $1/2$ if n is even, and to $1/2 + (-1)^{(n-1)/2}\varepsilon_n$ for odd n;
- $\varepsilon_n \to 0$ as $n \to +\infty$.

In other words, replacing addition by XOR, you are doing a good approximation unless there are "few" odd summands.

2 P. Sarkar's results

In this section, we recall the notations and results obtained by P. Sarkar [1], who himself refers a paper by Meier and Staffelbach [2].

2.1 Notations

Let $X^{(1)}, \ldots, X^{(n)}$ be integers, and $X_i^{(k)}$ is the i-th bit of the binary expansion of $X^{(k)}$. 0-th bit is the least significant bit. Arithmetic addition on integers is denoted by $+$ while XOR is denoted by $\oplus$. Following P. Sarkar's notations, we define, for $i \geq 0$:

- $L_i^{(n)} = X_i^{(1)} \oplus \ldots \oplus X_i^{(n)}$;
- $S_i^{(n)}$ is the i-th bit of $X^{(1)} + \cdots + X^{(n)}$;
- $B_i^{(n)} = X_i^{(1)} + \cdots + X_i^{(n)}$;
- $A_i^{(n)}$ is the carry output by the i-th column of the addition. It is convenient to set $A_{-1}^{(n)} = 0$.

The following draw is the traditional "young child" representation of the arithmetic addition:

$$\begin{array}{cccccccl}
 & \ldots & A_i^{(n)} & \ldots & A_0^{(n)} & & & Carry \\
 & \ldots & X_i^{(1)} & \ldots & X_1^{(1)} & X_0^{(1)} & & Summand\#1 \\
+ & \ldots & X_i^{(2)} & \ldots & X_1^{(2)} & X_0^{(2)} & & Summand\#2 \\
+ & \ldots & & & & & & \ldots \\
+ & \ldots & X_i^{(n)} & \ldots & X_1^{(n)} & X_0^{(n)} & & Summand\#n \\
\hline
 & \ldots & S_i^{(n)} & \ldots & S_1^{(n)} & S_0^{(n)} &
\end{array}$$

while the following is a similar representation for binary addition. Of course, addition is columnwise and no carry appears.

$$\begin{array}{cccccccl}
 & \ldots & X_i^{(1)} & \ldots & X_1^{(1)} & X_0^{(1)} & & Summand\#1 \\
\oplus & \ldots & X_i^{(2)} & \ldots & X_1^{(2)} & X_0^{(2)} & & Summand\#2 \\
\oplus & \ldots & & & & & & \ldots \\
\oplus & \ldots & X_i^{(n)} & \ldots & X_1^{(n)} & X_0^{(n)} & & Summand\#n \\
\hline
 & \ldots & L_i^{(n)} & \ldots & L_1^{(n)} & L_0^{(n)} &
\end{array}$$

2.2 Medley of results of [1] and comments

Lemma 1 *For $i \geq 0$, $S_i^{(n)} = (A_{i-1}^{(n)} + B_i^{(n)}) \bmod 2$ and $A_i^{(n)} = (A_{i-1}^{(n)} + B_i^{(n)}) \div 2$.*

Comment The two properties are easy to feel when you consider them together and think of how a column addition is performed. Indeed, when you process column i, you first have to add up all symbols $X_i^{(1)}, \ldots, X_i^{(n)}$ and the carry $A_{i-1}^{(n)}$ coming from the previous column, obtaining the intermediate value $v = A_{i-1}^{(n)} + B_i^{(n)}$. Then, young pupils say you "put" something and "retain" something else. Actually, what you "put" is v mod 2, and is $S_i^{(n)}$, the i-th symbol of the result, and what you "retain" is $v \div 2$, and is the i-th carry. □

Note that the number of summands n may be more than 2, so that the current carry may exceed 1. In fact, we have (the proof is by induction):

Lemma 2 *For $i \geq -1$, $A_i^{(n)} \in [0, n-1]$.*

The main concern is to study the correlation between $L_i^{(n)}$ and $S_i^{(n)}$, that is the probability $\gamma_i^{(n)} = \Pr(L_i^{(n)} = S_i^{(n)})$. In [1], it is established that:

Lemma 3

$$\gamma_i^{(n)} = Pr(A_{i-1}^{(n)} \text{ is even}) = \sum_{even\ s} Pr(A_{i-1}^{(n)} = s)$$

and

Theorem 1 *For fixed n, the sequence of random variables $(A_i^{(n)})_{i \geq -1}$ form a homogeneous discrete-time finite-state Markov chain, with:*

- *state space equal to $\{0, \ldots, n-1\}$;*
- *transition matrix P_n, where, for $0 \leq s, t \leq n-1$,*

$$P_n(s,t) = Pr(A_i^{(n)} = t | A_{i-1}^{(n)} = s) = 2^{-n} \binom{n+1}{2t-s+1}.$$

The proof of the second result makes use of the easy observations on $B_i^{(n)}$: $B_i^{(n)} = X_i^{(1)} + \cdots + X_i^{(n)}$ is an integer lying in the interval $[0, n]$. Because it is the sum of independant random variables satisfying Bernoulli's law with parameter 1/2 (unbiased bits), $B_i^{(n)}$ follows a binomial law $\mathcal{B}(n, 1/2)$, such that $\Pr(B_i^{(n)} = k) = \binom{n}{k} 2^{-n}$.
The Markov chain is proved to be irreducible and aperiodic, which is known to imply that there is unique positive stationary distribution. The remarkable job in [1] is to prove that the latter is given by the mean of Eulerian triangle numbers, so that there are - surprising and beautiful - connections with sequences coming from arithmetic and combinatorial theory:

Theorem 2 *(1) The Markov chain $(A_i^{(n)})_{i \geq -1}$ is irreducible and aperiodic.*
(2) The - positive - stationary distribution of the Markov chain $(A_i^{(n)})_{i \geq -1}$ is given by $(\alpha_0^{(n)}, \ldots, \alpha_{n-1}^{(n)})$, where

$$\alpha_s^{(n)} = \frac{1}{n!} \left\langle {n \atop s} \right\rangle, \quad (0 \leq s \leq n-1).$$

Here $\left\langle \begin{matrix} n \\ s \end{matrix} \right\rangle$ denotes the Eulerian triangle numbers, defined for $n \geq 2, 0 \leq s \leq n-1$, as the number of permutations of $\{1, \ldots, n\}$ with s ascents. Two asymptotic results are derived:

Theorem 3 *(1) Convergence when $i \to +\infty$ for fixed n:*

$$\lim_{i \to +\infty} \gamma_i^{(n)} = \gamma^{(n)} = \frac{1}{n!} \sum_{even\ s} \left\langle \begin{matrix} n \\ s \end{matrix} \right\rangle = \frac{1}{2} \left(1 + \frac{2^{n+1}(2^{n+1}-1) b_{n+1}}{(n+1)!} \right);$$

In particular,

- *if n is even, $\gamma^{(n)} = 1/2$;*
- *if n is odd, $\gamma^{(n)} = \frac{1}{2}(1 + \varepsilon^{(n)})$, with $\varepsilon^{(n)} > 0$ (resp. < 0) for $n \equiv 5 \bmod 4$ (resp. $n \equiv 3 \bmod 4$).*

(2) Convergence when $n \to +\infty$: $\lim_{n \to +\infty} \gamma^{(n)} = 1/2$, and $\gamma^{(n)} - 1/2 = O((2/\pi)^n)$.

Here b_{n+1} denotes the $(n+1)$-th Bernoulli number. Recall the first values of this famous sequence:

b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	$\ldots$
1	$-\frac{1}{2}$	$\frac{1}{6}$	0	$-\frac{1}{30}$	0	$\frac{1}{42}$	0	$-\frac{1}{30}$	0	$\frac{5}{66}$	0	$\ldots$

It will derive from our work in section 3 that the convergence $\lim_{i \to +\infty} \gamma_i^{(n)} = \gamma^{(n)}$ is also exponential: $O(2^{-i})$ for even n and $O(2^{-2i})$ for odd n.

Independently, Meier and Staffelbach obtained in [2] the spectrum of P_n (the proof is by induction on n).

Theorem 4 *The eigenvalues of P_n are $1, 1/2, \ldots, 1/2^{n-1}$.*

Finally, section 5 in [1] gives elements to compute explicit formulas for some small fixed values of n, that is $\gamma_i^{(n)}$ as a function of i. This is the point where we propose some improvement.

3 Our improvement for explicit formulas

3.1 Formal algebraic approach

3.1.1 Obtaining $\gamma_i^{(n)}$ as a function of i for fixed n

Probability distribution for $(\Pr(A_{i-1}^{(n)} = s))_{s=0,\ldots,n-1}$ is given inductively by the Markov chain. Once again, we make use of notations of [1] and set $\beta_{i,s}^{(n)} = \Pr(A_{i-1}^{(n)} = s)$. We have, for $i \geq 0$,

$$(\beta_{i,0}^{(n)}, \ldots, \beta_{i,n-1}^{(n)}) = (\beta_{i-1,0}^{(n)}, \ldots, \beta_{i-1,n-1}^{(n)}) P_n$$

and, by an immediate induction,

$$\begin{aligned}(\beta^{(n)}_{i-1,0},\ldots,\beta^{(n)}_{i-1,n-1}) &= (\beta^{(n)}_{-1,0},\ldots,\beta^{(n)}_{-1,n-1})P_n^i \\ &= (1,0,\ldots,0)P_n^i\end{aligned}$$

where P_n^i is the i-th power of P_n. In other words, the probability distribution $(\beta^{(n)}_{i-1,0},\ldots,\beta^{(n)}_{i-1,n-1})$ at step $i-1$ is the first row of the matrix P_n^i. Note that we retrieve that when $i \to +\infty$, $(\beta^{(n)}_{i,0},\ldots,\beta^{(n)}_{i,n-1})$ has a limit equal to the stationary distribution[1]. Relatedly, the studied probability $\gamma_i^{(n)} = \sum_{\text{even } s} \Pr(A_{i-1}^{(n)} = s)$ is equal to the sum of even-index components of the first row of the matrix P_n^i. We feel convenient, although apparently uselessly artificial, to write it as a row vector-column vector product:

$$\gamma_i^{(n)} = (\beta^{(n)}_{i-1,0},\ldots,\beta^{(n)}_{i-1,n-1}) \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}$$

where $\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}$ is a column vector whose components are 1 (resp. 0) for even (resp. odd) indices[2]. Finally, we state the

Theorem 5 *For $n \geq 2$ and $i \geq 0$, $\gamma_i^{(n)} = (1,0,\ldots,0)P_n^i \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}$.*

This a compact closed formula, but with no immediate practical interest since we have to compute P_n^i to get $\gamma_i^{(n)}$. Now, we make use of the knowledge of the spectrum of P_n. Since there are n pairwise distinct eigenvalues, P_n is diagonalisable, *i.e.* there exist $Q \in GL(n,\mathbb{Q})$ such that

$$P_n = Q^{-1} \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1/2 & \ldots & 0 \\ & & \ldots & \\ 0 & 0 & \ldots & 1/2^{n-1} \end{pmatrix} Q.$$

In this form, the rows of Q give the left eigenvectors of P_n, in particular the first row of Q is the stationary distribution of P_n (up to normalization). It also follows that, for any $i \geq 0$:

$$P_n^i = Q^{-1} \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1/2^i & \ldots & 0 \\ & & \ldots & \\ 0 & 0 & \ldots & 1/2^{(n-1)i} \end{pmatrix} Q.$$

[1] Perron-Frobenius' theorem actually says, among other things, that P_n^i tends to some positive matrix whose rows are all equal to the unique - positive - invariant vector.

[2] recall our indices range in $\{0,\ldots,n-1\}$.

Consequently, every coefficient in P_n^i is a linear combination of $1, 2^{-i}, \ldots, 2^{-(n-1)i}$, and so is $\gamma_i^{(n)}$: there exist $g_0^{(n)}, \ldots, g_{n-1}^{(n)} \in \mathbb{Q}$ such that

$$\forall i \geq 0, \quad \gamma_i^{(n)} = \sum_{s=0}^{n-1} g_s^{(n)} 2^{-si}$$

Such a sequence is a linear recurring sequence whose characteristic (and minimal) polynomial is the one of P_n, that is $\prod_{s=0}^{n-1}(X - 2^{-s})$. Now, it remains to determine coefficients $g_0^{(n)}, \ldots, g_{n-1}^{(n)}$. We use the classical methodolgy of linear recurring sequences, that is compute directly the n first values of $\gamma_i^{(n)}$, for $i = 0, \ldots, n-1$, then express them formally and solve the linear system. This approach avoids us to compute matrices Q and Q^{-1}. Here we get the following linear system:

$$\left\{ \begin{array}{rcl} \gamma_0^{(n)} & = & g_0^{(n)} + g_1^{(n)} + \cdots + g_{n-1}^{(n)} \\ \gamma_1^{(n)} & = & g_0^{(n)} + g_1^{(n)} 2^{-1} + \cdots + g_{n-1}^{(n)} 2^{-(n-1)} \\ \cdots & & \\ \gamma_{n-1}^{(n)} & = & g_0^{(n)} + g_1^{(n)} 2^{-(n-1)} + \cdots + g_{n-1}^{(n)} 2^{-(n-1)^2} \end{array} \right.$$

such that

$$\begin{array}{rcl} (\gamma_0^{(n)}, \gamma_1^{(n)}, \ldots, \gamma_{n-1}^{(n)}) & = & (g_0^{(n)}, \ldots, g_{n-1}^{(n)}) \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & 2^{-1} & \ldots & 2^{-(n-1)} \\ \vdots & \vdots & & \vdots \\ 1 & 2^{-(n-1)} & \ldots & 2^{-(n-1)^2} \end{bmatrix} \\ & = & (g_0^{(n)}, \ldots, g_{n-1}^{(n)})\ V(1, 2^{-1}, \ldots, 2^{-(n-1)}) \end{array}$$

where "V" stands for Vandermonde (here a square matrix). Since the 2^{-s} are pairwise different, this Vandermonde matrix is invertible. Let us now sum up:

Theorem 6 *Let $\gamma_i^{(n)} = Pr(L_i^{(n)} = S_i^{(n)})$. Fix $n \geq 2$. The following closed formula gives $\gamma_i^{(n)}$ as a function of i:*

$$\forall i \geq 0, \quad \gamma_i^{(n)} = g_0^{(n)} + g_1^{(n)} 2^{-i} + \cdots + g_{n-1}^{(n)} 2^{-(n-1)}$$

where

$$(g_0^{(n)}, \ldots, g_{n-1}^{(n)}) = (\gamma_0^{(n)}, \gamma_1^{(n)}, \ldots, \gamma_{n-1}^{(n)})\ V(1, 2^{-1}, \ldots, 2^{-(n-1)})^{-1}$$

and the n first values $\gamma_0^{(n)}, \gamma_1^{(n)}, \ldots, \gamma_{n-1}^{(n)}$ are computed directly by expression in theorem 5.

Note that our method applies in the same way to the whole matrix P_n^i (resp. the probability distribution $(\beta_{i-1,0}^{(n)}, \ldots, \beta_{i-1,n-1}^{(n)})$) , because it satisfies itself the linear recurring equation of characteristic polynomial $\prod_{s=0}^{n-1}(X - 2^{-s})$. Namely, there exist matrices $G_0^{(n)}, \ldots, G_{n-1}^{(n)} \in \mathrm{Mat}_n(\mathbb{Q})$ such that

$$\forall i \geq 0, \quad P_n^i = G_0^{(n)} + G_1^{(n)} 2^{-i} + \cdots + G_{n-1}^{(n)} 2^{-(n-1)i}.$$

We can formally write:

$$(I_n, P_n, \ldots, P_n^{n-1}) = (G_0^{(n)}, \ldots, G_{n-1}^{(n)})\ V(1, 2^{-1}, \ldots, 2^{-(n-1)})$$

and thus

$$(G_0^{(n)}, \ldots, G_{n-1}^{(n)}) = (I_n, P_n, \ldots, P_n^{n-1})\ V(1, 2^{-1}, \ldots, 2^{-(n-1)})^{-1},$$

the latter meaning each $G_s^{(n)}$ is equal to some linear combination of $I_n, P_n, \ldots, P_n^{n-1}$.

3.2 Complexity estimation

3.2.1 Computing $\gamma_i^{(n)}$ and al.

For given n, theorem 6 naturally derives in an algorithm to numerically compute the closed formula. Let us determine its complexity. First step is to compute the $\gamma_0^{(n)}, \ldots, \gamma_{n-1}^{(n)}$ using theorem 5: it costs $O(n^3)$ (n row vector-matrix products, computing iteratively $(1, 0, \ldots, 0)P_n^i$) . Then, we have to compute the inverse of the Vandermonde matrix, which is $O(n^2)$ (see for example [4])[3]. The overall complexity is therefore $O(n^3)$.

If we compute a closed formula of the complete matrix P_n^i (resp. the probability distribution $(\beta_{i-1,0}^{(n)}, \ldots, \beta_{i-1,n-1}^{(n)})$), the complexity goes to $O(n^4)$ (resp. remains the same). Indeed, first step is the direct computation of $P_n^2, \ldots, P_n^{n-1}$, thus costs $O(n^4)$ (resp. is as above, $O(n^3)$). Second step is the computation of the inverse of the Vandermonde matrix, as abovee. Third step is the computation of n linear combinations of $n \times n$ matrices (resp. of n-dimensionnal vectors), that is $O(n^4)$ (resp.$O(n^3)$).

Remark It is worth to note, among other things, that computing explicit formula for $(\beta_{i-1,0}^{(n)}, \ldots, \beta_{i-1,n-1}^{(n)})$ has essentially the same cost as computing $\gamma_i^{(n)}$ with all this approach. This is not so surprising because what we do, in fact, is implicitely computing $(\beta_{i-1,0}^{(n)}, \ldots, \beta_{i-1,n-1}^{(n)})$ and then using the identity:

$$\gamma_i^{(n)} = (\beta_{i-1,0}^{(n)}, \ldots, \beta_{i-1,n-1}^{(n)}) \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}.$$

3.3 Experimental results

We have run our algorithm on small values of n. Results are given in appendix A. Note that coefficient $g_0^{(n)}$ is the limit of the studied probability $\gamma_i^{(n)}$ when $i \to +\infty$, and we retrieve P. Sarkar's numerical results for the first values of $\gamma^{(n)}$. Moreover, our experimental results suggest some conjecture on the coefficients of the linear combination. Before the statement, we need some notation. We aim study the sign of elements of the sequence $(g_1^{(n)}, \ldots, g_{n-1}^{(n)})$. We write, for instance, $(g_1^{(n)}, \ldots, g_{n-1}^{(n)}) = (+\ 0\ -\ 0\ \ldots)$ to mean $g_1^{(n)} > 0, g_2^{(n)} = 0, \ldots$. Now we can state the conjecture, as follows:

[3]Note that the generic inversion algorithm would suffice here since the dominating complexity of is $O(n^3)$, at first step.

Conjecture 1 *For $n \geq 2$,*

- *if* $n \equiv 2 \bmod 4$, $(g_1^{(n)}, \ldots, g_{n-1}^{(n)}) = (+\ 0\ -\ \ldots\ 0\ +)$;
- *if* $n \equiv 3 \bmod 4$, $(g_1^{(n)}, \ldots, g_{n-1}^{(n)}) = (0\ +\ 0\ -\ \ldots\ 0\ +)$;
- *if* $n \equiv 4 \bmod 4$, $(g_1^{(n)}, \ldots, g_{n-1}^{(n)}) = (-\ 0\ +\ \ldots\ 0\ +)$;
- *if* $n \equiv 5 \bmod 4$, $(g_1^{(n)}, \ldots, g_{n-1}^{(n)}) = (0\ -\ 0\ +\ \ldots\ 0\ +)$;

The conjecture has been checked for $2 \leq n \leq 65$ by numerical computations.

3.4 (Pseudo) Open problem: fully explicit formula

At this point, we have written $\gamma_i^{(n)}$ as a function of $i \geq 0$, for fixed n, say $f_n(i)$. We would like to write $\gamma_i^{(n)}$ as an explicit function of n and i.
In our first post, we said that we had no idea to tackle this issue. But, see section 5, the problem indeed has a solution that was suggested to us a few days ago. It happens to prove Conjecture 1 too.

4 Study of the q-ary addition

4.1 Introduction

In this section, which is independent on the previous one, we generalize Sarkar's approach to any numeration basis. Namely, we assume that summands are integers written is base q, and we try to approximate the symbols of the arithmetic addition by modulo q addition of the symbols of the summands. For sake of simplicity, we will call "digit" such these symbols, as they are called in the specific case of base $q = 10$. Moreover, the latter is indeed a good numerical example to make up one mind of how the things work.

4.2 Notations

Let $X^{(1)}, \ldots, X^{(n)}$ be integers, and $X_i^{(k)}$ is the i-th digit of the q-ary expansion of $X^{(k)}$. 0-th digit is the least significant digit. Now variables $X_i^{(k)}$ are in $\{0, \ldots, q-1\}$. To save additional notations, integers addition and modulo q addition will both be denoted by $+$, and to avoid any confusion, we will explicitely write "mod q" in the latter case. We keep the same capital letters to denote similar objects, such that, for $i \geq 0$:

- $L_i^{(n)} = X_i^{(1)} + \cdots + X_i^{(n)} \bmod q$;
- $S_i^{(n)}$ is the i-th digit of $X^{(1)} + \cdots + X^{(n)}$;
- $B_i^{(n)} = X_i^{(1)} + \cdots + X_i^{(n)}$;
- $A_i^{(n)}$ is the carry output by the i-th column of the addition. It is convenient to set $A_{-1}^{(n)} = 0$.

The following drawings represent both additions.
Arithmetic addition

$$\begin{array}{cccccccc}
 & \cdots & A_i^{(n)} & \cdots & A_0^{(n)} & & & Carry \\
 & \cdots & X_i^{(1)} & \cdots & X_1^{(1)} & X_0^{(1)} & & Summand\#1 \\
+ & \cdots & X_i^{(2)} & \cdots & X_1^{(2)} & X_0^{(2)} & & Summand\#2 \\
+ & \cdots & & & & & & \cdots \\
+ & \cdots & X_i^{(n)} & \cdots & X_1^{(n)} & X_0^{(n)} & & Summand\#n \\
\hline
 & \cdots & S_i^{(n)} & \cdots & S_1^{(n)} & S_0^{(n)} &
\end{array}$$

q-ary modular addition

$$\begin{array}{cccccccc}
 & \cdots & X_i^{(1)} & \cdots & X_1^{(1)} & X_0^{(1)} & & Summand\#1 \\
+ & \cdots & X_i^{(2)} & \cdots & X_1^{(2)} & X_0^{(2)} & & Summand\#2 \\
+ & \cdots & & & & & & \cdots \\
+ & \cdots & X_i^{(n)} & \cdots & X_1^{(n)} & X_0^{(n)} & & Summand\#n \\
\hline
 & \cdots & L_i^{(n)} & \cdots & L_1^{(n)} & L_0^{(n)} &
\end{array}$$

We aim to study the correlation between $L_i^{(n)}$ and $S_i^{(n)}$, that is $\gamma_i^{(n)} = \Pr(L_i^{(n)} = S_i^{(n)})$. Note that here both $L_i^{(n)}$ and $S_i^{(n)}$ are digits, *i.e.* numbers in $\{0, \ldots, q-1\}$.

4.3 Similar results as in the binary addition

Lemma 4 *For $i \geq 0$*

$$\begin{aligned}
S_i^{(n)} &= (A_{i-1}^{(n)} + B_i^{(n)}) \bmod q \\
A_i^{(n)} &= (A_{i-1}^{(n)} + B_i^{(n)}) \div q
\end{aligned}$$

The proof is essentially the same as in the one presented above: "put" something and "retain" something else. □

We now study the values of the carry sequence. Of course, some things have changed and the various variables now range in more general subsets. For example, $B_i^{(n)}$ may take all integer values in $\{0, \ldots, n(q-1)\}$. Among all these modifications, the following result on $A_i^{(n)}$ is somehow remarkable: it establishes that the values are in $\{0, \ldots, n-1\}$, as previously, independently on the value of q.

Lemma 5

$$For\ i \geq -1, \quad A_i^{(n)} \in [0, n-1]$$

Proof We reproduce the induction. The property is true for $i = -1$. Assume that it holds at step $i-1$ ($i \geq 0$). Then

$$A_i^{(n)} = \left\lfloor \frac{A_{i-1}^{(n)} + B_i^{(n)}}{q} \right\rfloor \leq \left\lfloor \frac{n-1+n(q-1)}{q} \right\rfloor = \left\lfloor n-1+\frac{q-1}{q} \right\rfloor = n-1,$$

and property holds at step i. □

The following result explicitly relates the studied correlation with the values of $A_i^{(n)}$.

Lemma 6 *For $n \geq 2, i \geq 0$, we have*

$$\gamma_i^{(n)} = Pr(A_{i-1}^{(n)} \equiv 0 \bmod q) = \sum_{s \equiv 0 \bmod q} Pr(A_{i-1}^{(n)} = s)$$

Proof We have $S_i^{(n)} = A_{i-1}^{(n)} + X_i^{(1)} + \cdots + X_i^{(n)} \bmod q = A_{i-1}^{(n)} + L_i^{(n)} \bmod q$. Moreover, $S_i^{(n)}$ and $L_i^{(n)}$ are in $\{0, \ldots, n-1\}$. Therefore, $L_i^{(n)} = S_i^{(n)} \Leftrightarrow A_{i-1}^{(n)} = 0 \bmod q$, and the conclusion follows. □

To study the probability distribution of values of $A_i^{(n)}$, we follow the same methodology as [1]. The main difference is that $X_i^{(k)}$ are no longer Bernoulli variables, hence $B_i^{(n)}$ has no longer binomial distribution. The following section focuses on the new distribution.

4.4 Multi-uniform probability law

Let $X_1, \ldots, X_n$ discrete random variables, independent identically distributed (i.i.d.), following the uniform distribution law over $\{0, \ldots, q-1\}$ (denoted by $\mathcal{U}(q)$). Let $Y = X_1 + \cdots + X_n$. The case $q = 2$ leads to the classical Bernoulli and binomial laws. In the general case, we are able to give some closed expression for the distribution of Y. For this purpose, generating functions are a powerful tool. Recall that, for a (discrete $\mathbb{N}$-valued) variable X, the generating function ϕ_X is a formal series in $\mathbb{R}[[z]]$, defined by $\phi_X(z) = E[z^X] = \sum_k \Pr(X = k) z^k$.

Theorem 7 *Let $X_1, \ldots, X_n$ i.i.d. random variables with uniform distribution $\mathcal{U}(q)$. Then the probability distribution of random variable $Y = X_1 + \cdots + X_n$ is given by*

$$\begin{aligned} Pr(Y = k) &= \sum_{\substack{k_0 \geq 0, \ldots, k_{q-1} \geq 0, \\ k_0 + \cdots + k_{q-1} = n, \\ k_1 + 2k_2 + \cdots + (q-1)k_{q-1} = k}} \frac{1}{q^n} \frac{n!}{k_0! k_1! \ldots k_{q-1}!} \quad \textit{for } 0 \leq k \leq n(q-1), \\ &= 0 \qquad \textit{otherwise.} \end{aligned}$$

We call this law the *multiuniform law with parameters q and n* and denote it $\mathcal{MU}(q, n)$. Note that $\mathcal{MU}(2, n)$ coincide with binomial law with parameters n and $1/2$.

Proof Since $X_1, \ldots, X_n$ are independent random variables, we have $\phi_Y(z) = \phi_{X_1}(z) \ldots \phi_{X_n}(z)$. (assumption "identically distributed" is not necessary for this specific equality). For each X_k, the generating function is $\phi_{X_k}(z) = \frac{1}{q}\left(1 + z + \cdots + z^{q-1}\right)$, so that

$$\begin{aligned} \phi_Y(z) &= \frac{1}{q^n}(1 + z + \cdots + z^{q-1})^n \\ &= \sum_{\substack{k_0 \geq 0, \ldots, k_{q-1} \geq 0, \\ k_0 + \cdots + k_{q-1} = n}} \frac{1}{q^n} \frac{n!}{k_0! k_1! \ldots k_{q-1}!} z^{k_1 + 2k_2 + \cdots + (q-1)k_{q-1}} \end{aligned} \tag{1}$$

Expression of $\Pr(Y = k)$ is given by the coefficient of z^k in equation (1). □

4.5 The Markov chain of carries

We fix some $n \geq 2$. Random variable $B_i^{(n)}$ follows multiuniform $\mathcal{MU}(q, n)$ distribution, for which a – complex ! – closed formula have been found. Note that it is independant on i. Consequently, equality $A_i^{(n)} = (A_{i-1}^{(n)} + B_i^{(n)}) \div q$ implies that $(A_i^{(n)})_{i \geq -1}$ is a homogeneous discrete-time finite-state Markov chain of order 1. The initial probability distribution, for $i = -1$, is $(1, 0, \ldots, 0)$. Let us compute the transition matrix: for every $s, t \in \{0, \ldots, n-1\}$,

$$\{A_i^{(n)} = t | A_{i-1}^{(n)} = s\} \Leftrightarrow \left\lfloor \frac{s + B_i^{(n)}}{q} \right\rfloor = t$$

$$\Leftrightarrow t \leq \frac{s + B_i^{(n)}}{q} < t + 1 \Leftrightarrow qt - s \leq B_i^{(n)} < q(t+1) - s$$

We derive the

Theorem 8 *Let $P_{n,q}$ be the transition matrix of the Markov chain $(A_i^{(n)})_{i \geq -1}$. For $0 \leq s, t \leq n-1$,*

$$P_{n,q}(s,t) = \sum_{qt - s \leq k_1 + 2k_2 + \cdots + (q-1)k_{q-1} \leq qt - s + q - 1} \frac{1}{q^n} \binom{n}{k_1, \ldots, k_{q-1}}.$$

where we use the following notation for multinomial (here q-nomial) coefficient:

$$\binom{n}{k_1, \ldots, k_{q-1}} = \begin{cases} \dfrac{n!}{k_1! \ldots k_{q-1}! (n - k_1 - \cdots - k_{q-1})!} & \text{if } \begin{cases} k_1, \ldots, k_{q-1} \geq 0, \\ k_1 + \cdots + k_{q-1} \leq n, \end{cases} \\ 0 \quad \text{otherwise.} \end{cases}$$

Proof From definition of $P_{n,q}$ and theorem 7, we have

$$P_{n,q}(s,t) = \Pr(A_i^{(n)} = t | A_{i-1}^{(n)} = s) = \Pr(qt - s \leq B_i^{(n)} \leq qt - s + q - 1)$$

$$= \sum_{\substack{k_0 \geq 0, \ldots, k_{q-1} \geq 0, \\ k_0 + \cdots + k_{q-1} = n, \\ qt - s \leq k_1 + 2k_2 + \cdots + (q-1)k_{q-1} \leq qt - s + q - 1}} \frac{1}{q^n} \frac{n!}{k_0! k_1! \ldots k_{q-1}!}$$

$$= \sum_{\substack{k_1 \geq 0, \ldots, k_{q-1} \geq 0, \\ k_1 + \cdots + k_{q-1} \leq n, \\ qt - s \leq k_1 + 2k_2 + \cdots + (q-1)k_{q-1} \leq qt - s + q - 1}} \frac{1}{q^n} \frac{n!}{(n - k_1 - \cdots - k_{q-1})! k_1! \ldots k_{q-1}!}.$$

The last equality comes from the last but one by a change of indices in the summation (removal of k_0). It leads to the desired form applying our notation for q-nomials. □

4.6 Experimental - and conjectural - results

We have done some computations for small values of n and q, namely $4 \leq n \leq 7$ and $2 \leq q \leq 7$. They suggest the following conjecture.

Conjecture 2 *(1) The eigenvalues of $P_{n,q}$ are $1, q^{-1}, \ldots, q^{-(n-1)}$.*
(2) Left eigen vectors of $P_{n,q}$ do not depend on q. In matricial setting, there exists $Q \in GL(n, \mathbb{Q})$ that does not depend on q such that

$$P_{n,q} = Q^{-1} \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1/q & \ldots & 0 \\ & & \ldots & \\ 0 & 0 & \ldots & 1/q^{n-1} \end{pmatrix} Q.$$

Comments We believe that (1) may be proved by induction in a similar way as in [2]. But the generalization of the proof is not straightforward at all. Expression for $P_{n,q}(s,t)$ is much more complex than in the binary case, and imply delicate handlings of q-nomial coefficients.

On the other hand, we have absolute no idea nor initiating key to prove (2), which we find really amazing.

If the first part of the conjecture is true, then it is straightforward to obtain closed expressions for $\gamma_i^{(n)}$, for fixed n by the same methodology. Probability $\gamma_i^{(n)}$ is as a linear combination of $1, q^{-i}, \ldots, q^{-(n-1)i}$, hence has some limit $\gamma^{(n)}$ as $i \to +\infty$.

As a further work, we plan to perform experimental investigations on the form of these closed formulas. For example, among other things, we will be able to examine values of $\gamma^{(n)}$ to see if

- they are close or equal to $1/q$ (our reasonable guess);
- they have some limit close or equal to $1/q$ when $n \to +\infty$.

5 Very late news

After the first post of our work, we have been told that our conjectures had actually been studied and proved in a paper by Holte [3]. Consequently, they can in fact be raised as theorems. Although it is a little disappointing not to be original, we have decided to leave our work on the archive, for the sake of common knowledge. The paper [3] is scientifically rich and very pleasant to read. We recommend it. It contains very clever computational tricks, far more efficient than ours. Let us quote, for example, an alternative expression for coefficients in Markov matrix of q-ary case. Instead of involving multinomial coefficients, Holte writes, up to notations, something like

$$\begin{aligned} (1 + z + \cdots + z^{q-1})^n &= (1 - z^q)^n (1 - z)^{-n} \\ &= \left(\sum_{r=0}^{+\infty} \binom{n}{r} (-z^q)^r \right) \left(\sum_{s=0}^{+\infty} \binom{n+s-1}{s} z^s \right) \end{aligned}$$

such that everything can be expressed with combinations of binomial coefficients. As another example, the same trick leads to an alternative expression of the multi-uniform probability law:

$$\Pr(Y = k) = \frac{1}{q^n} \sum_{r=0}^{\lfloor k/q \rfloor} (-1)^r \binom{n}{r} \binom{n+k-rq-1}{n-1}.$$

Ernst Schulte-Geers, who brought Holte's paper to us, has further investigated Holte's results and suggests us a solution for fully explicit formula, our "pseudo open problem" in section 3.4, that turns to solve also our Conjecture 1. The formula is the following:

$$\gamma_i^{(n)} = \frac{1}{2} + \sum_{j=0}^{n-1} \frac{2^n(2^{n-j+1}-1)}{n!(n-j+1)} \begin{bmatrix} n \\ n-j \end{bmatrix} b_{n-j+1} 2^{-ij},$$

where b_k are, as above, the Bernoulli numbers, and $\begin{bmatrix} n \\ k \end{bmatrix}$ are the Stirling numbers of first kind. For the sake of completeness, we recall two possible definitions of the latter:

- (combinatorial) $\begin{bmatrix} n \\ k \end{bmatrix}$ is the number of ways to arrange n elements into k cycles;
- (algebraic) $\begin{bmatrix} n \\ k \end{bmatrix}$ satisfy the polynomial identity

$$X(X+1)\dots(X+n-1) = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix} X^k.$$

6 Conclusion

In this paper, we investigate approximations of arithmetic addition by q-ary addition. We first deal with the case $q = 2$, and taking P. Sarkar's work [1] as starting point, we explain a systematic way to obtain closed formulas. Then we show that the general q-ary case has important similarities with the binary one, the main of which being the Markov chain formed by the carries. We give expression of the matrix with the help of q-nomial coefficients. Experimental computations lead to re discover some conjectures on the Markov matrix, that in fact were proven in Holte's paper.

Acknowledgements

We wish to thank Palash Sarkar for paying attention to our sollicitations, and encouraging us to post our work. We wish to thank Ernst Schulte-Geers for bringing Holte's paper to us and suggesting the solution for our "open problem" in section 3.4.
Finally, we would like to apologize to J.M. Holte for our ignoring his paper in our first post, which was not intended of course.

References

[1] P. Sarkar: On Approximating Addition by Exclusive OR, eprint 2009/047.

[2] O. Staffelbach, W. Meier: Cryptographic Significance of the Carry for Ciphers based on Integer Addition, Crypto 1990, LNCS volume 537, pages 601-614, Springer.

[3] J.M. Holte: Carries, Combinatorics and an Amazing Matrix, The American Mathematical Monthly, Vol 104 N.2 (feb. 1997).

[4] `http://lumimath.univ-mrs.fr/~jlm/travaux/livretab/node23.html`.

A Numerical results for binary addition

n	$\gamma_i^{(n)}, \quad \forall i \geq 0$
2	$\frac{1}{2} + \frac{1}{2}2^{-i}$
3	$\frac{1}{3} + \frac{2}{3}2^{-2i}$
4	$\frac{1}{2} - \frac{1}{2}2^{-i} + 2^{-3i}$
5	$\frac{17}{30} - \frac{7}{6}2^{-2i} + \frac{8}{5}2^{-4i}$
6	$\frac{1}{2} + \frac{1}{3}2^{-i} - \frac{5}{2}2^{-3i} + \frac{8}{3}2^{-5i}$
7	$\frac{149}{315} + \frac{10}{9}2^{-2i} - \frac{232}{45}2^{-4i} + \frac{32}{7}2^{-6i}$
8	$\frac{1}{2} - \frac{17}{90}2^{-i} + \frac{28}{9}2^{-3i} - \frac{469}{45}2^{-5i} + 8\ 2^{-7i}$
9	$\frac{2897}{5670} - \frac{221}{270}2^{-2i} + \frac{1069}{135}2^{-4i} - \frac{59062}{2835}2^{-6i} + \frac{128}{9}2^{-8i}$
10	$\frac{1}{2} + \frac{31}{315}2^{-i} - \frac{17}{6}2^{-3i} + 19\ 2^{-5i} - \frac{2606}{63}2^{-7i} + \frac{128}{5}2^{-9i}$
11	$\frac{154543}{311850} + \frac{496}{945}2^{-2i} - \frac{11611}{1350}2^{-4i} + \frac{124252}{2835}2^{-6i} - \frac{128824}{1575}2^{-8i} + \frac{512}{11}2^{-10i}$
12	$\frac{1}{2} - \frac{691}{14175}2^{-i} + \frac{682}{315}2^{-3i} - \frac{32351}{1350}2^{-5i} + \frac{278762}{2835}2^{-7i} - \frac{762212}{4725}2^{-9i} + \frac{256}{3}2^{-11i}$
13	$\frac{6102919}{12162150} - \frac{13129}{42525}2^{-2i} + \frac{108314}{14175}2^{-4i} - \frac{5348489}{85050}2^{-6i} + \frac{9191704}{42525}2^{-8i} - \frac{16509056}{51975}2^{-10i} + \frac{2048}{13}2^{-12i}$