Constellations Cross Circular auto-Correlation C4-sequences

Emmanuel Boutillon

Constellations Cross Circular auto-Correlation C4-sequences

Emmanuel Boutillon,
Université Bretagne Sud and IMT-Atlantique, Lab-STICC UMR 6285, CNRS
Email: emmanuel.boutillon@univ-ubs.fr

***Abstract*—This paper introduces a novel type of sequences called C4-sequences. C4-sequences share similar optimal auto-correlation properties with Zadoff-Chu sequences. However, C4-sequences offer the additional advantage of being also optimal (in the sense of minimal Euclidean distance between sequences) for four truncation lengths, providing flexibility in adapting to different channel conditions without compromising performance. Moreover, unlike Zadoff-Chu sequences, the points of a constellation associated with a C4-sequence are not limited to the unit circle. This opens up possibilities for achieving shaping gain, leading to enhanced spectral efficiency. By combining a truncated C4-sequence modulation as an inner code with a fixed-rate non-binary outer code, flexible and performant rate-adaptive communication systems can also be achieved. Finally, the notion of C4-sequences can be generalized.**

***Index Terms*—Low SNR, Rate-adaptive, CCSK, truncated, sequence, short message.**

I. Introduction

The Cyclic Code Shift Keying (CCSK) modulation is a widely recognized spreading technique [1], utilized to enhance the spectral efficiency of a spreading sequence of length q. It achieves this by employing its q circularly rotated versions to encode $m = \log_2(q)$ bits per sequence transmission. In addition, [1] suggested truncating the CCSK sequence to its first l elements in order to increase the spectral efficiency. Recently, Marchand et al. introduced a coded-modulation scheme that integrates a fixed-rate non-binary outer code with a variable-rate inner code based on variable-length truncated CCSK modulation. They utilize a binary CCSK sequence in [2] and a q-ary CCSK sequence in [3] to devise efficient rate-adaptive communication schemes. In [4], the q-CCSK sequences found in [3] are generalized to define the class of C4-sequence. This paper is an extension of [4]. It presents a systematic mathematical construction method for generating optimal q-ary CCSK sequences of length q, which are referred to as C4-sequences. The term "C" stands for the first letter "C" of the four words: "Constellation", "Cross", "Circular", and "Correlation". It also refers to the four truncation lengths that yield an inner code with an optimal distance property, namely $q/4$, $q/2$, $3q/4$, and q. In addition, the paper shows that when the sequence is not truncated (i.e., a sequence of length q), C4-sequences exhibit the same autocorrelation property as the well-known Zadoff-Chu (ZC) sequences [5], [6] or the chirp spread spectrum (CSS) modulation used in LoRaWAN system (Long Range wide area network [7]). The current paper gives the demonstrations that were not given in [4]. It also describes a method to optimize C4-sequence, and presents new simulation results and theoretical results on the asymptotic spectral efficiency of C4-sequences and the distances between two distinct truncated C4-sequences. Finally, it generalizes C4-sequences to C3-sequences or C5-sequences, or more generally, to any Cn sequences, with n an integer greater than 1.

The rest of the paper is organized as follows. Section II defines the C4-sequence and presents its main properties. Section III is dedicated to the optimization of a C4-sequence for a specific objective. Section IV introduces the truncated C4-sequence and its use in a concatenated coded modulation system consisting of a fixed rate non-binary outer code and a variable length truncated C4-sequence as inner code. Section V deals with the distance properties of a set of C4-sequences for multi-user applications, while Section VI generalizes C4-sequences to arbitrary Cn sequences. Finally, Section VII concludes the paper.

Notations: The complex vector $\mathbf{x} = (x(0), x(1), \ldots, x(n), \ldots)$ is a q-periodic infinite vector of complex numbers. The vector $\mathbf{x}_a$ represents the vector $\mathbf{x}$ left-shifted by a positions, i.e., for all n, $x_a(n) = x(n+a)$. The vector $\mathbf{x}_a^{a+l-1}$ denotes the truncated vector obtained by taking the first l values of $\mathbf{x}_a$, i.e., $\mathbf{x}_a^{a+l-1} = (x(n+a))_{n=0,1,\ldots,l-1}$. The notation $\langle \mathbf{x}, \mathbf{y} \rangle$ represents the complex scalar product over a period between the vectors $\mathbf{x}$ and $\mathbf{y}$. It is defined as $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=0}^{q-1} x(n)y(n)'$, where $y(n)'$ denotes the complex conjugate of $y(n)$. The inter-correlation vector between $\mathbf{x}$ and $\mathbf{y}$ is denoted $\mathbf{R_{xy}}$, where its τ^{th} component $R_{\mathbf{xy}}(\tau)$ equals to $R_{\mathbf{xy}}(\tau) = \langle \mathbf{x}, \mathbf{y}_\tau \rangle = \sum_{n=0}^{q-1} x(n)y(n+\tau)'$. Finally, $\mathcal{R}(x)$ represents the real part of x.

II. Definition and properties of the C4-sequences

This section reminds the definition of C4-sequences along with certain characteristics [4]. In this study, C4-sequences of length $q = 2^m$ are under consideration, where m is a small integer. Given a C4-sequence of length q, it is feasible to generate q distinct sequences denoted as $\mathbf{x}_a$, where a ranges from 0 to $q-1$, thereby encoding $\log_2(q) = m$ bits of information. The variable p is introduced as $p = q/4$, or equivalently $p = 2^{m-2}$, which holds specific significance. It is worth mentioning that there exist several equivalent methods to define a C4-sequence, including its auto-correlation function, discrete Fourier transform (DFT) function, or directly through its time-domain distance property. In this paper, a

C4-sequence is defined based on its circular auto-correlation property. Suppose $\mathbf{x}$ represents a sequence of length q. Its circular auto-correlation, denoted by $\mathbf{R_{xx}}$, is a vector of length q defined as

$$R_{\mathbf{xx}}(\tau) = \langle \mathbf{x}, \mathbf{x}_\tau \rangle, \ \ \tau = 0, 1, \ldots, q-1. \tag{1}$$

Definition 2.1: A complex sequence of constellation points $\mathbf{x} = (x(n))_{n=0,1,\ldots,q-1}$, $q \geqslant 4$, is said to be a C4-sequence if and only if its circular auto-correlation $\mathbf{R_{xx}}$ vector verifies

$$R_{\mathbf{xx}}(\tau) = \begin{cases} qj^{-ck}, \text{for } \tau = kq/4, \ k = 0, 1, 2, 3 \\ 0 \text{ otherwise,} \end{cases} \tag{2}$$

with c a sign value, i.e., $c \in \{-1, 1\}$ and j the imaginary number verifying $j^2 = -1$. By convention, when $c = 1$, the C4-sequence is referred to as clockwise C4-sequence since the non-null values of $R_{\mathbf{xx}}(\tau)$ take sequentially the values $R_{\mathbf{xx}}(0) = q$, $R_{\mathbf{xx}}(p) = -jq$, $R_{\mathbf{xx}}(2p) = -q$ and $R_{\mathbf{xx}}(3p) = jq$ (i.e., a clockwise rotation direction). Symmetrically, when $c = -1$, the non-null value of $R_{\mathbf{xx}}(\tau)$ takes the value $R_{\mathbf{xx}}(0) = q$, $R_{\mathbf{xx}}(p) = jq$, $R_{\mathbf{xx}}(2p) = -q$ and $R_{\mathbf{xx}}(3p) = -jq$. This type of sequences is thus referred to as a counter-clockwise C4-sequences.

Theorem 2.2: A length q sequence $\mathbf{x}$ is a C4-sequence if and only if its Discrete Fourier Transform (DFT) $\mathbf{X} = \mathcal{F}(\mathbf{x})$ verifies, for all $k = 0, 1, \ldots, q-1$

$$|X(k)|^2 = \begin{cases} 4q & \text{if } (k+c) \bmod 4 = 0, \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

with $c \in \{-1, 1\}$.

Proof: Consider a C4-sequence $\mathbf{x}$ of length q. By computing the circular auto-correlation $\mathbf{R_{xx}}$ defined in (2) in the frequency domain, the following expression is obtained

$$\mathbf{R_{xx}} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{x}) \odot \mathcal{F}(\mathbf{x})'), \tag{4}$$

where $\odot$ represents the term-by-term component multiplication of the two vectors. Considering the DFT of both terms in (4) yields $\mathcal{F}(\mathbf{R_{xx}}) = (\mathcal{F} \circ \mathcal{F}^{-1})(\mathcal{F}(\mathbf{x}) \odot \mathcal{F}(\mathbf{x})')$, and thus

$$\mathcal{F}(\mathbf{R_{xx}}) = \mathcal{F}(\mathbf{x}) \odot \mathcal{F}(\mathbf{x})' = \mathbf{X} \odot \mathbf{X}'. \tag{5}$$

Using the formal expression of the k^{th} terms $\mathcal{F}(R_{\mathbf{xx}})(k)$ of $\mathcal{F}(\mathbf{R_{xx}})$ and by permuting the left and the right terms, (5) gives

$$X(k)X(k)' = |X(k)|^2 = \sum_{\tau=0}^{q-1} R_{\mathbf{xx}}(\tau) e^{\frac{-2\pi j \tau k}{q}}. \tag{6}$$

According to (2), $R_{\mathbf{xx}}(\tau)$ contains only 4 non-null terms for $\tau = 0$, $q/4$, $q/2$ and $3q/4$, thus, (6) gives

$$|X(k)|^2 = \sum_{m=0}^{3} qj^{-cm} e^{\frac{-2\pi j m \frac{q}{4} k}{q}} \tag{7}$$

$$= q \sum_{m=0}^{3} j^{-(k+c)m}. \tag{8}$$

According to (8), $|X(k)|^2$ is equal to the product of q with the sum of the first 4 terms of a geometric series with a common ratio $\rho = j^{-(k+c)}$. This sum equals 4 if the common ratio ρ is equal to 1, which occurs when $(k+c) \bmod 4 = 0$, and the sum equals 0 otherwise. Reciprocally, if $\mathbf{X} = \mathcal{F}(\mathbf{x})$ verifies (3), then (4) implies that $\mathbf{R_{xx}}$ verifies (2), which gives $\mathbf{x}$ as a C4-sequence □

Theorem 2.2 thus provides an explicit method for constructing a length-q C4-sequence $\mathbf{x}$. In fact, for $k + c = 0 \bmod 4$, $|X(k)|^2 = 4q$ implies that $X(k)$ is a point of the complex circle with radius $\sqrt{4q}$, so it can be expressed as $X(k) = \sqrt{4q} e^{\frac{2\pi j s(k)}{q}}$ with $s(k) \in [0, q[$. Starting from a seed vector $\mathbf{s}$ of length $q/4$, algorithm 1 defines the C4-sequence construction function $\mathbf{x} = G(\mathbf{s})$. In Algorithm 1, the operator $\text{kron}(\mathbf{a}, \mathbf{b})$ represents the Kronecker product between vectors $\mathbf{a}$ and $\mathbf{b}$. The function $I_0(x)$ is the 0-indicator function, i.e. it takes the value 1 if $x = 0$, otherwise 0. Thus, $c = 1$ (clockwise C4-sequence) yields a Kronecker product performed on the vector $[0, 0, 0, 1]$, while $c = -1$ (counter-clockwise C4-sequence) yields a Kronecker product performed on the vector $[0, 1, 0, 0]$.

Algorithm 1 Generation of a C4-sequence of length q by the function $\mathbf{x} = G(\mathbf{s})$

Input A seed vector $\mathbf{s}$ of size $p = q/4$ composed of $q/4$ reals on the interval $[0, q[$, a value of c in the set $\{-1, 1\}$.
Output A clockwise ($c = 1$) or counter-clockwise ($c = -1$) C4-sequence $\mathbf{x}$ of length q

```
for k ← 0 to q/4 − 1 do
    E_s(k) ← √(4q) × e^{2πj s(k)/q}
end for
X ← kron(E_s, [0, I_0(c+1), 0, I_0(c−1)])
x ← F^{-1}(X)
Return x
```

Figure 1 shows the clockwise length-32 C4-sequence $\mathbf{x} = G(\mathbf{s})$ obtained with the seed sequence $\mathbf{s} = (25, 23, 0, 11, 11, 24, 8, 22)$, where $G(\mathbf{s})$ is the function defined in Algorithm 1. The plot labels the first five elements $x(0), x(1), \ldots, x(5)$ and the last element $x(31)$ of the C4-sequence. Each successive pair of points in $\mathbf{x}$ is connected by a line, and the last point $x(31)$ is also connected to the first point $x(0)$.

Prior to presenting a theorem that explicitly asserts the optimality of the C4-sequences, several lemmas are established.

Lemma 2.3: The square of the ℓ_2-norm of a C4-sequence $\mathbf{x}$ of length q is given by $\|\mathbf{x}\|^2 = q$. This implies that the average energy of the components of $\mathbf{x}$ is equal to 1.

Proof: This constitutes a direct application of Parseval's theorem, which asserts that $\sum_{n=0}^{q-1} |x(n)|^2 = \frac{1}{q} \sum_{k=0}^{q-1} |X(k)|^2$. According to Theorem 2.2, $\|\mathbf{X}\|^2$ contains precisely $\frac{q}{4}$ non-zero values, each equal to $4q$. Thus, $\|\mathbf{X}\|^2 = \frac{q}{4} \times 4q = q^2$, thereby leading to $\|\mathbf{x}\|^2 = q$ □

Lemma 2.4: Let $\mathbf{x}$ be a C4-sequence of length q, Then, for all n and for all $k \in \{0, 1, 2, 3\}$, $x(n + kq/4) = j^{-kc} x(n)$.

Proof: The proof of this lemma is given in APPENDIX I.

This lemma implies that $\mathbf{x}$ exhibits a 4-fold rotational symmetry, meaning it remains unaltered under a rotation of

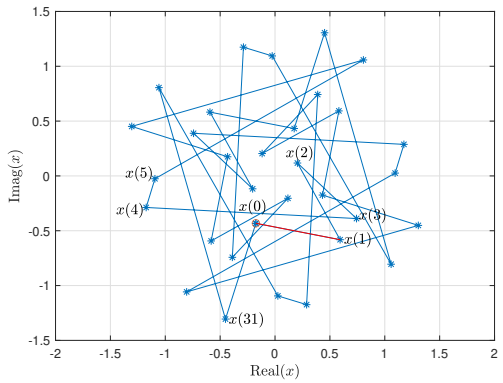

Fig. 1. Example of randomly generated C4-sequence

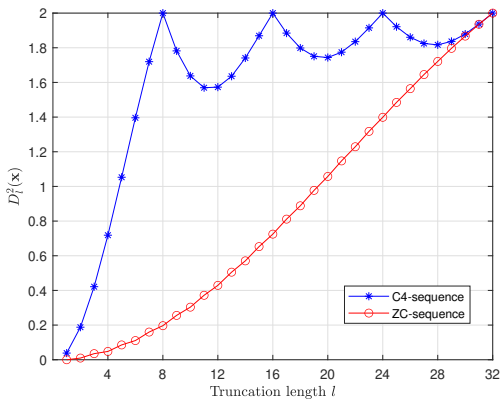

Fig. 2. Normalized square minimum distance $D_l^2(\mathbf{x})$ as a function of l for the C4-sequence given in Fig. 1.

$\pi/2$, as depicted in Fig. 1.

Definition 2.5: Let us consider a vector (or sequence) $\mathbf{x}$ of length q. The notation $\mathbf{x}_a^{a+l-1}$ denotes the subsequence of length l of $\mathbf{x}$ that spans from index a to index $a+l-1$.

From the definition of a length l truncated sequence, the following theorem is derived:

Theorem 2.6: Let $\mathbf{x}$ be a C4-sequence of length q and let $p = q/4$. For any $l \in \{p, 2p, 3p, 4p\}$ and any pair (a, b) of integers between 0 and $q-1$, $a \neq b \Rightarrow \left\|\mathbf{x}_a^{a+l-1} - \mathbf{x}_b^{b+l-1}\right\|^2 \geqslant 2l$. Additionally, if $b - a \neq 2p \pmod q$ then $\left\|\mathbf{x}_a^{a+l-1} - \mathbf{x}_b^{b+l-1}\right\|^2 = 2l$.

Proof: The proof is given in APPENDIX II □

Let the Normalized Minimum Square (NMS) distance $D_l^2(\mathbf{x})$ between two sequences from the set $\{\mathbf{x}_a^{a+l-1}\}_{a=0,1,\ldots,q-1}$ [3] be defined as

$$D_l^2(\mathbf{x}) = \frac{1}{l} \min_{a,b,a\neq b} \{\left\|\mathbf{x}_a^{a+l-1} - \mathbf{x}_b^{b+l-1}\right\|^2\}. \tag{9}$$

According to Theorem 2.6, $D_l^2(\mathbf{x}) = 2$ for $l \in \{p, 2p, 3p, 4p = q\}$. Fig. 2 illustrates the variation of the NMS-Distance $D_l^2(\mathbf{x})$ as a function of the truncation length l for the C4-sequence $\mathbf{x}$ shown in Fig. 1. It also gives, for comparison, the evolution of $D_l^2(\mathbf{z})$, with $\mathbf{c}$ the ZC sequence [5], [6] defined by $\mathbf{c} = (\exp(\pi j n^2/32))_{n=0,1,\ldots,31}$ (note that it is also equivalent to the length-32 CSS sequence).

In summary, C4-sequences are easy to construct and are optimal for truncation lengths p, $2p$, $3p$, and $4p$. Before discussing the application of truncated sequences, the following section focuses on the optimization of the C4-sequence for a given objective.

III. Optimization of C4-sequences

As described in Algorithm 1, it is easy to construct a C4-sequence of size q from a vector random vector of size $p = q/4$. However, each C4-sequence has its own distinct characteristics. They can thus be optimized according to a criterion (or set of criteria) that is application dependent. In this paper, three objective functions are considered: $\psi_D(\mathbf{x})$ the weighted sum of the NMS distances, $\psi_I(\mathbf{x})$ the mutual information (MI) of the $l = 1$ truncated C4-sequence for a given signal-to-noise ratio (SNR) and $\psi_U(\mathbf{x})$, the peak to average power ratio (PAPR) of the C4-sequence.

A. C4-sequence optimization with Greedy algorithm

Considering an objective function $\psi(\mathbf{x})$ to be maximized, it is possible to use a greedy search algorithm on the seed vector $\mathbf{s}$ to maximize $\psi(\mathbf{x})$, with $\mathbf{x} = G(\mathbf{s})$. Algorithm 2 introduces the algorithm used in the paper. First, the step θ is fixed to a size $\theta = p$, then a modification of each element of $s(i)$ of $\mathbf{s}$ is tested with $s(i)+\theta$ and $s(i)-\theta$. If a modification improves the objective function, it is kept. If a local optima is reached, the value of θ is halved and the process is repeated. The process ends when the value of θ is small enough (arbitrarily set to 2^{-4} in Algorithm 2).

Note that Algorithm 2 is one of many possible algorithms for optimization and may not be the most efficient. However, it is relatively simple to implement and fast to execute. Its performance can be significantly improved by running it several times with different initial seed sequences and selecting the best result.

B. Maximizing NMS distances for small l values

In a first experiment, the cost function $\psi_D(\mathbf{x})$ is given as

$$\psi_D(\mathbf{x}) = 6D_1^2(\mathbf{x}) + 3D_2^2(\mathbf{x}) + 2D_3^2(\mathbf{x}) + D_6^2(\mathbf{x}), \tag{10}$$

to achieve a trade-off between maximization of NMS distance with truncation lengths of $l = 1$, 2, 3 and 6. The best C4-sequence $\mathbf{x}_D$ obtained with this cost function has a score of $\psi_D(\mathbf{x}_D) = 2.85$. The values of $D_l^2(\mathbf{x}_D)$ for the targeted truncation lengths l are provided in Table I.

C. Optimization of the Mutual-information for $l = 1$

In a second experiment, we utilize the cost function $\psi_I(\mathbf{x}, \text{SNR})$ defined as the MI obtained with the C4-sequence with truncation length $l = 1$ in the complex AWGN channel with a Signal-to-Noise Ratio (SNR). Since this sequence comprises the set of q distinct points that support the C4-sequence, it is termed the C4-constellation. The evaluation of the cost function $\psi_I(\mathbf{x}, \text{SNR})$ is achieved using the method described in [8].

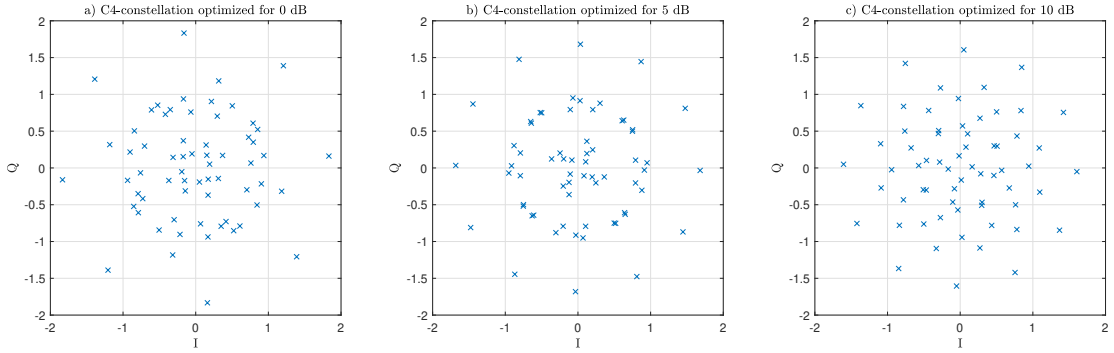

Fig. 3. Constellations associated to C4-sequences optimized for 0 dB, 5 dB and 10 dB of SNR.

Algorithm 2 Greedy search algorithm

Input An initial seed vector $\mathbf{s}$ of size $p = q/4$, the nature of the C4-sequence $\mathbf{x}$ to optimize (clockwise ($c = 1$) or counter-clockwise ($c = -1$)) and the objective function $\psi(\mathbf{x})$ to be maximized.

Output Optimized C4-sequence $\mathbf{x}_{opt}$ for the objective function $\psi(\mathbf{x})$.

```
θ ← p; % Initial step of the greedy search.
s_opt ← s; x_opt ← G(s_opt); ψ_opt ← ψ(x_opt);
while θ > 2^{-4} do
    improved ← true;
    while improved do
        improved = false;
        for i = 0 to p − 1 do
            s_n ← s;
            s_n(i) ← s(i) − θ; x_n ← G(s_n);
            if ψ(x_n) > ψ_opt then
                improved ← true; ψ_opt ← ψ(x_n);
                s_opt ← s_n; x_opt ← x_n;
            end if
            s_n(i) ← s(i) + θ; x_n = G(s_n);
            if ψ(x_n) > ψ_opt then
                improved = true; ψ_opt ← ψ(x_n);
                s_opt ← s_n; x_opt ← x_n;
            end if
            s ← s_opt;
        end for
    end while
    θ ← θ/2;
end while
Return x_opt
```

Fig. 3 shows 3 different C4-constellations optimized at different SNR. Fig. 3.a) shows the obtained constellation optimized at 0 dB of SNR. For this SNR, the MI is 0.9998 bit/s/Hz, very close to the channel capacity of 1 bit/s/Hz. Note that 8 points are regularly distributed around an outer circle. Fig. 3.b) shows the obtained constellation optimized for 5 dB of SNR. For this constellation, the achieved MI is 2.0536 bit/s/Hz, again very close to the channel capacity of 2.057 bit/s/Hz. The shape of the constellation is different from the 0 dB case (outer circle with 12 points). Finally, Fig. 3.c) shows the optimized constellation for 10 dB of SNR, with a MI of 3.4192 bit/s/Hz for a channel capacity of 3.4594 bit/s/Hz. Again, the shape of the constellation is different, with the dots distributed more regularly in space.

D. Optimization of the PAPR: Unitary C4-sequence

Peak-to-average power ratio (PAPR) can be an important feature of a low-cost/low-power wireless communication scheme. In fact, a low PAPR is desirable because it allows the RF power amplifier to be used at its maximum efficiency. In the case of a C4-sequence $\mathbf{x}$, the average energy is one by construction, so the PAPR $\beta(\mathbf{x})$ of $\mathbf{x}$ is defined as $\beta(\mathbf{x}) = \max_{i=0,1,\ldots,q-1}(|x(i)|^2)$. Thus, the minimum PAPR is obtained when all points of the C4-sequence are on the unit circle: in this case, $\beta(\mathbf{x}) = 1$ (unitary C4-sequence). Using the heuristic optimization process with the cost function $\psi_U(\mathbf{x}) = -\beta(\mathbf{x})$ allows generating C4-sequences with a PAPR slightly higher than 1 (typically 1.12), but not to reach the minimum value. However, in [4] a formal technique is proposed to construct unitary C4-sequences directly with PAPR equal to 1. This result shows that the implemented greedy search algorithm (Algo. 2) is not optimal.

The construction of a unitary C4-sequence $\mathbf{x}_u$ can be obtained as $\mathbf{x}_u = G(\mathbf{s}_u)$ (see Algo. 1), where $\mathbf{s}_u$ is a constrained seed sequence, called a unitary seed sequence (since it generates a unitary C4-sequence). For $q = 2^{2t}$, (with t a positive integer) the k^{th} element $s_u(k)$, $k = 0, 1, \ldots, 2^{2t-2}$ of the unitary seed vector $\mathbf{s}_u$ can be defined by

$$s_u(k) = d(r_k) + q_k\gamma(r_k)2^{t+1},\ k = 0, 1, \ldots, 2^{2t-2} - 1, \quad (11)$$

with $r_k = k \bmod 2^{t-1}$ and $q_k = (k - r_k)/2^{t-1}$ (thus, $k = q_k 2^{t-1} + r_k$), $\mathbf{d}$ a real vector of size 2^{t-1} taking its values in the interval $[0, q[$ and γ a permutation over the set $\{0, 1, \ldots, 2^{t-1} - 1\}$.

Theorem 3.1: If $\mathbf{s}_u$ is a unitary sequence defined by (11), then $\mathbf{x}_u = G(\mathbf{s}_u)$ is a unitary C4-sequence. Moreover, the n^{th} term of $\mathbf{x}_u$ is given as

$$x_u(n) = \exp\left(\frac{2\pi j(n + d(\gamma^{-1}(-n)) + 4n\gamma^{-1}(-n))}{q}\right), \quad (12)$$

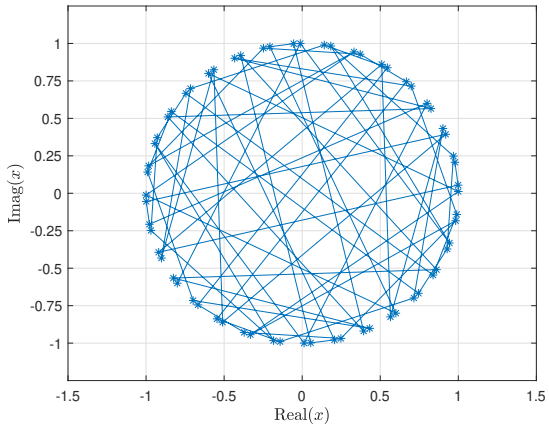

Fig. 4. Optimized Unitary C4-sequence $\mathbf{x}_U$.

with $0 \leq \gamma^{-1}(-n) < 2^{t-1}$ the unique solution of the equation $\gamma(\gamma^{-1}(-n)) = -n \mod 2^{t-1}$.

Proof: The proof of theorem 3.1 is given in APPENDIX III □

While a seed sequence $\mathbf{s}$ has $p = q/4$ degrees of freedom (the dimension of the vector $\mathbf{s}$), a unitary seed sequence is more constrained, since it is itself generated by a vector $\mathbf{d}$ of size 2^{t-1} (and a permutation γ) when $q = 2^{2t}$. Note that when $q = 2^{2t+1}$, the degrees of freedom of $\mathbf{s}_u$ are respectively 2 and 4 for $q = 32$ and $q = 128$, respectively [4]. However, for the sizes $q = 2^{2t+1}$ with $t > 3$, the general construction of a unitary C4-sequence from a unitary seed sequence is still an open problem.

Thus, it is possible to adapt Algo. 2 to also optimize unitary C4-sequences by replacing modifications on the seed vector $\mathbf{s}$ directly by modifications on the vector $\mathbf{d}$, which generates a unitary seed vector $\mathbf{s}_u$. For example, using the objective function $\psi_D(\mathbf{x})$ (see section III.B), we obtained the unitary C4-sequence $\mathbf{x}_U$ of length $q = 64$ shown in Fig. 4, with $\psi(\mathbf{x}_U) = 2.791$ ($\mathbf{s}_u$ is generated with $\mathbf{d} = [0.445, 37.878, 16.445, 61.878]$, $\gamma = \{1, 2, 3, 0\}$) and $c = -1$). The NMS distances $D_l^2(\mathbf{x}_U)$ for the targeted truncation lengths are also given in the Table I.

Finally, it is worth mentioning that the ability to construct unitary C4-sequences provides an alternative/complement to the well-known ZC [5], [6] sequences used in the Physical Random Access Channel (PRACH) of the 3GPP standard [9] and the CSS sequence used in LoRaWAN [7].

The next section presents the association of a truncated-C4-sequence (T-C4-sequence) with an outer non-binary code.

IV. Truncated C4-sequences for single-user applications

A T-C4-sequence can be used alone to encode a few bits message. It can also be used in a concatenated coding scheme as an inner coder combined with an outer non-binary code.

A. Modulation with T-C4-sequence

As mentioned in the introduction, a length $q = 2^m$ C4-sequence, and its truncated version, can be used to transmit m bits of information $(b_{m-1}, b_{m-2}, \ldots, b_1, b_0)$ by using the length l sequence $\mathbf{x}_M^{l+M}$ as the modulation sequence, with the integer index $M = \sum_{i=0}^{m-1} 2^i b_i$. The truncation length l allows to adjust the spectral efficiency to the channel conditions. Table I characterizes the C4-sequences $\mathbf{x}_D$ and $\mathbf{x}_U$ defined in sections III.B and III.D, respectively. It gives the NMS distance $D_l^2(\mathbf{x})$ for truncation lengths $l = 1$, 2, 3 and 6, as well as the resulting MI (in bit/s/Hz) in the AWGN channel. These characteristics are compared with the use of standard modulations, denoted by the generic term $\mathbf{s}$, to transmit 6 bits of information. For instance, $\mathbf{s}$ refers to 64-QAM for $l = 1$, a pair of 8-PSK symbols for $l = 2$, a triplet of QPSK symbols for $l = 3$, and 6 BPSK symbols for $l = 6$.

TABLE I
COMPARISON OF CLASSICAL MODULATIONS WITH A T-C4-SEQUENCES FOR TRANSMITTING A LENGTH-6 MESSAGE. THE MI IS EXPRESSED IN BIT/S/HZ.

	$l=1$	$l=2$	$l=3$	$l=6$
$D_l^2(\mathbf{s})$	0.0952	0.2929	0.6667	0.6667
$D_l^2(\mathbf{x}_D)$	0.0120	0.2984	0.4621	0.9506
$D_l^2(\mathbf{x}_U)$	0.0018	0.1877	0.5546	1.1077
MI($\mathbf{s}$) @ 0 dB	0.992	1.962	2.916	4.329
MI($\mathbf{x}_D$) @ 0 dB	0.997	1.975	2.908	4.929
MI($\mathbf{x}_U$) @ 0 dB	0.981	1.957	2.925	4.951
MI($\mathbf{s}$) @ 5 dB	1.993	3.724	5.155	5.857
MI($\mathbf{x}_D$) @ 5 dB	2.029	3.855	5.143	5.984
MI($\mathbf{x}_U$) @ 5 dB	1.863	3.665	5.167	5.987
MI($\mathbf{s}$) @ 10 dB	3.269	5.355	5.981	≈ 6
MI($\mathbf{x}_D$) @ 10 dB	3.311	5.504	5.975	≈ 6
MI($\mathbf{x}_U$) @ 10 dB	2.746	5.167	5.980	≈ 6

According to Table I, for a truncation length of $l = 1$, the minimum distance of the classical constellation (i.e. a 64-QAM) is significantly higher than for the C4-sequence $\mathbf{x}_D$ and $\mathbf{x}_U$, however, this higher distance does not directly translate to a higher MI for $\mathbf{x}_D$. The advantage of the C4-sequence becomes predominant for $l = 6$, as both the NMS-distance and MI significantly surpass those of classical constellations.

B. Principle of concatenated scheme

As proposed in [2], [3], it is feasible to concatenate an outer Non-Binary Error Correcting Code (NB-ECC) over GF(q) with an inner code composed of a Truncated-C4-sequence, as shown in Fig. 5. In this configuration, the outer code takes k GF(q) symbols (i.e., m-tuple binary vector, with $m = \log_2(q)$) to generate n GF(q) symbols (coding rate $R = k/n$). Each of the GF(q) symbols is then employed to modulate a Truncated-C4-sequence of length l, where l is a parameter that allows flexibility to precisely match the spectral efficiency to the channel condition. The total spectral efficiency of this coding scheme is thus

$$S(l) = \frac{Rm}{l} \text{ bit/s/Hz.} \tag{13}$$

The decoding of the T-C4-sequence involves computing the Log Likelihood Ratio (LLR) for all the q possible codewords of the T-C4-sequence sequence, based on the length-l received sequence. These LLRs are then utilized by the outer code to recover the transmitted message.

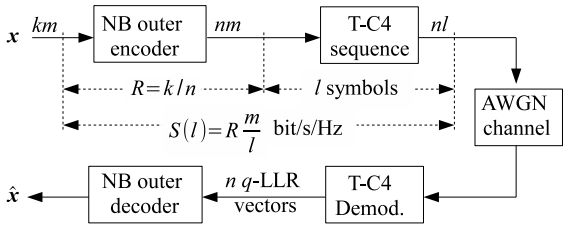

Fig. 5. Concatenation of an NB outer code and a Truncated-C4-sequence.

C. Asymptotic spectral efficiency

In this section, the asymptotic spectral efficiency of a T-C4-sequence of length $q = 2^m$ is analyzed when it is combined with an external non-binary code of rate R. To ensure that the outer code can reliably decode the transmitted codeword with an arbitrary low probability of error, the average amount of received information associated with a transmitted codeword should be greater than the information contained in the message itself. Consider a code rate of R, where a message of size k contains km bits of information. The length of the encoded message is $n = k/R$ m-ary symbols. The average MI between a transmitted T-C4-sequence of length l and the received sequence through the AWGN channel at a given SNR is defined as $\mu_{\text{SNR}}(l)$. The maximum amount of information available for the outer code is then $\mu_{\text{SNR}}(l)n$. To ensure reliable decoding, $\mu_{\text{SNR}}(l)n$ is required to be greater than km, which gives us the following relation between the outer code rate R and the MI $\mu_{\text{SNR}}(l)$

$$R < \frac{km}{\mu_{\text{SNR}}(l)n}. \tag{14}$$

The minimum MI bound μ_{SNR} is thus given as $\mu_{\text{SNR}} = Rm$. Since $\mu_{\text{SNR}}(l) > \mu_{\text{SNR}}$, the spectral efficiency $S_{\text{SNR}}(l)$ is upper-bounded by $S_{\text{SNR}} = Rm/l$ bits/s/Hz.

D. Estimation of the maximum spectral efficiency

Let's consider an AWGN channel with a fixed SNR (the subscript "SNR" is omitted in the following). Also assume a constant rate R outer code in the asymptotic mode. The aim is to determine the maximum spectral efficiency achievable with a T-C4-sequence. To achieve the required MI μ at the receiver side, the minimum truncation length l_m is given as

$$l_m = \arg\min_l \{\mu(l) \geqslant \mu\}. \tag{15}$$

When l_m goes from 1 to 2, the spectral efficiency is immediately divided by 2. This quantization effect is detrimental to the precise adaptation of the spectral efficiency to the channel condition. To solve this problem, the method proposed in [2] is used. For each SNR, a proportion α of T-C4-sequences of length l_m is mixed with a proportion $(1-\alpha)$ of T-C4-sequences of length $l_m - 1$ so that $\alpha\mu(l_m) + (1-\alpha)\mu(l_m - 1) = \mu$. Since $\mu(l_m) \geqslant \mu > \mu(l_m - 1)$, the solution α is unique. The average length is therefore

$$\bar{l}_m = \alpha l_m + (1-\alpha)(l_m - 1), \tag{16}$$

and the associated spectral efficiency is $S(\bar{l}_m) = Rm/\bar{l}_m$.

For a given SNR and a specific truncation length, the MI $\mu(l)$ can be estimated through a Monte Carlo simulation by averaging the received MI over multiple trials (typically 10^5 trials). This approach enables the estimation of asymptotic spectral efficiency of the non-binary code combined with a Truncated-C4-sequence. Figure 6 illustrates the resulting capacity for three different coding rates: $R = 1/3$, $R = 1/2$, and $R = 2/3$. The plot shows that the overall system achieves higher efficiency when the outer code rate is set to $R = 1/3$ compared to $R = 1/2$ or $R = 2/3$. To highlight the results

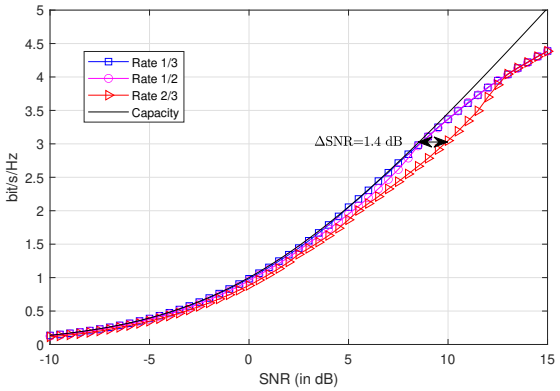

Fig. 6. Capacity of concatenated non-binary code with T-C4-sequences for several coding rates

presented in Fig. 6, Fig. 7 displays the distance ΔSNR (in dB) between the Shannon capacity and the maximum spectral efficiency. The performance of the Zadoff-Chu sequences is also given to illustrate the better performance of the C4-sequence. As shown in Fig. 7, with a fixed outer code rate of

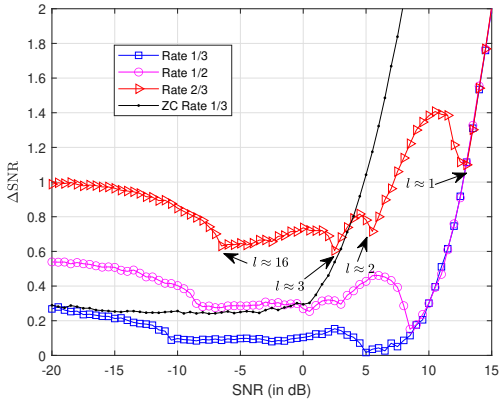

Fig. 7. Distance comparison to the channel capacity between NB-T-C4-sequence and ZC sequence.

1/3, the performance asymptotically approaches the channel capacity of less than 0.2 dB in the range of -15 dB to 10 dB. In addition, the C4-sequence significantly outperforms the ZC sequence. The figure needs further comment. For high SNR, $l_m = 1$ and thus using a sequence of length $l_m - 1$ as in (16) just means that the symbols are punctured. It is interesting to observe that the spectral efficiency for the 3 coding rates

merge at high SNR: they all then follow the spectral efficiency of the C4-constellation alone. Since the C4-sequence used for this simulation is the one optimized for 5 dB, for rate 1/3, the overall spectral efficiency is very close to the channel capacity. Between $l = 1$ and $l = 2$ a degradation is visible (see the "dome shape" for $R = 2/3$), which shows that the mixing technique given in (16), although effective, is not optimal. The same attenuated dome shape also appears between $l = 2$ and $l = 3$.

E. Simulation results with a single parity check of GF(64)

In this section, an outer code consisting of a single parity check of degree 4 over GF(64) is considered. This code allows encoding a message of 3 GF(64) symbols (corresponding to 18 bits of information) into a codeword of 4 GF(64) symbols. For a given truncation length l, the spectral efficiency $S(l)$ of this coding scheme is thus given by $S(l) = \frac{9}{2l}$, as stated in (13).

Fig. 8 shows the performance of the C4-sequences $\mathbf{x}_D$, $\mathbf{x}_U$ and the ZC sequence $\mathbf{c}$ for truncation lengths $l \in \{1, 2, 3, 6, 12, 24, 60\}$, resulting in spectral efficiencies ranging from 4.5 bit/s/Hz ($l = 1$) to 0.075 bit/s/Hz ($l = 60$) as indicated by (13). This concatenated coding scheme can be effectively used to fine-tune the spectral efficiency to the channel condition. It is also well suited for use in a hybrid automatic request communication scheme. In case of a decoding failure, the receiver can request the transmitter to send the subsequent symbols of the truncated sequences to effectively increase the truncation length. Note that the performance of the T-C4-sequence associated with a stronger outer code (a regular non-binary low-density parity check code) is available in [3].

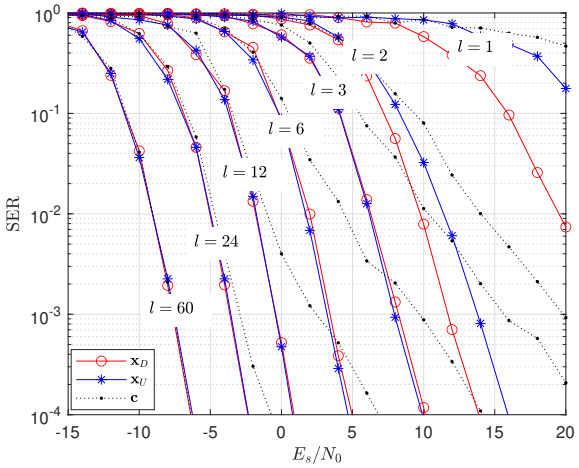

Fig. 8. Performance of an outer code composed of a degree 4 parity check over GF(64) and a Truncated-C4-sequence as the inner code.

Fig. 8 shows that the ZC-sequence and the C4-sequences have equivalent performance for high truncation lengths, but when l is small, the T-C4-sequences outperform the truncated ZC sequence.

V. Distance between C4-sequences for multi-user applications

In the context of a multi-user application, it is interesting to determine the properties of the distance between two different C4-sequences, with or without truncation. This section states the problem and gives some preliminary results. The minimum NMS distance $D_l^2(\mathbf{x}, \mathbf{y})$ between two T-C4-sequences of length l, $\mathbf{x}$ and $\mathbf{y}$, can be defined from (9) as

$$D_l^2(\mathbf{x}, \mathbf{y}) = \frac{1}{l} \min_{a,b} \{ \left\| \mathbf{x}_a^{a+l-1} - \mathbf{y}_b^{b+l-1} \right\|^2 \}. \tag{17}$$

A. Distance between C4-sequences

When the full length sequence are considered, the q-periodicity of the C4-sequences $\mathbf{x}$ and $\mathbf{y}$ transforms (17) to

$$D_q^2(\mathbf{x}, \mathbf{y}) = \frac{1}{q} \min_{\tau} \{ \|\mathbf{x} - \mathbf{y}_\tau\|^2 \}. \tag{18}$$

$$= \frac{1}{q} (\|\mathbf{x}\|^2 + \|\mathbf{y}_\tau\|^2 - 2 \max_{\tau} \{ \mathcal{R}(\langle \mathbf{x}, \mathbf{y}_\tau \rangle) \}). \tag{19}$$

Let $\mathbf{R}_{\mathbf{x},\mathbf{y}}$ be the inter-correlation vector between $\mathbf{x}$ and $\mathbf{y}$ defined as $\mathbf{R}_{\mathbf{x},\mathbf{y}} = (\langle \mathbf{x}, \mathbf{y}_\tau \rangle)_{\tau=0,1,\ldots,q-1}$. Therefore, (19) is equivalent to

$$D_q^2(\mathbf{x}, \mathbf{y}) = 2 - \frac{2}{q} \max\{ \mathcal{R}(\mathbf{R}_{\mathbf{x},\mathbf{y}}) \}. \tag{20}$$

The computation of the inter-correlation function in the frequency domain gives

$$\mathbf{R}_{\mathbf{x},\mathbf{y}} = \mathcal{F}^{-1}(\mathbf{X} \odot \mathbf{Y}')\}, \tag{21}$$

with $\mathbf{X} = \mathcal{F}(\mathbf{x})$ and $\mathbf{Y} = \mathcal{F}(\mathbf{y})$. By defining

$$\mathbf{Z} = \frac{1}{\sqrt{4q}} (\mathbf{X} \odot \mathbf{Y}'), \tag{22}$$

and $\mathbf{z} = \mathcal{F}^{-1}(\mathbf{Z})$, we get $\mathbf{R}_{\mathbf{x},\mathbf{y}} = \sqrt{4q}\ \mathbf{z}$. Thus, (19) can be reformulated as

$$D_q^2(\mathbf{x}, \mathbf{y}) = 2 - \frac{4}{\sqrt{q}} \max\{ \mathcal{R}(\mathbf{z}) \}). \tag{23}$$

In summary, the study of the minimum square distance $D_q^2(\mathbf{x}, \mathbf{y})$ boils down to studying the vector $\mathbf{z}$, or more precisely, to studying its maximum real value $\max\{\mathcal{R}(\mathbf{z})\}$. There are two distinct cases: either $\mathbf{x}$ and $\mathbf{y}$ belong to the same type of C4-sequences (i.e., both clockwise or both counter-clockwise), or they do not.

In the latter case, let us assume, without loss of generality, that $\mathbf{x}$ is a clockwise C4-sequence and $\mathbf{y}$ a counter-clockwise C4-sequence. According to (3), $\mathbf{X}$ has non-null values at positions k where $k + 1 \bmod 4 = 0$, while $\mathbf{Y}$ has non-null values at positions k' where $k' - 1 \bmod 4 = 0$. This implies that $\mathbf{Z}$ equals the null vector (see (22)), and thus, $\mathbf{z}$ is also the null vector. In this case, $\max \mathcal{R}(\mathbf{z}) = 0$, and thus, $D_q^2(\mathbf{x}, \mathbf{y}) = 2$.

When both C4-sequences $\mathbf{x}$ and $\mathbf{y}$ are either clockwise or counter-clockwise, the situation is less favorable. Without loss of generality, let us consider two clockwise C4-sequences in the sequel. According to (3) and (22), $\mathbf{Z}$ verifies

$$\begin{cases} |Z(k)|^2 & = \left(\frac{4q}{\sqrt{4q}}\right)^2 = 4q \text{ if } k+1 \bmod 4 = 0, \\ & = 0 \text{ otherwise.} \end{cases} \quad (24)$$

Thus, according to Theorem 2.2, $\mathbf{z}$ is a C4-sequence. The determination of the minimum maximum value of the real part of a C4-sequence is an open problem. However, if $\mathbf{z}$ is a unitary C4-sequence, then $\max\{\mathcal{R}(\mathbf{z})\} \leq 1$, and thus, $D_q^2(\mathbf{x},\mathbf{y}) \geq 2(1-\frac{2}{\sqrt{q}})$. Note that the relationship between $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ in (22) can be transformed. One way is to express $\mathbf{Y}$ as a function of $\mathbf{X}$ and $\mathbf{Z}$ by multiplying both terms of the equation by $\frac{1}{2q}\mathbf{X}$. Thus, from any sequence $\mathbf{X}$, it is possible to construct a sequence $\mathbf{Y}$ that satisfies $D_q^2(\mathbf{x},\mathbf{y}) \geq 2(1-\frac{2}{\sqrt{q}})$ by choosing $\mathbf{z}$ as the unitary sequence.

To conclude, this section presents two preliminary results about the NMS-distance between two C4-sequences of length q: If one C4-sequence is clockwise and the other is counter-clockwise, then the NMS-distance between the two sequences is 2. Considering two C4-sequences of the same type, it is possible to guarantee an NMS-distance of $2(1-\frac{2}{\sqrt{q}})$. However, the determination of the minimum distance in a set of more than two C4-sequences remains an open problem. In the next section, the generalization of C4-sequences is presented.

VI. Generalization of C4-sequences

The construction method for C4-sequences can be generalized to generate C3 or C5 sequences, and more generally to generate any Cn sequences, where n is a strictly positive integer, by modifying Algorithm 1 to Algorithm 3.

Algorithm 3 Generation of a Cn-sequence of length q by the function $\mathbf{x} = G(\mathbf{s})$

Input A seed vector $\mathbf{s}$ of size $p = q/n$ composed of q/n reals on the interval $[0, q[$, the order n of the Cn-sequence and index c, $1 < c < n$ co-prime with n.
Output A Cn-sequence $\mathbf{x}$ of length q

```
for k ← 0 to q/n − 1 do
    E_s(k) ← √(nq) × exp(2πj s(k)/q)
end for
M ← [0, 0, ..., 0]; % null vector of size n
M(c) ← 1
X ← kron(E_s, M)
x ← F^{-1}(X)
Return x
```

The study of Cn constellations gives sequences with interesting properties. Fig. 9.a gives an example of a C3-sequence, Fig. 9.b its associated NMS-Distance and Fig. 9.c its autocorrelation function (see the 3-branch star shape). The C3-sequence possesses a 3-fold symmetry and its NMS-Distance gets the optimal value for 3 different truncation lengths, $l = q/3$, $l = 2q/3$ and $l = q$. Fig. 10 gives an example of a C5-sequence. The C5-sequence possesses a 5-fold symmetry (see Fig. 10.a), an autocorrelation function with a 5-branch star shape (see Fig. 10.c). Finally, the maximum NMS-distance for the C5-sequence is lower than 2 and equals to $\alpha = 2(1-\cos(2\pi/5)) = 1.382$. The NMS-Distance equals α for truncation lengths of $l = q/5, 2q/5, 3q/5, 4q/5$, and q (see Fig. 10.b). In the general case, the NMS-distance of a Cn sequence takes its maximum value $\alpha = \min(2, 2(1-\cos(2\pi/n))$ on truncation lengths $l = kq/n$, with $k = 1, 2, \ldots, n$.

VII. Conclusion

This paper defines the notion of C4-sequences. C4-sequences share similar optimal autocorrelation properties with ZC or CSS sequences. However, C4-sequences offer the additional advantage of having optimal properties (in terms of minimum Euclidean distance between sequences) for several truncation lengths. Moreover, unlike ZC sequences, they are not restricted to having their points on the unitary circle.

C4-sequences can have several applications in a communication system. First, they can be used alone as an alternative to classical sequences (ZC sequences, CSS sequences). Second, the constellation associated with a C4-sequence can be shaped to maximize the MI through the AWGN channel, thus providing a geometric shaping gain. Third, the concatenation of a non-binary outer code with a T-C4-sequence as an inner code represents a very efficient and flexible communication scheme. While the outer code is fixed, the choice of the truncation length provides a versatile tool to closely adapt the overall coding rate to the channel conditions. It's worth mentioning that this flexibility can be effectively exploited in a hybrid automatic request (H-ARQ) communication system. Fourth, they offer a high degree of freedom in their construction, which can be exploited for multi-user applications. Finally, C4-sequences can be extended to C3-sequences or C5-sequences.

Acknowledgment

This work has been funded by the french ANR-21-CE25-0006 (https://ai4code.projects.labsticc.fr/). The author would like to thank Cédric Marchand and Alexandru Olteanu for their indirect, but fundamental, contribution: their machine learning development to optimize truncated CCSK sequences gives the mathematical clues to construct the C4-sequences.

References

[1] G.M. Dillard, M. Reuter, J. Zeiddler, and B. Zeidler. Cyclic code shift keying: a low probability of intercept communication technique. *IEEE Transactions on Aerospace and Electronic Systems*, 39(3):786–798, 2003.

[2] Cédric Marchand and Emmanuel Boutillon. Rate-adaptive Inner Code for Non-Binary Decoders. In *2021 11th International Symposium on Topics in Coding (ISTC)*, pages 1–5, 2021.

[3] Cédric Marchand and Emmanuel Boutillon. Rate-adaptive cyclic complex spreading sequence for Non-Binary Decoders. In *International Symposium on Topics in Coding (ISTC'2023), Brest*, 2023.

[4] Emmanuel Boutillon. C4-Sequences: Rate Adaptive Coded Modulation for Few Bits Message. In *International Symposium on Topics in Coding (ISTC'2023), Brest*, 2023.

[5] D. Chu. Polyphase codes with good periodic correlation properties (Corresp.). *IEEE Transactions on Information Theory*, 18(4):531–532, 1972.

[6] R. Frank, S. Zadoff, and R. Heimiller. Phase shift pulse codes with good periodic correlation properties (Corresp.). *IRE Transactions on Information Theory*, 8(6):381–382, 1962.

[7] Lora-Alliance. LoRaWAN TM 101 A Technical Introduction. http://www.lora-alliance.org. Technical Marketing Workgroup 1.0, November 2015, Accessed: 2024-04-23.

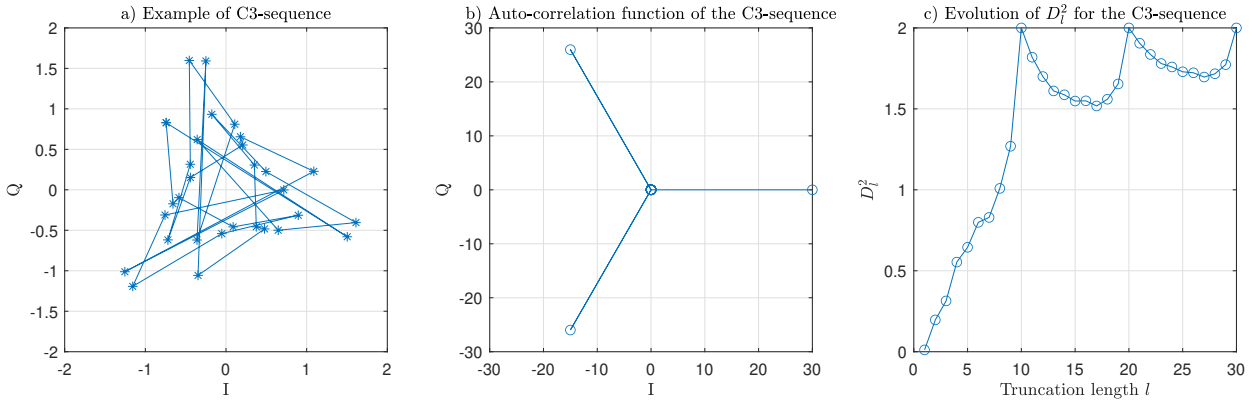

Fig. 9. Example of C3-sequence

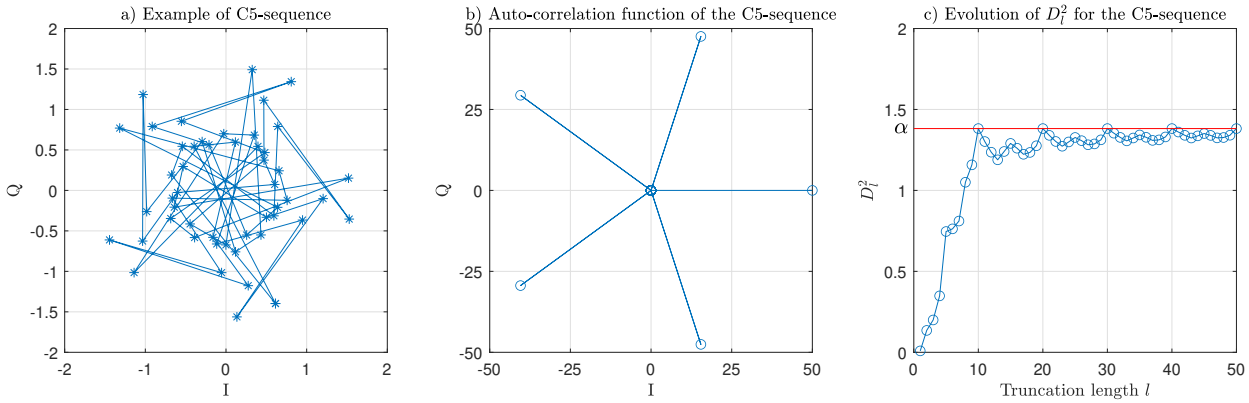

Fig. 10. Example of C5-sequence

[8] G. Foschini, R. Gitlin, and S. Weinstein. Optimization of Two-Dimensional Signal Constellations in the Presence of Gaussian Noise. *IEEE Transactions on Communications*, 22(1):28–38, 1974.

[9] 5G; NR; Physical channels and modulation (3GPP TS 38.211 version 16.7.0 Release 16). https://www.etsi.org/deliver/etsi_ts/138200_138299/138211/16.07.00_60/ts_138211v160700p.pdf. Accessed: 2024-04-11.

APPENDIX I: PROOF OF LEMMA 2.4

Let us consider a C4-sequence $\mathbf{x}$ of length q. From $\mathbf{x}$, we can define the vector $\bar{\mathbf{x}}$ defined as:

$$\bar{\mathbf{x}} = \text{kron}([1, j^{-c}, j^{-2c}, j^{-3c}], \mathbf{x}_0^{p-1}). \tag{25}$$

From the definition of $\bar{\mathbf{x}}$, we get, for all $n = 0, 1, \ldots, p-1$ and $k = 0, 1, 2$ and 3, $\bar{x}(n+kp) = j^{-kc}x(n)$, thus, showing $\bar{\mathbf{x}} = \mathbf{x}$ will achieve the proof.

Let us study $\bar{\mathbf{X}} = \mathcal{F}(\bar{\mathbf{x}})$, the DFT of vector $\bar{\mathbf{x}}$. The u^{th} element $\bar{X}(u)$ is (remind that $e^{2\pi j \frac{p}{q}} = j$)

$$\bar{X}(u) = \sum_{n=0}^{q-1} \bar{x}(n) e^{-2\pi j \frac{un}{q}} \tag{26}$$

$$= \sum_{n=0}^{p-1} \sum_{k=0}^{3} \bar{x}(n+pk) e^{-2\pi j \frac{u(n+kp)}{q}} \tag{27}$$

$$= \sum_{n=0}^{p-1} x(n) e^{-2\pi j \frac{un}{q}} \sum_{k=0}^{3} j^{-(c+u)k} \tag{28}$$

The terms $w(u) = \sum_{k=0}^{3} j^{-(c+u)k}$ is only a function of u, with $w(u) = 4$ if $(u+c) \bmod 4 = 0$, 0 otherwise. Thus,

$$\bar{X}(u) = 0 \text{ when } (u+c) \bmod 4 \neq 0. \tag{29}$$

According to theorem 2.2, the DFT $\mathbf{X}$ of $\mathbf{x}$ verifies also the same condition as (29). Thus, the linearity of the DFT operator applied to $\delta = \mathbf{x} - \bar{\mathbf{x}}$, gives $\Delta = \mathcal{F}(\delta)$ verifying

$$\Delta(u) = 0 \text{ when } (u+c) \bmod 4 \neq 0. \tag{30}$$

Moreover, by construction, the first $q/4$ first coordinates of δ are equal to 0, thus the set of $3p$ equations $\Delta(u) = 0$, for u verifying $(u+c) \bmod 4 \neq 0$ gives

$$\sum_{n=p}^{q-1} \delta(n) e^{-2\pi j \frac{un}{q}} = 0. \tag{31}$$

The set of $3p$ equations (31) can be expressed in a matrix form as $\mathbf{V}\delta_p^{q-1} = 0$ with $\mathbf{V}$ the $3p \times 3p$ matrix obtained from the DFT matrix $\mathbf{U} = (e^{\frac{-2\pi jnu}{q}})_{0 \leq n,u < q}$ by pruning the first p lines and the p columns of indices u verifying $u + c = 0 \bmod 4$ of $\mathbf{U}$. Let $\mathbf{I}_p$ be the diagonal matrix of size $p \times p$, one can verify that $\mathbf{V} \times \mathbf{V}'$ gives the matrix

$$\mathbf{V} \times \mathbf{V}' = p(\mathbf{H} \otimes \mathbf{I}_p), \tag{32}$$

with $\otimes$ the Kronecker product and $\mathbf{H}$ the 3×3

$$\mathbf{H} = \begin{bmatrix} 3 & -j & 1 \\ j & 3 & -j \\ 1 & j & 3 \end{bmatrix} \tag{33}$$

Since $\mathbf{H}$ is invertible (its determinant is equal to 16), $\mathbf{V} \times \mathbf{V}'$ is an invertible matrix, and thus, $\mathbf{V}$ is also invertible. The equation $\mathbf{V}\delta_{q/4}^{q-1} = 0$ has thus a unique null solution $\delta_{q/4}^{q-1} = 0$. In summary, all the components of the vector δ are equal to zero. Thus, $\bar{\mathbf{x}} = \mathbf{x}$, which achieve the proof $\square$

APPENDIX II: PROOF OF THEOREM 2.6

Let us first give the proof in the case of $l = q$. Due to the q-periodicity of the $\mathbf{x}$ C4-sequence

$$\begin{aligned} \|\mathbf{x}_a - \mathbf{x}_b\|^2 &= \|\mathbf{x} - \mathbf{x}_{b-a}\|^2 \\ &= \|\mathbf{x}\|^2 + \|\mathbf{x}_{b-a}\|^2 - 2\mathcal{R}(\mathbf{R}_{\mathbf{xx}}(b-a)) \\ &= 2q - 2\mathcal{R}(\mathbf{R}_{\mathbf{xx}}(b-a)) \end{aligned} \tag{34}$$

By definition, when $a \neq b$, $\tau = b - a \neq 0$. By definition of the C4-sequence, (4) gives $\mathbf{R}_{\mathbf{xx}}(\tau) = 0$ when $\tau \in [\![1, q]\!] - \{2p\}$, and $\mathbf{R}_{\mathbf{xx}}(\tau) = -q$ when $\tau = 2p$, which proves the theorem for the value of $l = q$. The demonstration for the cases $l = p$ is obtained by considering the computation of $\|\mathbf{x}_a - \mathbf{x}_b\|^2$ using the 4-fold symmetry property of $\mathbf{x}$. According to Lemma 2.4, for any integer k, $|x(n+kp+a) - x(n+kp+b)|^2 = |j^{-kc}x(n+a) - j^{-kc}x(n+kp+b)|^2 = |x(n+a) - x(n+b)|^2$, thus, for any couple (a, b),

$$\begin{aligned} \|\mathbf{x}_a - \mathbf{x}_b\|^2 &= \sum_{n=0}^{q-1} |x(n+a) - x(n+b)|^2 \\ &= \sum_{n=0}^{p-1} \sum_{k=0}^{3} |x(n+kp+a) - x(n+kp+b)|^2 \\ &= 4 \sum_{n=0}^{p-1} |x(n+a) - x(n+b)|^2 \\ &= 4 \left\| \mathbf{x}_a^{a+p-1} - \mathbf{x}_b^{b+p-1} \right\|^2 \end{aligned} \tag{35}$$

Thus, (35) shows that $\left\| \mathbf{x}_a^{a+p-1} - \mathbf{x}_b^{b+p-1} \right\|^2$ equals $2q/4 = 2p$ when $a - b \neq q/2$, and is greater than $2p$ otherwise, which proves the theorem for $l = p$. Finally, the proof for $l = 2p$ and $l = 3p$ can be obtained similarly than the case $l = p$ $\square$

APPENDIX III: PROOF OF THEOREM 3.1

Since the sequence $\mathbf{s}_u$ is real, $\mathbf{x}_u = G(\mathbf{s}_u)$ is a C4-sequence by construction. Let us compute explicitly the n^{th} terms $x_u(n)$, of $\mathbf{x}_u$ using the function $G(\mathbf{s}_u)$ given in algorithm 1. By definition of the IDFT, we get

$$x_u(n) = \frac{1}{q} \sum_{k=0}^{q-1} X_u(k) e^{\frac{2\pi jkn}{q}} \tag{36}$$

.

By construction of $\mathbf{x}$, all the terms not congruent to 1 modulo 4 are equal to 0. Moreover, by construction, $X_u(4k+1) = \sqrt{4q} e^{\frac{2\pi j s_u(k)}{q}}$ (see equation (3)), thus (36) is equivalent to

$$x_u(n) = \frac{1}{q} \sum_{k=0}^{2^{2t-2}-1} \sqrt{4q} e^{\frac{2\pi j s_u(k)}{q}} e^{\frac{2\pi j(4k+1)n}{q}}. \tag{37}$$

The factors independent of index k in (37) can be factorized (note that, $\frac{\sqrt{4q}}{q} = \frac{2^{t+1}}{2^{2t}} = \frac{1}{2^{t-1}}$), giving

$$x_u(n) = \frac{e^{\frac{2\pi jn}{q}}}{2^{t-1}} \sum_{k=0}^{2^{2t-2}-1} e^{\frac{2\pi j(s_u(k)+4kn)}{q}}. \tag{38}$$

.

It is possible to decompose the k indices as $k = 2^{t-1}q_k + r_k$, with $0 \leq q_k < 2^{t-1}$, $0 \leq r_k < 2^{t-1}$, then, to replace $s_u(2^{t-1}q_k + r_k)$ by its expression in (11), giving

$$\begin{aligned} x_u(n) &= \frac{e^{\frac{2\pi jn}{q}}}{2^{t-1}} \sum_{r_k=0}^{2^{t-1}-1} \sum_{q_k=0}^{2^{t-1}-1} e^{\frac{2\pi j(d(r_k)+q_k\gamma(r_k)2^{t+1}+4(2^{t-1}q_k+r_k)n)}{q}} \\ &= \frac{e^{\frac{2\pi jn}{q}}}{2^{t-1}} \sum_{r_k=0}^{2^{t-1}} e^{\frac{2\pi j(d(r_k)+4r_kn)}{q}} A_t(r_k, n) \end{aligned} \tag{39}$$

.

with

$$A_t(r_k, n) = \sum_{q_k=0}^{2^{t-1}-1} e^{\frac{2\pi jq_k(\gamma(r_k)2^{t+1}+2^{t+1}n)}{q}} \tag{40}$$

$$= \sum_{q_k=0}^{2^{t-1}-1} e^{\frac{2\pi jq_k(\gamma(r_k)+n)}{2^{t-1}}} \tag{41}$$

$$= \sum_{q_k=0}^{2^{t-1}-1} \rho^{q_k}, \tag{42}$$

with $\rho = e^{\frac{2\pi j(\gamma(r_k)+n)}{2^{t-1}}}$. Let us define $\gamma^{-1}(-n)$ the unique solution to the equation $\gamma(r_k) = -n \bmod 2^{t-1}, r \in \{0, 1, \ldots, 2^{t-1} - 1\}$. If $r_k = \gamma^{-1}(-n)$, then $\gamma(r_k) + n = 0 \mod 2^{t-1}$), which gives $\rho = 1$, and thus, $A_t(r_k, n) = 2^{t-1}$. Otherwise, if $r \neq \gamma^{-1}(-n)$, then $\rho \neq 1$, and thus,

$$A_t(r_k, n) = \frac{1 - \rho^{2^{t-1}}}{1 - \rho}, \tag{43}$$

which gives $A_r(r_k, n) = 0$ since $\rho^{2^{t-1}} = e^{2\pi j(\gamma(r_k)+n)} = 1$.

Going back to (39), we obtain

$$x_u(n) = e^{\frac{2\pi j(n+d(\gamma^{-1}(-n))+4\gamma^{-1}(-n)n)}{q}}; \tag{44}$$

Thus, for any n, $x(n)$ belongs to the unitary circle $\square$