Improved DC-Free Run-Length Limited 4B6B Codes for Concatenated Schemes

ELIE NGOMSEU MAMBOU1, (Member, IEEE), THIBAUD TONNELLIER2, AND WARREN J. GROSS2, (Senior Member, IEEE)
1Department of Electrical and Electronic Engineering, University of Johannesburg, Auckland Park 2006, South Africa
2Department of Electrical and Computer Engineering, McGill University, Montréal, QC H3A 0E9, Canada
Corresponding author: Elie Ngomseu Mambou (emambou@uj.ac.za)

ABSTRACT In this letter, we introduce a class of improved DC-free 4B6B codes in terms of error correction capabilities for a serially concatenated architecture. There are billions of different codebooks that can be derived from the 16 codewords contained in the traditional 4B6B code as per the IEEE 802.15.7 standard for visible light communication (VLC). These codebooks can be classified based on distances properties which determine their error correction performances. The traditional 4B6B code is suitable for hard-decision decoding, however, when a soft decoder is used like in a serially concatenated architecture, that code becomes obsolete. Simulations show that the proposed 4B6B code concatenated with forward error correction (FEC) codes, has better performance compared to state-of-the-art schemes such as the original 4B6B code, the enhanced Miller code, the Manchester code, the 5B10B code and the (0,4) 2/3 RLL code.

INDEX TERMS DC-free, maximum likelihood decoding, run-length limited, 4B6B code, Manchester code, on-off keying modulation, visible light communication.

I. INTRODUCTION
Visible light communication (VLC) consists of transmitting information through light by using light emitting diodes (LED) as transmitter and photo detector (PD) as receivers. Due to the saturation and interference within the RF bandwidth, VLC becomes a potential alternative for optical wireless communication. The IEEE 802.15.7 standard for VLC [1] recommends the modulation of optical signals through run-length limited (RLL) codes at high data rates. These codes produce DC-balanced state by maintaining the average illumination intensity constant and mitigate the flickering of the channel. The flickering refers to the light blinking which is perceptible to the human eye, and is caused by a sequence of consecutive zeros or ones, also called run-length. RLL codes also have various other applications [2], [3]. Due to their low error correction capability, RLL codes are often serially concatenated with a forward error correction (FEC) code.

The design of good RLL codes remains a challenge for VLC channels. Several related works have been recorded in the literature. In [4], a FEC-aware design of RLL codes in VLC was proposed. The generator matrix structure together with channel selection methods of a polar code are used to pre-determine frozen indexes. However, this method does not extend well to high-rate codes, due to the lack of free frozen bits positions.

In [5], the eMiller code was introduced as an enhanced version of the Miller code for run-length control in VLC. It was reported that the eMiller code offers a better performance than some conventional RLL codes. However, the spectral efficiency of the eMiller code is less than that of the 4B6B code.

A 5B10B code was proposed in [6]. Although it was reported to improve the error performance compared to the Manchester and the 4B6B codes, its spectral efficiency is similar to that of the Manchester code. Moreover, the computational complexity of decoding the 5B10B code is much higher than that of the state-of-the-art alternatives.

In [7], a class of rate \((n-1)/n\) RLL codes was described. A matrix transition was designed for \((0,k)\) RLL codes for \(k = [3,7]\) with \((0,k)\) being the minimum and maximum of consecutive zeros in a codeword respectively. The performance of this code improves with increasing run-length \(k\), which could break the flicker-free property of VLC systems. Most of the work on FEC codes for VLC systems focuses on improving channel reliability while considering conventional RLL codes. In this paper, we propose a class of improved
DC-free 4B6B codes with a better decoding performance for concatenated schemes. All $16! = 2.1 \times 10^{15}$ codebooks generated from permutations of the 16 codewords from the original 4B6B code are classified based on some distance metrics in order to obtain the optimal codebooks. This proposition could significantly improve existing VLC systems based on the original 4B6B code. For instance, flicker-free FEC codes were designed based on Knuth’s balancing algorithm for VLC systems in [8]; the flickering mitigation of the encoded Knuth’s prefix was done with the Manchester code and could be improved by using the proposed 4B6B code.

The remainder of this paper is structured as follows. Section II presents some preliminaries. The proposed 4B6B code is introduced in Section III. Section IV discusses decoding performance and computational complexities. Finally, the paper is concluded in Section V.

II. BACKGROUND

VLC channels aim to provide lighting while allowing communication. The flickering mitigation and dimming control must be taken into consideration for optimal communication. The maximum flickering time period (MFTP) refers to the maximum duration that the light intensity can vary without human eye perception; and its inverse sets the lowest clock rate frequency for flickering mitigation. A system frequency above 200 Hz, equivalent to an MFTP of less than 5 ms, is imperceptible to human eyes [9].

Fig. 1 depicts the block diagram for a traditional VLC system composed by a concatenation of FEC and RLL codes. A word $u = [u_0, u_1, \ldots, u_{K-1}]$ of length $K$ is encoded by a FEC code; the obtained codeword is segmented and sent to an $aB\beta B$ RLL encoder which maps each $\alpha$ coded bits into a codeword of size $\beta$; the resulting signal is transmitted through a VLC AWGN channel as $y = [y_0, y_1, \ldots, y_{N-1}]$ via OOK modulation. Although more complex modulation schemes such as colour shift keying (CSK) are often considered, OOK modulation remains widely used for optical communications and recommended by VLC standards. In VLC systems, the transmitted signal is conveyed by light emitting diodes (LED) and received by photodetectors (PD). At the receiver, $r = [r_0, r_1, \ldots, r_{N-1}]$ of length $N$ is demodulated and soft information are extracted at the RLL decoder based on a posteriori probability (APP) decoding since RLL codes are not linear. Finally, the message $\hat{u}$ is recovered after FEC decoding.

The most used RLL codes in VLC are Manchester, 4B6B and 8B10B codes. Table I provides the conventional 4B6B code as described in [1]. This is a mapping of every 4-bit word (4B) into a 6-bit codeword (6B). Note that among the 20 possible 6-bit codewords of Hamming weight of 3, 4 codewords are excluded: 000111 and 110000 are the idle time pattern, and 110100 and 001011 are used as the preamble pattern. The posteriori bit probability based on OOK modulation through AWGN channel is given by:

$$p(r_i|y_i = \lambda) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(r_i - \lambda)^2}{2\sigma^2}},$$

where $\sigma^2$ is the noise variance and $\lambda = [0, 1]$.

Polar codes are used as FEC codes for the scope of this work but any other linear FEC code could be considered. Polar codes [10] are based on the channel polarization which consists of recursively transforming two copies of a channel into two synthetic channels such that one becomes stochastically upgraded and the other one, stochastically degraded. The matrix generator of a polar code of length $M$ can be derived from the $m$th Kronecker power, $\otimes$, of $F_2 = [1 \ 0 \ 1 \ 1] = F_2^2$. The successive cancellation (SC) decoder of polar codes has a complexity of $M \log_2 M$, and consists of traversing a binary tree of depth $m + 1$ for $M$ leave nodes, based on a left-branch-first priority.

III. PROPOSED DC-FREE RUN-LENGTH LIMITED CODES

The weight distribution of a code can be represented by a polynomial called the weight enumerating function (WEF) and given by $A(D) = \sum_{d=0}^{D} A_d D^d$ where $A_d$ is the number or multiplicity of codewords with weight (equivalent to Hamming distance from the all-zero word) $d$ [11]. Furthermore, the input-output weight enumerating function (IOWEF), $B(W, D)$ of the 4B6B code maps the WEF of the 4-bit word input to the 6-bit codeword output:

$$B(W, D) \triangleq \sum_{w=0}^{\alpha} \sum_{d=0}^{\beta} B_{w,d} W^w D^d.$$

As presented in [12], due to the non-linearity of the code, an average has to be performed over $B_{w,d}$ values, where $B_{w,d}$ is the number of codewords with Hamming distance $d$ derived from words of weight $w$; $\alpha$ and $\beta$ are the lengths of the input and output code respectively. The IOWEF of the original 4B6B code is given by

$$B(W, D) = D^6(W^0) + D^4(2.5W^2 + 2.25W^4 + 1.75W^6 + 0.5W^8) + D^6(0.125W^2 + 0.75W^3 + 0.125W^4).$$
In order to preserve the codewords used for synchronization and the run-length property of the original 4B6B code, we cannot modify the codewords. Thus, the distances and associated multiplicities are fixed.

However, by applying permutations to the codebook, i.e. by mapping different codewords to different words, it is possible to change the IOWEF of the code. There are \( 16! = 2.09228 \times 10^{13} \) different possible permutations, each giving a different 4B6B code satisfying certain distance properties. In order to facilitate the search, we propose to define the metric \( M^d \) as follows:

\[
M^d = \sum_w (A_w \times w), \quad \text{where } A_w = \sum_{d=0}^\beta B_{w,d} D^d. \quad (4)
\]

The idea behind this metric is that it will get its lower values when close words are mapped to close codewords. First, \( M^2 \) must be minimized, then \( M^4 \), and finally \( M^6 \). In order to solve this optimization problem, a backtracking algorithm was used. An efficient implementation can be considered using Knuth’s Algorithm X [13]. Indeed, some constraints could be imposed to force the metric to obtain its lower values.

Each codeword has 7 codewords distant by a Hamming distance of 2, 7 other codewords distant by a Hamming distance of 4, and 1 codeword with a distance of 6. In addition, each word considers 4 words distant by 1, 6 words distant by 2, 4 words distant by 3, and 1 word distant by 4. Thus, we want to ensure that words at a distance of 1 map to codewords that have a distance of 2 between each other, and that the word distant to 4 maps to a codeword distant to 6. For the others locations, the constraint can be relaxed with the codewords being distant from 2 or 4. As an example, the different constraints for the 8th codeword are depicted in Fig. 2. We obtained 768 codebooks that satisfied the aforementioned constraints. All these codes have the same IOWEF described as follows:

\[
B(W, D) = D^0(W^0) + D^2(4W^1 + 3W^2) + D^4(3W^2 + 4W^3) + D^6(W^4). \quad (5)
\]

This profile corresponds to the minimum metric of \( M^2 = 10 \). It can be observed that (5) has fewer components compared to (3). A brute-force search was performed to confirm that this is indeed the minimum metric. It was also verified that this metric leads to the best decoding performance for concatenated schemes. Table 2 presents an instance of the proposed 4B6B code with the minimum metric. The proposed 4B6B code is a permutation of the original 4B6B code, therefore the spectral efficiency of 0.67 bit/s/Hz as well as the maximum run-length of \( k = 4 \) remain the same for both codes. The original and the proposed 4B6B codes can be viewed as \((d, k)\) RLL codes where \( d = 0 \) and \( k = 4 \). Since there are 20 balanced codewords of length 6 and only 16 are used for the 4B6B code, the \( M^2 \) metric could be further reduced by choosing other encoding mapping that will result in a higher run-length. This could further improve the error correction performance of the proposed 4B6B code at the cost of attenuating its flicker-free property.

The analytical performance of the BER and FER for the uncoded proposed 4B6B are highlighted in (6) and (7) respectively [11],

\[
\begin{align*}
\text{BER} &= 0.6 \times \sum_w w \times \frac{A_w}{K} \times \text{erfc}\left(\sqrt{d \times R \times E_b/N_0}\right) \quad (6) \\
\text{FER} &= 0.6 \times \sum_d d \times \frac{A_d}{K} \times \text{erfc}\left(\sqrt{d \times R \times E_b/N_0}\right) \quad (7)
\end{align*}
\]

where \( \text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt \) is the complementary error function; \( R \) is the code rate, \( E_b/N_0 \), the energy per bit to noise power spectral density ratio and \( K \) the payload length.

### IV. ANALYSIS AND DISCUSSIONS

#### A. ERROR CORRECTION PERFORMANCE

Simulation results are performed over an AWGN channel considering the OOK modulation. The MAP decoding is considered for all DC-free RLL codes. Fig. 3 shows the comparison in terms of bit-error rate (BER) and frame-error rate (FER) between the original 4B6B code and the proposed 4B6B code. The maximum likelihood (ML) decoder is considered for both codes. One can observe that the performance in terms of FER is equivalent for both schemes. However, regarding the BER, gains of at least 1.5 dB are obtained with

![Graphical representation of distance constraints.](image)

**FIGURE 2.** Graphical representation of distance constraints.

![Uncoded 4B6B codes for a code length of \( N = 6 \).](image)

**FIGURE 3.** Uncoded 4B6B codes for a code length of \( N = 6 \).

---

**TABLE 2. Proposed 4B6B code.**

| Codebook | BER | FER |
|----------|-----|-----|
| 0000     | 0.0111 | 10^{-2} |
| 0001     | 0.0111 | 10^{-2} |
| 0010     | 0.0111 | 10^{-2} |
| 0100     | 0.0111 | 10^{-2} |
| 0101     | 0.0111 | 10^{-2} |
| 1000     | 0.0111 | 10^{-2} |
| 1001     | 0.0111 | 10^{-2} |
| 1010     | 0.0111 | 10^{-2} |
| 1100     | 0.0111 | 10^{-2} |

---

The table above presents the proposed 4B6B codes with their respective BER and FER values.
the proposed 4B6B code compared to the original 4B6B code for BER values above $10^{-1}$. The gains decrease with higher BER values. However, the high BER region is the region of interest for this code, since its purpose is to be used in an FEC-coded scheme. Indeed, from the point of view of the FEC decoder, the samples are independent and therefore, more than the FER, it is the BER that matters. Also, the coding gain is realized by the FEC code and even a high BER at the FEC decoder input may result in a very low BER at its output. Hence, the gain shown in Fig. 3 foresees a significant gain in the case of a FEC-coded scheme.

In Fig. 4, the performances of various convolutional coded RLL codes are presented. A rate 1/2 convolutional code with the generator $(13, 15)_8$ is considered. The convolutional code $\text{CC}(128, 64)$ coupled with the proposed 4B6B code present gains of 0.25, 1.2 and 1.21 dB over $\text{CC}(128, 64) + 4B6B$ original, $\text{CC}(96, 48) + \text{Manchester code}$ and $\text{CC}(96, 48) + \text{eMiller code}$ at the BER of $10^{-5}$. The proposed scheme performs better than $\text{CC}(128, 64) + (0, 4)\ 2/3\ \text{RLL}$ [7] for BER greater or equal to $7 \times 10^{-6}$. However, the $\text{CC}(96, 48) + 5B10B$ code [6] outperforms the proposed code by 1 dB at the BER of $10^{-5}$.

### B. COMPUTATIONAL COMPLEXITY ANALYSIS

We now compare the computational complexity of the proposed scheme with the state-of-the-art RLL schemes. We focus only on decoding RLL codes with a soft-inputs soft-outpus SISO algorithm.

The decoding of the Manchester code only requires one subtraction. For the 4B6B code, the APP decoding of each 6-bit block under the max-log approximation is computed in three steps. First, probabilities for each potential codeword are computed; the first 8 probabilities require 8 additions where 6 data are added for each codeword using 5 adders; the 8 other probabilities are obtained by inverting the first 8 ones which takes 1 adder for each; this gives a subtotal of 48 adders. Then, the marginal probability for each bit is calculated: maximums are found considering this bit as 0 or 1, and are subtracted. The maximum of 8 data is found, requiring 7 comparators;
and 8 maximums are calculated, giving a subtotal of 56 comparisons. Finally, 4 subtractions are required to obtain the 4 logarithmic likelihood ratios (LLRs). This leads to a total of 108 operations, thus 27 operations are required per decoded bit. The same reasoning translates to the 5B10B code leading to a total of 309 operations where a single decoded bit takes 62 operations.

For the BCJR decoding, for each trellis section, the branch metric and two state metrics (forward and backward) have to be computed. Then the output is given by adding those 3 metrics for each possible value of information bit or tuple of information bits. Considering the \((0, k)\) 2/3 RLL code under max-log approximation, the decoding steps consist of the following. For the branch metric, there are 8 possible transmitted codewords; 4 of them are opposite of the other 4; this requires 3 LLRs to be combined through 2 adders or 8 adders per section. Then, for the forward metric, all possible combinations of previous forward metrics coupled with branch metrics reaching a node are build. There are 4 incoming branches to each node which takes 4 adders, then the maximum among the 4 candidates is obtained by using 3 comparators; this gives a subtotal of 7 operations to get one forward metric or 56 operations for all the 8 states. The backward metric requires 56 operations similar to the forward one, only the links between nodes are different. For the LLR outputs, since one trellis section counts 32 paths, 32 sums are required based on 3 inputs each leading to 64 adders. Then, the maximum of 16 bits is computed through 15 comparators; so, 60 comparators are needed to cover the 4 cases: first bit is 0 or 1, second bit is 0 or 1. This produces a total of 244 operations for decoding 2 bits or 122 operations per bit. Similarly, the eMiller code through BCJR decoding algorithm requires 56 operations for decoding 2 bits.

Table 3 presents the computational complexity of decoding DC-free RLL codes. This counts the number of operations to be performed for obtaining a single decoded bit through either APP or BCJR decoding algorithm. All 2-input operations such as addition, subtraction or comparison are treated as equivalent. The Manchester code has the smallest number of operations but its spectral efficiency is only 0.5 bit/s/Hz against 0.67 bit/s/Hz for the 4B6B code, also its error correction performance is lower than that of the proposed 4B6B code. The 5B10B code [6] has the higher complexity that is 2.3 times that of the proposed scheme; therefore even though the error correction capabilities of the 5B10B code is attractive, its complexity is very high and increases with its code length. Moreover, the decoding complexity of the 4B6B code is 4.51 and 1.03 times lower than that of the \((0, 4)\) 2/3 RLL code [7] and the eMiller code [5] respectively. In comparison, the number of operations per bit for the SC decoding equals \(\log_2 N \approx 8\) for \(N = 192\), making RLL decoding non-negligible. Overall, our proposed scheme presents a good trade-off between complexity and error correction performance compared to state-of-the-art schemes.

Fig. 6 shows the complexity comparison of RLL codes versus SC decoding in function of the code length.

The SC decoding presents a complexity of \(N \log_2 N\) operations, where \(N\) is the code length. The Manchester decoding has the lowest complexity. It can be observed that the complexity decoding of 4B6B and 8B10B code is much larger than that of the SC decoding. This can be explained by the fact that RLL codes are non-linear codes which can only be decoded through APP where an exhaustive search is conducted within a codebook.

**V. CONCLUSION**

A new class of improved 4B6B codes was introduced for VLC systems based on metrics related to the distance profile performance. The proposed 4B6B code is a simple permutation of the original 4B6B code. Simulation results showed that the proposed 4B6B code outperforms state-of-the-art RLL codes for a concatenated scheme in terms of error correction capabilities. In particular, for the case of polar coded RLL codes, gains of at least 0.3 dB were recorded for a target BER of \(10^{-4}\). Outstandingly, a gain of 1.4 dB was observed when compared to the original and standardized 4B6B code, while the same complexity and spectral efficiency are maintained. This improvement is related to the construction of the proposed 4B6B RLL code in the sense that the inner 4B6B RLL code was optimally decoded before to be sent to the outer FEC decoder. For future work, optimising other RLL codes like 6B8B and 8B10B codes for VLC systems should be considered; also investigating on the concatenation of the proposed RLL code with other codes used in optical networks such as staircase codes.

**REFERENCES**

[1] IEEE Standard for Local and Metropolitan Area Networks—Part 15.7: Short-Range Wireless Optical Communication Using Visible Light, IEEE Standard 802.15.7-2011, Sep. 2011, pp. 1–309.
[2] K. A. Immink, Codes for Mass Data Storage Systems, 2nd ed. The Netherlands: Shannon Foundation Publishers, 2004.
[3] E. N. Mambou and T. G. Swart, “A construction for balancing non-binary sequences based on gray code prefixes,” IEEE Trans. Inf. Theory, vol. 64, no. 8, pp. 5961–5969, Aug. 2018.
[4] H. Wang and S. Kim, “Design of polar codes for run-length limited codes in visible light communications,” IEEE Photon. Technol. Lett., vol. 31, no. 1, pp. 27–30, Jan. 1, 2019.

[5] X. Lu and J. Li, “New Miller codes for run-length control in visible light communications,” IEEE Trans. Wireless Commun., vol. 17, no. 3, pp. 1798–1810, Mar. 2018.

[6] V. A. Reguera, “New RLL code with improved error performance for visible light communication,” 2019, arXiv:1910.10079.

[7] Z. Li, H. Yu, B. Shan, D. Zou, and S. Li, “New run-length limited codes in on–off keying visible light communication systems,” IEEE Wireless Commun. Lett., vol. 9, no. 2, pp. 148–151, Feb. 2020.

[8] E. Ngomseu Mambou, T. Tonnellier, S. A. Hashemi, and W. J. Gross, “Efficient flicker-free FEC codes using Knuth’s balancing algorithm for VLC,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2019, pp. 1–6.

[9] IEEE Recommended Practices for Modulating Current in High Brightness LEDs for Mitigating Health Risks to Viewers, IEEE Standards 179-2015, IEEE, Jun. 2015, pp. 1–80, vol. 15, no. 8.

[10] E. Arikan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” IEEE Trans. Inf. Theory, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[11] S. Benedetto and E. Biglieri, Principles of Digital Transmission: With Wireless Applications. Norwell, MA, USA: Kluwer, 1999.

[12] Z. Dhannoon and R. F. H. Fischer, “Performance of run-length-limited codes in visible-light communications,” in Proc. 12th Int. ITG Conf. Syst., Commun. Coding (SCC), 2019, pp. 1–6.

[13] D. E. Knuth, “Dancing links,” 2000, arXiv:cs/0011047.

[14] NR: Multiplexing and Channel Coding (Release 15), document TS 38.212 V15.2.0 (2018-06), 3GPP, Jan. 2018. [Online]. Available: http://www.3gpp.org/DynaReport/38-series.htm

---

**Elie Ngomseu Mambou** (Member, IEEE) received the B.Eng. degree in electrical and electronic engineering and the B.Sc. degree in information technology, in 2014, the master’s degree in electrical and computer engineering from McGill University, and the M.Eng. (cum laude) and Ph.D. degrees in electrical and electronic engineering from the University of Johannesburg, South Africa, in 2016 and 2021, respectively. He is currently a Research Associate at the University of Johannesburg. His research interests include encoding/decoding, optimization, error correction, and error detection algorithms for constant weight codes, LDPC codes, turbo codes, and polar codes. In 2017, he was awarded the Chancellor’s Medal Award for most outstanding master’s degree graduate and the University of Johannesburg Golden Key International Society Chapter Award.

**Thibaud Tonnellier** received the M.Eng. degree from the Electrical Engineering School of Bordeaux (ENSEIRB-MATMECA), Bordeaux, France, in 2013, and the Ph.D. degree in electronics from the University of Bordeaux, Bordeaux, in 2017. He is currently a Postdoctoral Fellow with the ISIP Laboratory, McGill University, Montréal, QC, Canada. His research interests include channel coding, design, and implementation of digital signal processing systems, and computer architectures.

**Warren J. Gross** (Senior Member, IEEE) received the B.A.Sc. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 1996, and the M.A.Sc. and Ph.D. degrees from the University of Toronto, Toronto, ON, Canada, in 1999 and 2003, respectively. He is currently a James McGill Professor and the Chair with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada. His research interests include the design and implementation of signal processing systems and custom computer architectures. He served as the Chair for the IEEE Signal Processing Society Technical Committee on Design and Implementation of Signal Processing Systems. He served as the General Co-Chair for IEEE/ACM NANOARCH 2021, IEEE GlobalSIP 2017, and IEEE SiPS 2017, and the Technical Program Co-Chair for SiPS 2012. He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and as a Senior Area Editor. He is a Licensed Professional Engineer in the Province of Ontario.

* * *