Transactions Papers

Genetic Algorithm Aided Design of Component Codes for Irregular Variable Length Coding

R. G. Maunder and L. Hanzo

***Abstract*—In this paper we propose a novel Real-Valued Free Distance Metric (RV-FDM) for comparing the error correction capabilities of Variable Length Error Correction (VLEC) codebooks that have the same integer-valued free distance lower bounds. We demonstrate that VLEC codebooks having higher RV-FDMs tend to have EXtrinsic Information Transfer (EXIT) functions with more pronounced 'S'-shapes. Furthermore, we show that higher-accuracy EXIT chart matching can be achieved if the component EXIT functions of an irregular code exhibit more variety. This motivates the employment of our novel genetic algorithm for designing the component VLEC codes of irregular variable length coding, that have particular EXIT functions, in addition to exhibiting desirable bit entropies and decoding complexities.**

***Index Terms*—Variable length codes, joint source and channel coding, trellis codes, information rates.**

I. Introduction

As detailed in [1], the error correction capability of a Variable Length Error Correction (VLEC) codebook C may be characterised by its free distance lower bound of

$$\bar{d}_{\text{free}}(C) = \min(d_{b_{\min}}(C), d_{d_{\min}}(C) + d_{c_{\min}}(C)), \quad (1)$$

where $d_{b_{\min}}(C)$ was defined as the minimum block-distance between any pair of equal-length codewords in the VLEC codebook C, whilst $d_{d_{\min}}(C)$ and $d_{c_{\min}}(C)$ were termed as the minimum divergence- and convergence-distances between any pair of unequal-length codewords, respectively in [1].

Note that VLEC codebooks having different error correction capabilities may have the same Integer-Valued Free Distance Lower Bound (IV-FDLB) $\bar{d}_{\text{free}}(C)$, preventing their comparison. For this reason, the IV-FDLB $\bar{d}_{\text{free}}(C)$ is unsuitable for use as the objective function or fitness function of a Genetic Algorithm (GA) [2] designed to generate VLEC codebooks with particular error correcting capabilities. This motivates the employment of the Real-Valued Free Distance Metric (RV-FDM) $D(C)$ of Section II to characterise the error correction capability of a VLEC codebook C. Since this RV-FDM is defined within the real-valued domain, it facilitates the unambiguous comparison of the VLEC codebooks' error correction capabilities, even if they happen to have the same IV-FDLB. Section II also demonstrates that the RV-FDM of a VLEC codebook influences the shape of its EXtrinsic Information Transfer (EXIT) function [3].

This is of relevance to Irregular Variable Length Codes (IrVLCs) [4], since they employ various component VLEC codebooks with different EXIT functions to encode particular fractions of the input source symbol stream. These fractions or weights may be specifically selected in order to shape the composite IrVLC EXIT function, since this is obtained as the weighted average of the various component EXIT functions. When an IrVLC is serially concatenated [5], [6] with a regular channel code for the purpose of joint source and channel coding [7], EXIT chart matching [8] may be employed in order to create a narrow EXIT chart tunnel. This implies that iterative decoding convergence to an infinitesimally low Symbol Error Ratio (SER) is supported at near-capacity coding rates [9]. However, in Section V, we shall show that the accuracy of EXIT chart matching and hence the resultant near-capacity performance is commensurate with the variety exhibited by the component EXIT function shapes.

This motivates the employment of the novel GA of Section III for designing VLEC codebooks having particular EXIT functions. This design objective differs from that of the 'state-of-the-art' Heuristic Algorithms (HAs) used for designing the VLEC codebooks that were proposed in [10] and later refined in [11], [12]. More specifically, these algorithms attempt to maximise the coding rate $R(C)$ of a VLEC codebook C satisfying certain minimum integer-valued block distances, as well as divergence- and convergence-distances of $d_{b_{\min}}(C)$, $d_{d_{\min}}(C)$ and $d_{c_{\min}}(C)$ [1], respectively. However, the nature of these HAs does not facilitate the direct control or prediction of the resultant VLEC codebook's EXIT function shape. Therefore - as expected - the design of a suite of component VLEC codebooks having a wide range of EXIT function shapes for employment in IrVLCs typically requires a significant amount of 'trial-and-error' based human interaction.

In Section IV, our GA and the HA of [12] are employed for designing suites of VLEC codebooks for use in IrVLCs. The EXIT chart matching accuracy that may be obtained using these codebook suites is compared in Section V. Finally, we offer our conclusions in Section VI.

Paper approved by T. M. Duman, the Editor for Coding Theory and Applications of the IEEE Communications Society. Manuscript received May 25, 2007; revised November 21, 2007.

L. Hanzo is with the University of Southampton - ECS, Salisbury Road, Southampton SO17 1BJ, UK (e-mail: lh@ecs.soton.ac.uk).

The financial support of the EPSRC, Swindon UK and the EU under the auspices of the PHOENIX and NEWCOM projects is gratefully acknowledged.

Digital Object Identifier 10.1109/TCOMM.2009.05.070234

II. The Free Distance Metric

In this section we introduce our RV-FDM $D(C)$, which allows the comparison of the error correction capability of two VLEC codebooks having equal IV-FDLBs $\bar{d}_{\text{free}}(C)$.

As described in Section I, the IV-FDLB $\bar{d}_{\text{free}}(C)$ of a VLEC codebook C depends on the constituent pairs of 'similar' codewords, which have the minimal block-, divergence- and convergence-distances, $d_{b_{\min}}(C)$, $d_{d_{\min}}(C)$ and $d_{c_{\min}}(C)$ respectively. The remaining so-called 'dissimilar' VLEC codeword pairs do not contribute to the IV-FDLB $\bar{d}_{\text{free}}(C)$, since they have block-, divergence- and convergence-distances that are higher than $d_{b_{\min}}(C)$, $d_{d_{\min}}(C)$ and $d_{c_{\min}}(C)$, respectively. Hence, we assume that the error correction capability of a VLEC codebook may be adequately characterised by modifying the IV-FDLB $\bar{d}_{\text{free}}(C)$ to yield the RV-FDM according to

$$D(C) = \bar{d}_{\text{free}}(C) + F(C), \tag{2}$$

where $F(C) \in [0, 1)$ quantifies the average dissimilarity of the various VLEC codeword pairs, as we shall detail below.

The average dissimilarity of the various VLEC codeword pairs of a particular K-entry VLEC codebook $C = \{\mathbf{c}_k\}_{k=1}^{K}$ may be quantified by

$$F(C) = \frac{\sum_{k_1=1}^{K-1} \sum_{k_2=k_1+1}^{K} P(k_1) \cdot P(k_2) \cdot F(\mathbf{c}_{k_1}, \mathbf{c}_{k_2})}{\sum_{k_1=1}^{K-1} \sum_{k_2=k_1+1}^{K} P(k_1) \cdot P(k_2)}, \tag{3}$$

where $F(\mathbf{c}_{k_1}, \mathbf{c}_{k_2})$ quantifies the dissimilarity of the specific pair of VLEC codewords $\mathbf{c}_{k_1}$ and $\mathbf{c}_{k_2}$, as will be detailed below. Note that the calculation of $F(C)$ in (3) takes into account the probabilities of occurrence $P(k_1)$ and $P(k_2)$ of the specific VLEC codewords $\mathbf{c}_{k_1}$ and $\mathbf{c}_{k_2}$, respectively. These probabilities $P(k_1)$ and $P(k_2)$ must be considered, because the error correction capability of the VLEC codebook is related to the occurrence probabilities of those specific VLEC codewords that contribute to the IV-FDLB $\bar{d}_{\text{free}}(C)$. For example, a stronger error correcting capability can be expected, if only pairs of infrequently occurring VLEC codewords are similar. Note that in order to ensure that $F(C) \in [0, 1)$, VLEC codewords are not paired with themselves and a normalisation factor of $\sum_{k_1=1}^{K-1} \sum_{k_2=k_1+1}^{K} P(k_1) \cdot P(k_2) < 1$ is employed in the calculation of $F(C)$ in (3).

As described above, the dissimilarity of a specific pair of VLEC codewords $\mathbf{c}_{k_1}$ and $\mathbf{c}_{k_2}$ depends on the degree to which they contribute to the IV-FDLB. More specifically, this depends on the block, divergence and convergence distances between the codewords and whether it is the minimum block distance or the minimum convergence and divergence distances that dominate the IV-FDLB in (1). The dissimilarity $F(\mathbf{c}_{k_1}, \mathbf{c}_{k_2})$ may be determined according to Table I, in which $d_b(\mathbf{c}_{k_1}, \mathbf{c}_{k_2})$ is the block-distance between a pair of equal-length VLEC codewords $\mathbf{c}_{k_1}$ and $\mathbf{c}_{k_2}$, whilst $d_d(\mathbf{c}_{k_1}, \mathbf{c}_{k_2})$ and $d_c(\mathbf{c}_{k_1}, \mathbf{c}_{k_2})$ are the divergence- and convergence-distances between a specific pair of unequal-length codewords $\mathbf{c}_{k_1}$ and $\mathbf{c}_{k_2}$, respectively [1].

Let us now consider the effect that the RV-FDM $D(C)$ of a VLEC codebook C has on the corresponding EXIT function shape, which may be employed to characterise the *A Posteriori* Probability (APP) Soft-In Soft-Out (SISO) decoding performance [3]. Note from (2) that the value of a VLEC codebook's RV-FDM $D(C)$ is dominated by its IV-FDLB $\bar{d}_{\text{free}}(C)$ according to $\lfloor D(C) \rfloor = \bar{d}_{\text{free}}(C)$, since we have $F(C) \in [0, 1)$. As a result, a RV-FDM of $D(C) \geq 2$ is a sufficient condition for the corresponding EXIT function to reach the top-right hand $[E(C), E(C)]$ point of the EXIT chart, since $\bar{d}_{\text{free}}(C) \geq 2$ was shown to be a sufficient condition for this [13]. Here $E(C)$ is the entropy of the VLEC-encoded bits, which is typically assumed to be unity and is given by $E(C) = -\sum_{b=0}^{1} \frac{L_b(C)}{L(C)} \cdot \log_2 \frac{L_b(C)}{L(C)}$, where $L_b(C) = \sum_{k=1}^{K} P(k) \cdot L_b(\mathbf{c_k})$ and $L_b(\mathbf{c_k})$ is the number of bits in the VLEC codeword $\mathbf{c_k}$ that assume a value of $b \in \{0, 1\}$, while $L(C) = \sum_{k=1}^{K} P(k) \cdot L(\mathbf{c_k})$ and $L(\mathbf{c_k})$ is the total number of bits in the VLEC codeword $\mathbf{c_k}$.

TABLE I
The dissimilarity of a specific pair of VLEC codewords $\mathbf{c}_{k_1}$ and $\mathbf{c}_{k_2}$ depending on the degree to which they contribute to the IV-FDLB.

Condition	$F(\mathbf{c}_{k_1}, \mathbf{c}_{k_2})$
if $L(\mathbf{c_{k_1}}) = L(\mathbf{c_{k_2}})$ and $d_{b_{\min}}(C) \leq d_{d_{\min}}(C) + d_{c_{\min}}(C)$ and $d_b(\mathbf{c}_{k_1}, \mathbf{c}_{k_2}) = d_{b_{\min}}(C)$	0
if $L(\mathbf{c_{k_1}}) \neq L(\mathbf{c_{k_2}})$ and $d_{d_{\min}}(C) + d_{c_{\min}}(C) \leq d_{b_{\min}}(C)$ and $d_d(\mathbf{c}_{k_1}, \mathbf{c}_{k_2}) = d_{d_{\min}}(C)$ and $d_c(\mathbf{c}_{k_1}, \mathbf{c}_{k_2}) = d_{c_{\min}}(C)$	0
if $L(\mathbf{c_{k_1}}) \neq L(\mathbf{c_{k_2}})$ and $d_{d_{\min}}(C) + d_{c_{\min}}(C) \leq d_{b_{\min}}(C)$ and $\left(d_d(\mathbf{c}_{k_1}, \mathbf{c}_{k_2}) = d_{d_{\min}}(C) \text{ xor } d_c(\mathbf{c}_{k_1}, \mathbf{c}_{k_2}) = d_{c_{\min}}(C)\right)$	0.5
otherwise	1

Furthermore, just as VLEC codebooks having higher IV-FDLBs have EXIT functions exhibiting more pronounced 'S' shapes, so do VLEC codebooks having higher RV-FDMs. More specifically, the effect of the IV-FDLB may be observed by comparing the inverted EXIT functions provided in Figure 2 for the 0.5-rate VLEC codebooks C_{16} and C_{20}, which have IV-FDLBs of $\bar{d}_{\text{free}}(C_{16}) = \lfloor D(C_{16}) \rfloor = 4$ and $\bar{d}_{\text{free}}(C_{20}) = \lfloor D(C_{20}) \rfloor = 2$, respectively. Similarly, we may observe the effect of the RV-FDM in Figure 2 by comparing the inverted EXIT functions of the 0.65-rate VLEC codebooks C_{15} and C_{19}, which have the same IV-FDLB of $\bar{d}_{\text{free}}(C_{15}) = \bar{d}_{\text{free}}(C_{19}) = 2$, but different RV-FDMs of $D(C_{15}) = 2.981$ and $D(C_{19}) = 2.269$, respectively. In both comparisons, the VLEC codebook having the larger IV-FDLB or RV-FDM can be observed to have an EXIT function with two inflections and a pronounced 'S' shape, while the other can be seen to have a single inflection. This trend of obtaining EXIT functions having a more pronounced 'S' shape with stronger inflections for VLEC codebooks having higher RV-FDMs was observed throughout our experiments.

These results complement the property that the area beneath an inverted VLEC EXIT function is approximately equal to the coding rate $R(C)$ of the VLEC codebook C [9], [14]. Hence, the shape of a VLEC EXIT function depends on both the coding rate $R(C)$ and the RV-FDM $D(C)$ of the VLEC codebook C.

Note that we are typically interested in VLEC codebooks C having a RV-FDM of $D(C) \geq 2$, since they support iterative decoding convergence to the top-right hand $[E(C), E(C)]$ point of the EXIT chart [13]. While we may expect the error correction performance of a VLEC codebook to be commensurate with its RV-FDM $D(C) \geq 2$, this is of little concern, since our objective of achieving an infinitesimally low SER is fulfilled at the top-right hand $[E(C), E(C)]$ point of the EXIT chart, regardless of the RV-FDM [3]. Indeed, regardless of the particular RV-FDM value $D(C) \geq 2$ of a specific VLEC codebook, it is still useful in IrVLCs, since the accuracy of EXIT chart matching depends on the *variety* of the component EXIT functions, as we shall show in Section V.

III. Overview of the proposed Genetic Algorithm

In this section we assume a basic familiarity with GAs [2] and we introduce a novel GA that seeks an IrVLC component VLEC codebook C having a particular coding rate $R(C)$ and RV-FDM $D(C)$, which determine the shape of the EXIT function, as described in Section II. Furthermore, our GA seeks designs for the codebook C that yield a near-unity bit entropy $E(C)$ and a bit-based trellis representation [15] comprising only a low number $T(C)$ of trellis transitions, thus imposing a low decoding complexity. Note that our objective is not to design individual VLEC codebooks that are optimal, but rather to design a suite of component codebooks that facilitates near-capacity EXIT chart matching, as will be demonstrated in Section V.

This GA operates on the basis of a list of VLEC codebooks $\mathbf{C}$, which initially contains only the so-called parent VLEC codebook C_0. As the GA proceeds, the list $\mathbf{C}$ is populated with mutations of its constituent VLEC codebooks. In this way, the GA attempts to find VLEC codebooks C having coding rates $R(C)$ and RV-FDMs $D(C)$ that are either increased or decreased with respect to those of the parent VLEC codebook, tending towards the specified limits, R_{lim} and D_{lim} respectively. The benefits of a mutated VLEC codebook C are assessed using a composite quality metric $M(C)$, which considers the codebook's coding rate $R(C)$, RV-FDM $D(C)$, bit entropy $E(C)$ and bit-based trellis complexity $T(C)$. Naturally, a VLEC codebook having a coding rate $R(C)$ and a RV-FDM $D(C)$ close to their specified limits R_{lim} and D_{lim} respectively, as well as having a near-unity bit entropy $E(C)$ and a low bit-based trellis complexity $T(C)$, is regarded as having a high merit and has a high-valued quality metric $M(C)$.

The proposed heuristic quality metric of a VLEC codebook C is defined here as

$$M(C) = \begin{array}{cccc} \alpha_D\beta_D\frac{D(C)-D_{\mathrm{best}}}{D_{\mathrm{best}}} & + & \alpha_R\beta_R\frac{R(C)-R_{\mathrm{best}}}{R_{\mathrm{best}}} & + \\ \alpha_E\frac{E(C)-E_{\mathrm{best}}}{E_{\mathrm{best}}} & + & \alpha_T\frac{T_{\mathrm{best}}-T(C)}{T_{\mathrm{best}}} & \end{array}, \quad (4)$$

where α_D, α_R, α_E, α_T, β_D and β_R are described in Table II. The values of D_{best}, R_{best}, E_{best} and T_{best} are obtained by taking into consideration the RV-FDMs, coding rates, bit entropies and bit-based trellis complexities of all VLEC codebooks that have been incorporated into the list $\mathbf{C}$ so far. More specifically, D_{best} and R_{best} are the RV-FDM and coding rate that are closest to the specified limits D_{lim} and R_{lim}, respectively. Additionally, E_{best} and T_{best} are the highest bit entropy and lowest bit-based trellis complexity, respectively. These values are employed to normalise the RV-FDM, coding rate, bit entropy and bit-based trellis complexity of the VLEC codebook C and are constantly updated during the operation of the proposed GA. It is necessary to employ these constantly updated values, since it is often difficult to predict the values of the best RV-FDM, coding rate, bit entropy and bit-based trellis complexity that will be found during the operation of the proposed GA in advance.

At each stage of the proposed GA's operation, the list $\mathbf{C}$ is searched to find the worst constituent VLEC codebook C_{worst}, namely that having the lowest quality metric $C_{\mathrm{worst}} = \mathrm{argmin}_{C \in \mathbf{C}} M(C)$. Next, N number of mutations of this VLEC codebook C_{worst} are generated, as will be detailed below. A mutation C' is inserted into the list $\mathbf{C}$ if it:

- has a higher quality metric than the VLEC codebook C_{worst}, more specifically, if $M(C') > M(C_{\mathrm{worst}})$,
- has a RV-FDM $D(C')$ that does not exceed the specified limit D_{lim}, more specifically, if we have $\beta_D D(C') < \beta_D D_{\mathrm{lim}}$,
- has a coding rate $R(C')$ that does not exceed the specified limit R_{lim}, more specifically, if $\beta_R R(C') < \beta_R R_{\mathrm{lim}}$ and
- is different with respect to all VLEC codebooks in the list $\mathbf{C}$, i.e. we have $C' \notin \mathbf{C}$.

Finally, the VLEC codebook C_{worst} is removed from the list $\mathbf{C}$, unless it is the only codebook in the list, in which case the GA's operation is concluded and C_{worst} is output as the designed VLEC codebook.

The number of mutations N that are generated at each stage of the proposed GA's operation is constantly varied in an attempt to maintain a constant list length $L(\mathbf{C})$ equal to a specified target L_{tar}. This approach prevents the list $\mathbf{C}$ and hence the complexity of the proposed GA from growing without bound, whilst avoiding directly limiting the number of mutations of any VLEC codebook that may be admitted to the list $\mathbf{C}$. Following each stage of the proposed GA, the number of mutations to be generated in the next stage is calculated as $\max[1, \min(N_{\mathrm{next}}, N_{\mathrm{max}})]$, where N_{max} is a parameter that limits the complexity of the next stage and

$$N_{\mathrm{next}} = N\left(1 + 0.05\left([L_{\mathrm{tar}} - L(\mathbf{C})] - [L(\mathbf{C}) - L_{\mathrm{prev}}]\right)\right). \quad (5)$$

Here, $L(\mathbf{C}) - L_{\mathrm{prev}}$ is the change in list length that resulted from the current stage of the proposed GA, whilst $L_{\mathrm{tar}} - L(\mathbf{C})$ is the list length change that is desired during the next stage. Note that the constant value of 0.05 in (5) was experimentally found to successfully maintain a wide range of desired list lengths L_{tar}.

During the mutation of a particular VLEC codebook C, the proposed GA employs four different types of operation, as follows:

- A randomly selected bit value of either 0 or 1 is *inserted* into a randomly selected position in a randomly selected codeword of the VLEC codebook C.
- The value of a randomly selected bit in a randomly selected codeword of the VLEC codebook C is *toggled* from 0 to 1 or from 1 to 0, as appropriate.
- A randomly selected bit in a randomly selected codeword of the VLEC codebook C is *removed*.
- Two randomly selected codewords in the VLEC codebook C are *swapped*, exchanging their probabilities of occurrence.

TABLE II
PARAMETERS THAT CONTROL THE OPERATION OF THE PROPOSED GA.

C_0	is the parent VLEC codebook.
$\alpha_D \geq 0$, $\alpha_R \geq 0$, $\alpha_E \geq 0$ and $\alpha_T \geq 0$	specify the relative importance of optimizing the RV-FDM, coding rate, bit entropy and bit-based trellis complexity in (4), respectively.
$\beta_D \in \{-1,+1\}$ and $\beta_R \in \{-1,+1\}$	specify in (4) whether it is desirable to design VLEC codebooks that have RV-FDMs and coding rates respectively that are higher than or lower than those of the parent VLEC codebook. A value of +1 specifies that it is desirable to increase the RV-FDM or coding rate, whilst a value of -1 indicates that a reduction is desirable.
D_{lim} and R_{lim}	specify the maximum desirable RV-FDM when $\beta_D = +1$ and the maximum desirable coding rate when $\beta_R = +1$, respectively. By contrast, they specify the minimum desirable RV-FDM when $\beta_D = -1$ and the minimum desirable coding rate when $\beta_R = -1$.
L_{tar}	specifies the target length of the list $\mathbf{C}$. Note that improved optimizations can be expected for longer target list lengths, although this is associated with a greater GA computational complexity.
N_{max}	specifies the maximum number of mutations of any VLEC codebook that may be considered at each stage of the proposed GA. Note that improved choices can be expected for higher maximum mutation limits, although this is associated with a higher GA computational complexity.

During the mutation of a particular VLEC codebook C, M number of operations are performed on the VLEC codebook C, where the type of each operation is randomly selected from the above list. The number M of GA operations performed on the VLEC codebook C is randomly selected, where the probability that M takes a particular value $m \geq 1$ is given as $P(M = m) = 2^{-m}$. In this way, invoking a high number of mutations of the VLEC codebook C is possible, but infrequent, since this is less likely to result in improved quality metrics. However, this measure allows the GA to generate diverse codebooks in the interest of exploring the entire search-space sufficiently densely.

The proposed GA is detailed in Algorithm 1 and its parameters are described in Table II.

Algorithm 1 Genetic algorithm for the design of VLEC codebooks.

Initialise the list $\mathbf{C}$ with the parent VLEC codebook C_0, in order that it has a length of $L(\mathbf{C}) = 1$.
$\hat{D}_{\mathrm{best}} \leftarrow D(C_0)$, $\hat{R}_{\mathrm{best}} \leftarrow R(C_0)$, $\hat{E}_{\mathrm{best}} \leftarrow E(C_0)$, $\hat{T}_{\mathrm{best}} \leftarrow T(C_0)$
$N \leftarrow N_{\mathrm{max}}$
loop
    $D_{\mathrm{best}} \leftarrow \hat{D}_{\mathrm{best}}$, $R_{\mathrm{best}} \leftarrow \hat{R}_{\mathrm{best}}$, $E_{\mathrm{best}} \leftarrow \hat{E}_{\mathrm{best}}$, $T_{\mathrm{best}} \leftarrow \hat{T}_{\mathrm{best}}$
    $L_{\mathrm{prev}} \leftarrow L(\mathbf{C})$
    $C_{\mathrm{worst}} \leftarrow \mathrm{argmin}_{C \in \mathbf{C}} M(C)$
    for $n \leftarrow 1$ to N **do**
        Generate a random mutation C' of C_{worst}.
        if ($M(C') > M(C_{\mathrm{worst}})$ **and** $\beta_D D(C') < \beta_D D_{\mathrm{lim}}$ **and** $\beta_R R(C') < \beta_R R_{\mathrm{lim}}$ **and** $C' \notin \mathbf{C}$) **then**
            Insert C' into the list $\mathbf{C}$, in order that its length $L(\mathbf{C})$ is incremented.
            if $\beta_D D(C') > \beta_D \hat{D}_{\mathrm{best}}$ **then** $\hat{D}_{\mathrm{best}} \leftarrow D(C')$
            if $\beta_R R(C') > \beta_R \hat{R}_{\mathrm{best}}$ **then** $\hat{R}_{\mathrm{best}} \leftarrow R(C')$
            if $E(C') > \hat{E}_{\mathrm{best}}$ **then** $\hat{E}_{\mathrm{best}} \leftarrow E(C')$
            if $T(C') < \hat{T}_{\mathrm{best}}$ **then** $\hat{T}_{\mathrm{best}} \leftarrow T(C')$
        end if
    end for
    if $L(\mathbf{C}) = 1$ **then**
        Return the single remaining VLEC codebook in the list C_{worst}.
    else
        Remove C_{worst} from the list $\mathbf{C}$, in order that its length $L(\mathbf{C})$ is decremented.
    end if
    $N \leftarrow \max(1, \min(N_{\mathrm{next}}, N_{\mathrm{max}}))$
end loop

IV. DESIGN OF IRREGULAR VARIABLE LENGTH CODING SCHEMES

In this section we demonstrate the operation of the HA of [12] and the GA of Section III to design suites of VLEC codebooks for use in IrVLCs. All VLEC codebooks were designed to have RV-FDMs and IV-FDLBs of at least two, for the sake of supporting iterative decoding convergence to an infinitesimally low SER [13]. Furthermore, all VLEC codebooks were designed for the encoding of $K = 16$-ary source symbol values, having the probabilities of occurrence that result from the Lloyd-Max quantization [16] of independent Gaussian distributed source samples [4].

As described in Section I, employing the HA of [12] to design a suite of VLEC codebooks having a wide range of EXIT function shapes typically requires a significant amount of 'trial-and-error' based human interaction. As we will demonstrate in this section, this is avoided when employing the GA of Section III. Hence, in order to ensure a fair comparison, the trial-and-error was avoided when designing a suite of VLEC codebooks for use in IrVLCs using the HA of [12]. Instead, 11 VLEC codebooks $\{C_n\}_{n=1}^{11}$ having IV-FDLBs given by $\bar{d}_{\mathrm{free}}(C_n) = n + 1 \; \forall n \in [1 \ldots 11]$ were generated. This was achieved in accordance with (1) by specifying a target minimum block-distance of $d_{b_{\mathrm{min}}}(C_n) = \bar{d}_{\mathrm{free}}(C_n)$, a target minimum divergence-distance of $d_{d_{\mathrm{min}}}(C_n) = \lceil \bar{d}_{\mathrm{free}}(C_n)/2 \rceil$ and a target minimum convergence-distance of $d_{c_{\mathrm{min}}}(C_n) = \lfloor \bar{d}_{\mathrm{free}}(C_n)/2 \rfloor$ [12]. This approach is not labour-intensive, but results in a suite of VLEC codebooks having only a *limited* variety of EXIT function shapes, as will be shown in Section V.

Similarly, the GA of Section III was employed to generate a suite of 11 VLEC codebooks $\{C_n\}_{n=12}^{22}$ comprising two categories, namely $\{C_n\}_{n=12}^{18}$ and $\{C_n\}_{n=19}^{22}$. More specifically, during the generation of the first seven VLEC codebooks $\{C_n\}_{n=12}^{18}$, which have coding rates $R(C)$ in the range of $[0.25, 0.95]$, the GA attempted to maximise the RV-FDM $D(C)$ in order to provide a variety of EXIT functions having pronounced 'S'-shapes. By contrast, RV-FDMs $D(C)$ having a value as close to but no less than two were sought in order to provide a variety of EXIT functions without pronounced 'S'-shapes during the generation of the remaining four VLEC

TABLE III
PARAMETERS EMPLOYED BY THE PROPOSED GA DURING THE GENERATION OF THE 11 VLEC CODEBOOKS $\{C_n\}_{n=12}^{22}$. THE VARIABLES ARE DEFINED IN TABLE II.

C	C_0	β_D	β_R	$D_{\lim}$	$R_{\lim}$
C_{12}	C_1	$+1$	-1	∞	0.95
C_{13}	C_1	$+1$	-1	∞	0.85
C_{14}	C_1	$+1$	-1	∞	0.75
C_{15}	C_1	$+1$	-1	∞	0.65
C_{16}	C_3	$+1$	-1	∞	0.50
C_{17}	C_7	$+1$	-1	∞	0.35
C_{18}	C_{11}	$+1$	-1	∞	0.25
C_{19}	C_1	-1	-1	2	0.65
C_{20}	C_1	-1	-1	2	0.50
C_{21}	C_1	-1	-1	2	0.35
C_{22}	C_1	-1	-1	2	0.25

TABLE IV
COMPOSITION OF THE 11 VLEC CODEBOOKS $\{C_n\}_{n=12}^{22}$ DESIGNED USING THE GA OF SECTION III. FOR EACH VLEC CODEBOOK C, THE $K = 16$ CODEWORD LENGTHS $\{L(\mathbf{c_k})\}_{k=1}^{K}$ ARE PROVIDED TOGETHER WITH THE HEXADECIMAL REPRESENTATION OF THE ORDERED CONCATENATION OF THE $K = 16$ VLEC CODEWORDS $\{\mathbf{c_k}\}_{k=1}^{K}$.

C_{12}	6, 6, 5, 5, 3, 4, 3, 3, 3, 4, 4, 4, 5, 5, 6, 6, 317E29FEA2490F1EA1
C_{13}	6, 7, 5, 5, 4, 4, 3, 3, 5, 4, 4, 5, 5, 6, 7, 8, 140DE2BCB304A54F6D818
C_{14}	7, 8, 7, 6, 6, 4, 5, 4, 5, 4, 5, 3, 6, 5, 8, 7, 1761E3F330FB2A494F8AD1C
C_{15}	7, 7, 8, 7, 7, 5, 6, 4, 6, 4, 6, 5, 6, 7, 8, 7, 3540C30FF578C04E753C1862D
C_{16}	13, 11, 9, 11, 7, 8, 6, 5, 5, 6, 8, 7, 9, 11, 12, 13, 13192CCD6E0F556B3F80CA55438997333A69
C_{17}	15, 14, 16, 14, 12, 10, 9, 8, 8, 9, 10, 11, 13, 13, 15, 16, 1522D8CA63D49CD714DB2F0FF8007A4D358B39A6353669 C2B
C_{18}	25, 18, 17, 16, 15, 14, 13, 12, 11, 13, 14, 15, 28, 17, 18, 19, 1670497B3BC1D9DAEB73994EB37CF183F00038E2D9276B 2CE0FB5B6736BDA6AE3E8
C_{19}	14, 11, 11, 7, 9, 5, 3, 3, 3, 3, 5, 7, 5, 9, 16, 12, 1558553AA17EA7350C850E542AB42AA
C_{20}	11, 17, 8, 8, 9, 5, 6, 2, 2, 3, 12, 18, 5, 15, 11, 14, 0DA1B6DCD0DCD843181B61B6D8736D86DCDB7
C_{21}	13, 19, 7, 17, 10, 10, 7, 3, 3, 3, 13, 37, 6, 16, 16, 24, 36D1B6D9346DB7368D932C536C9B6D804764936D936DA3 6DB01
C_{22}	8, 14, 8, 9, 12, 6, 5, 2, 3, 2, 56, 62, 5, 15, 11, 11, 06E6DB9A1B0DB0C22236DB34343BEB6C36DB34343BEB6D B0E6DB0DB9B4

codebooks $\{C_n\}_{n=19}^{22}$, which have coding rates $R(C)$ in the range of $[0.25, 0.65]$. As described in Section II, all 11 VLEC codebooks $\{C_n\}_{n=12}^{22}$ are useful in IrVLCs, since we will show in Section V that the achievable EXIT chart matching accuracy is commensurate with the variety exhibited by the component EXIT function shapes. Naturally, maximal VLEC-encoded bit entropies and minimal bit-based trellis complexities were sought in all cases. The GA parameters employed to design the two categories of VLEC codebooks $\{C_n\}_{n=12}^{18}$ and $\{C_n\}_{n=19}^{22}$ are detailed in Table III.

During all iterations of the GA, a target list length of $L_{\text{tar}} = 1000$ was employed. Furthermore, in order to maintain a reasonable GA complexity, the value of $N_{\max}$ was chosen experimentally so that the total number of mutations considered was of the order of 10^8. Finally, the generation of VLEC codebooks using the GA of Section III was divided into two phases.

During the first phase of the GA's operation, each parent VLEC codebook C_0 was provided by one of the codebooks designed using the HA of [12] $\{C_n\}_{n=1}^{11}$, as shown in Table III. More specifically, for each of the seven VLEC codebooks $\{C_n\}_{n=12}^{18}$, the heuristically designed VLEC codebook from $\{C_n\}_{n=1}^{11}$ having the coding rate $R(C)$ that is closest to but not lower than the target coding rate $R_{\lim}$ was employed as the parent codebook C_0. By contrast, the high-rate VLEC codebook C_1 was employed as the parent codebook during the generation of each of the four VLEC codebooks $\{C_n\}_{n=19}^{22}$. Note that we elected to employ a parent codebook having a coding rate that is close to the specified target in order to reduce the GA's 'warm-up' duration, which adds or removes bits from the codewords as appropriate.

Note that each first phase of the GA's operation was employed to provide a VLEC codebook C having an optimized RV-FDM $D(C)$ and a coding rate $R(C)$ that is reduced from that of the parent code C_0 towards the target coding rate $R_{\lim}$. This was achieved using the parameter values of $\alpha_D = 1$ and $\alpha_R = \alpha_E = \alpha_T = 0$. Note that whilst this specific choice of the parameters appears to optimize only the RV-FDM $D(C)$, this naturally results in a reduced coding rate $R(C)$, as desired.

During the second phase of the GA's operation, the result of the first phase was employed as the parent code C_0. Furthermore, the parameter values of $\alpha_D = 6$, $\alpha_R = 2$, $\alpha_E = 3$ and $\alpha_T = 1$ were employed in order to jointly optimize the resultant RV-FDM $D(C)$, the coding rate $R(C)$, the VLEC-encoded bit entropy $E(C)$ and the bit-based trellis complexity $T(C)$. The composition of the 11 resultant VLEC codebooks is provided in Table IV.

V. SIMULATION RESULTS

In this section we compare the suitability of the two VLEC codebook suites that were introduced in Section IV for use in near-capacity IrVLC coding.

The APP SISO decoding performance of the VLEC codebooks $\{C_n\}_{n=1}^{11}$ designed using the HA of [12] and of those $\{C_n\}_{n=12}^{22}$ designed by the GA of Section III may be characterised by their inverted EXIT functions of Figures 1 and 2, respectively. During APP SISO decoding, all calculations were performed in the logarithmic domain and an eight-entry lookup table was employed for correcting the Jacobian approximation, requiring only Add, Compare and Select (ACS) computational operations [17]. Note that the number of ACS operations performed during APP SISO decoding can be measured and employed as a computational complexity metric, since this is characteristic of the complexity/area/speed trade-offs in systolic-array based chips. Hence, Figures 1 and 2 provide the average number $O(C)$ of ACS operations that was performed per source symbol during our EXIT chart simulations for each considered VLEC codebook C. Furthermore, Figures 1 and 2 provide the corresponding coding rates $R(C)$ and RV-FDMs $D(C)$. Note that the VLEC-encoded bit entropy $E(C)$ was found to be only a negligible distance from unity in the case of each VLEC codebook C considered.

As shown in Figure 1, the suite of VLEC codebooks $\{C_n\}_{n=1}^{11}$ generated using the HA of [12] provides only a

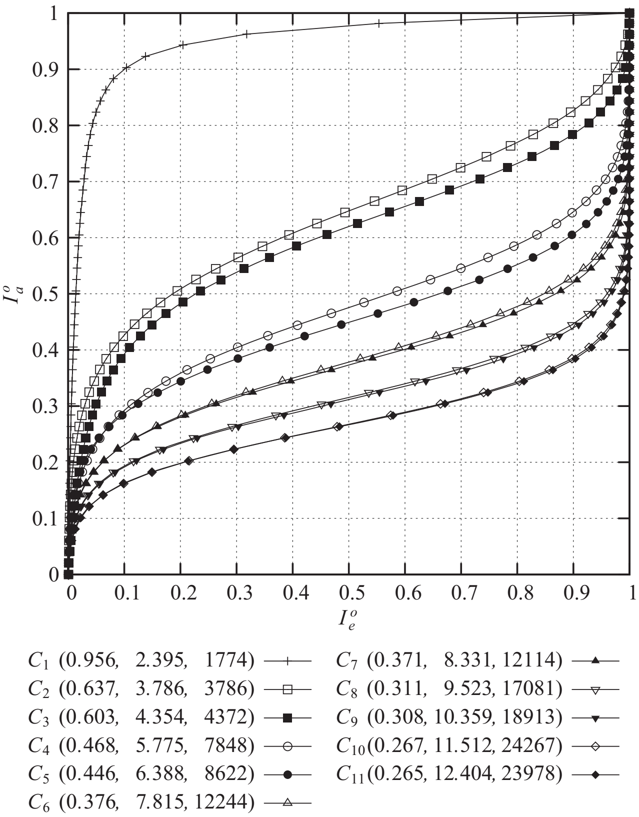


Fig. 1. Inverted EXIT functions, characterising the APP SISO decoding performance of the VLEC codebooks that were designed using the HA of [12]. Functions are labelled using the format $(R(C),\ D(C),\ O(C))$.

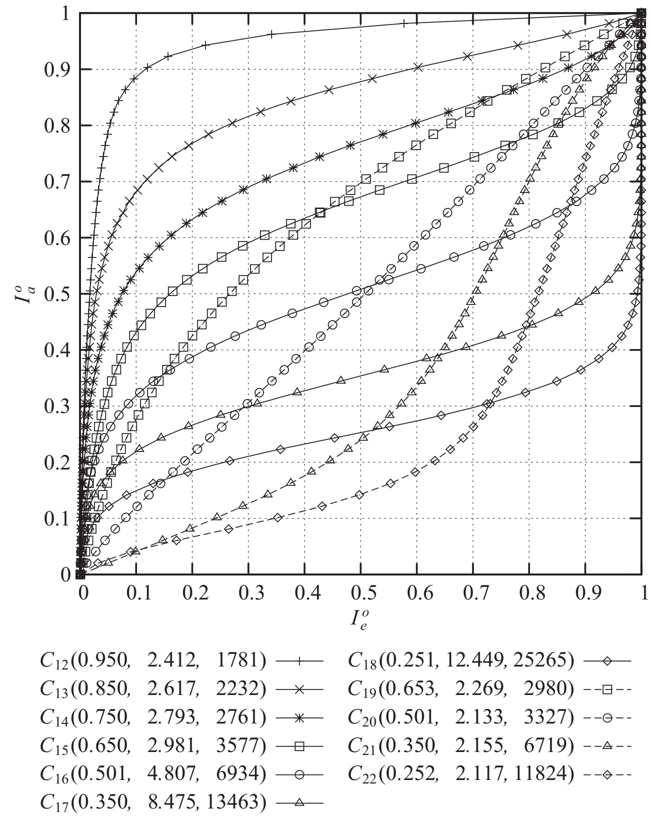


Fig. 2. Inverted EXIT functions, characterising the APP SISO decoding performance of the VLEC codebooks that were designed using the GA of Section III. Functions are labelled using the format $(R(C),\ D(C),\ O(C))$.

limited variety of 'S'-shaped EXIT functions, as predicted in Section IV. These pronounced 'S'-shapes are generated by the HA, when seeking a minimal coding rate for a particular IV-FDLB, which results in a relatively high RV-FDM for that specific coding rate, as described in Section II. By contrast, the GA of Section III provided a suite of VLEC codebooks $\{C_n\}_{n=12}^{22}$ having the wide variety of inverted EXIT function shapes shown in Figure 2, which are 'S'-shaped to varying degrees, as predicted in Section IV.

Note that for all VLEC codebooks designed to have relatively high RV-FDMs $\{C_n\}_{n=1}^{18}$, the RV-FDM $D(C)$ can be seen to increase, as the coding rate $R(C)$ decreases. By contrast, the RV-FDM $D(C)$ can be seen to decrease, as the coding rate $R(C)$ decreases in the case of the VLEC codebooks $\{C_n\}_{n=19}^{22}$ designed using the GA of Section III to have minimised RV-FDMs. These results can be explained by the increased degree of design freedom that is facilitated by having reduced coding rates $R(C)$.

Let us now comment on the APP SISO decoding computational complexity that is associated with the VLEC codebooks of Section IV, as recorded in Figures 1 and 2. Observe that for all the specific VLEC codebooks $\{C_n\}_{n=1}^{18}$ having relatively high RV-FDMs, the computational complexity of the APP SISO decoder can be seen to increase, as the coding rate is reduced. This is due to the higher codeword lengths and hence increased number of bit-based trellis transitions that are required at reduced coding rates. Whilst the computational complexities associated with the VLEC codebooks $\{C_n\}_{n=19}^{22}$ having relatively low RV-FDMs also increase as the corresponding coding rates are reduced, these are lower than those of the codebooks $\{C_n\}_{n=1}^{18}$ having relatively high RV-FDMs. Owing to the presence of these low-complexity components within the suite of VLEC codebooks $\{C_n\}_{n=12}^{22}$ generated using the GA of Section III, a corresponding IrVLC scheme can be expected to have a lower computational complexity than one designed using the components $\{C_n\}_{n=1}^{11}$ generated using the HA of [12].

Let us now consider the suitability of the VLEC codeword suites of Section IV for designing IrVLC schemes. We opted for considering IrVLC schemes that are serially concatenated and iteratively decoded in conjunction with memory-3 Unity Rate Codes (URCs) [18], having octal generator and feedback polynomials of 10_8 and 17_8, respectively. For the sake of simplicity, these URCs are employed to protect transmissions over Binary Phase Shift Keying (BPSK)-modulated narrowband uncorrelated Rayleigh fading channels. Note that the particular operating Signal to Noise Ratio (SNR) of the channel corresponds to a specific channel capacity, which represents an upper bound to the maximum IrVLC coding rate R that will permit iterative decoding convergence to an infinitesimally low SER [19].

Rayleigh fading channel SNRs in the range of -3 to 15 dB were considered and in each case appropriate IrVLC schemes were designed using either the suite of component VLEC codebooks $\{C_n\}_{n=1}^{11}$ of Section IV that were designed using the HA of [12] or the set of $\{C_n\}_{n=12}^{22}$ designed

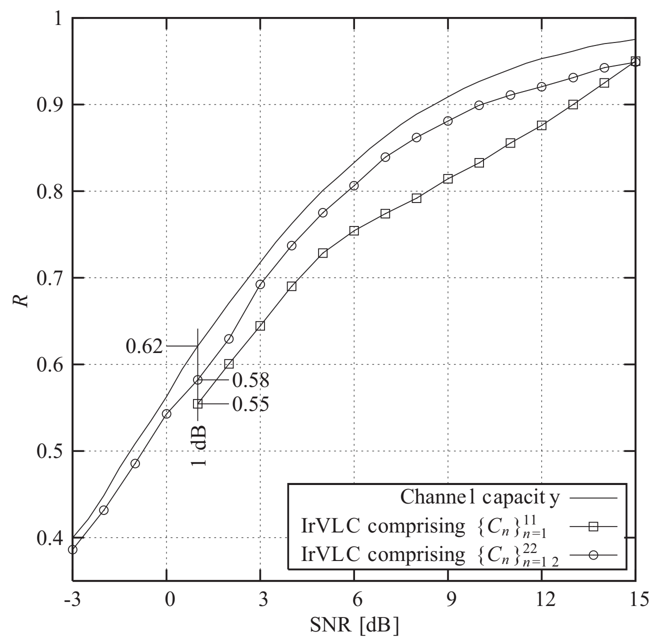


Fig. 3. The maximum coding rates R of IrVLC schemes based on the VLEC codebook suites $\{C_n\}_{n=1}^{11}$ and $\{C_n\}_{n=12}^{22}$ that yield open EXIT chart tunnels when serially concatenated with a URC that protects transmission over a BPSK-modulated Rayleigh fading channel having particular SNRs.

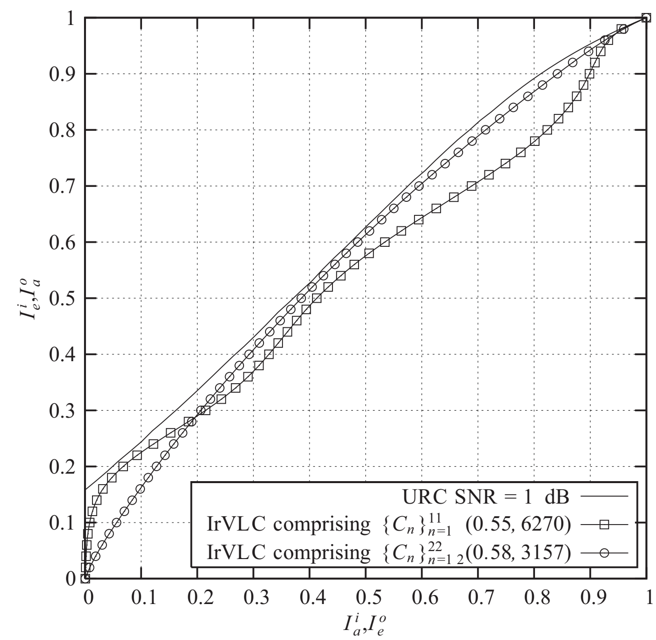


Fig. 4. Marginally open EXIT chart tunnels obtained for IrVLC schemes based on the VLEC codebook suites $\{C_n\}_{n=1}^{11}$ and $\{C_n\}_{n=12}^{22}$ when serially concatenated with a URC that protects transmission over a BPSK-modulated Rayleigh fading channel having an SNR of 1 dB.

using the GA of Section III. For each IrVLC scheme, the specific fractions of the source symbols that were encoded by each of the component VLEC codebooks were designed using the EXIT chart matching algorithm of [8]. For each case, Figure 3 plots the maximum IrVLC coding rate R for which a marginally open EXIT chart tunnel was created (provided this was possible), in addition to plotting the BPSK-modulated Rayleigh fading channel's capacity [19]. Note that the resultant open EXIT chart tunnels imply that iterative decoding convergence to an infinitesimally low SER could be achieved at these coding rates, provided that an infinite interleaver length could be employed.

As shown in Figure 3, the suite of component VLEC codebooks $\{C_n\}_{n=12}^{22}$ of Section IV that was designed using the GA of Section III is more suitable for use in IrVLCs than the set $\{C_n\}_{n=1}^{11}$ designed using the HA of [12], since it allows operation over a wider range of channel SNRs and at 'nearer-to-capacity' coding rates. This may be explained by the wide variety of component EXIT functions exhibited by the suite $\{C_n\}_{n=12}^{22}$, facilitating accurate EXIT chart matching to a wide range of inner decoder EXIT function shapes.

This is illustrated by the marginally open EXIT chart tunnels of Figure 4, which were obtained at a channel SNR of 1 dB for the IrVLCs comprising the suite of component VLEC codebooks $\{C_n\}_{n=1}^{11}$ of Section IV that was designed using the HA of [12] and the set of $\{C_n\}_{n=12}^{22}$ designed using the GA of Section III. In each case, the IrVLC coding rate R and the average number O of ACS operations that was performed per source symbol during APP SISO decoding is provided in Figures 4.

Observe in Figure 4 that the suite of component VLEC codebooks $\{C_n\}_{n=12}^{22}$ that was designed using the GA of Section III facilitates accurate EXIT chart matching and operation at the near-capacity coding rate of $R = 0.58$. Here, the shape of the corresponding inverted IrVLC EXIT function of Figure 4 was facilitated by invoking the component VLEC codebooks $\{C_n\}_{n=19}^{22}$, which have EXIT functions without a pronounced 'S'-shape, as shown in Figure 2. Since these low RV-FDM codebooks also have low decoding complexities, Figure 4 shows that the IrVLC comprising the VLEC codebooks $\{C_n\}_{n=12}^{22}$ has a decoding complexity O that is nearly 50% lower than that of the IrVLC comprising the suite of component VLEC codebooks $\{C_n\}_{n=1}^{11}$ that was designed using the HA of [12]. Note that a reduced decoding complexity was also observed at all other channel SNRs considered.

VI. Conclusions

In this paper we have proposed the novel RV-FDM of (2) for comparing the error correction capabilities of VLEC codebooks that have the same IV-FDLBs. We also demonstrated that VLEC codebooks having higher RV-FDMs tend to have EXtrinsic Information Transfer (EXIT) functions with more pronounced 'S'-shapes. This complements the property that the area below a VLEC EXIT function equals the corresponding coding rate and the property that a RV-FDM of at least two yields a VLEC EXIT function that reaches the top right hand corner of the EXIT chart. Furthermore, we argued that any VLEC codebook having a RV-FDM of at least two is useful in IrVLCs, since we showed that the accuracy of EXIT chart matching depends on the *variety* of the component EXIT function shapes. This motivated contriving our novel GA for designing IrVLC component VLEC codes having particular coding rates, RV-FDMs and hence EXIT functions, as well as desirable bit entropies and decoding complexities. Indeed, we demonstrated that our proposed approach towards IrVLC design can yield a 50% reduction in decoding complexity.

REFERENCES

[1] V. Buttigieg and P. G. Farrell, "Variable-length error-correcting codes," *IEE Proc. Commun.*, vol. 147, no. 4, pp. 211-215, Aug. 2000.

[2] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.

[3] S. ten Brink, "Convergence of iterative decoding," *Electron. Lett.*, vol. 35, no. 10, pp. 806-808, May 1999.

[4] R. G. Maunder, J. Wang, S. X. Ng, L.-L. Yang, and L. Hanzo, "On the performance and complexity of irregular variable length codes for near-capacity joint source and channel coding," to appear in *IEEE Trans. Wireless Commun.*.

[5] S. Benedetto and G. Montorsi, "Serial concatenation of block and convolutional codes," *Electron. Lett.*, vol. 32, no. 10, pp. 887-888, May 1996.

[6] ——, "Iterative decoding of serially concatenated convolutional codes," *Electron. Lett.*, vol. 32, no. 13, pp. 1186-1188, June 1996.

[7] J. Hagenauer and R. Bauer, "The turbo principle in joint source channel decoding of variable length codes," in *Proc. IEEE Inform. Theory Workshop*, Cairns, Queensland, Australia, Sept. 2001, pp. 33-35.

[8] M. Tüchler and J. Hagenauer, "EXIT charts of irregular codes," in *Proc. Conf. Information Sciences Systems*, Princeton, NJ, USA, Mar. 2002, pp. 748-753.

[9] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2657-2673, Nov. 2004.

[10] V. Buttigieg, "Variable-length error-correcting codes," Ph.D. dissertation, Department of Electronic Enginnering, University of Manchester, Manchester, U.K., 1995.

[11] C. Lamy and J. Paccaut, "Optimised constructions for variable-length error correcting codes," in *Proc IEEE Inform. Theory Workshop*, Paris, France, Mar. 2003, pp. 183-186.

[12] J. Wang, L.-L. Yang, and L. Hanzo, "Iterative construction of reversible variable-length codes and variable-length error-correcting codes," *IEEE Commun. Lett.*, vol. 8, no. 11, pp. 671-673, Nov. 2004.

[13] J. Kliewer, N. Görtz, and A. Mertins, "Iterative source-channel decoding with Markov random field source models," *IEEE Trans. Signal Processing*, vol. 54, no. 10, pp. 3688-3701, Oct. 2006.

[14] M. Adrat, J. Brauers, T. Clevorn, and P. Vary, "The EXIT-characteristic of softbit-source decoders," *IEEE Commun. Lett.*, vol. 9, no. 6, pp. 540-542, June 2005.

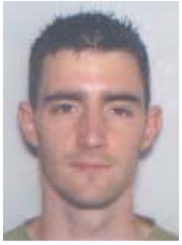
[15] V. B. Balakirsky, "Joint source-channel coding with variable length codes," in *Proc. IEEE International Symp. Inform. Theory*, Ulm, Germany, June 1997, p. 419.

[16] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, vol. 6, no. 1, pp. 7-12, Mar. 1960.

[17] L. Hanzo, T. H. Liew, and B. L. Yeap, *Turbo Coding, Turbo Equalisation and Space Time Coding for Transmission over Wireless Channels*. Chichester, UK: Wiley, 2002.

[18] D. Divsalar, S. Dolinar, and F. Pollara, "Serial concatenated trellis coded modulation with rate-1 inner code," in *Proc. IEEE Global Telecommun. Conf.*, vol. 2, San Francisco, CA, USA, Nov. 2000, pp. 777-782.

[19] J. G. Proakis, *Digital Communications*. McGraw-Hill, 1983.



Robert G. Maunder has studied with the School of Electronics and Computer Science, University of Southampton, UK, since October 2000. He was awarded a first class honours BEng in Electronic Engineering in July 2003, as well as a PhD in Communications and a lectureship in December 2007. Rob's research interests include joint source/channel coding, iterative decoding and irregular coding. He has published a number of IEEE papers in these areas.

Lajos Hanzo, Fellow of the Royal Academy of Engineering, received his first-class degree in electronics in 1976 and his doctorate in 1983. In 2004 he was awarded the Doctor of Sciences (DSc) degree by the University of Southampton, UK. During his career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the Department of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has co-authored 16 books, totalling 10 000 pages on mobile radio communications, published about 800 research papers, has acted as TPC Chair of numerous major IET and IEEE conferences, presented various keynote lectures and has been awarded a number of distinctions, most recently the WCNC'07 Best Paper Award and the 2007 ComSoc Wireless Technical Committee Achievement Award. Currently he heads an academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European IST Programme and the Mobile Virtual Centre of Excellence (VCE), UK. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. Lajos is also an IEEE Distinguished Lecturer of both the Communications as well as the Vehicular Technology Society, a Fellow of both the IEEE and the IET. He is an editorial board member of the Proceedings of the IEEE and a Governer of both the IEEE VT Society and ComSoc. He is appointed EIC of the IEEE Press. For further information on research in progress and associated publications, please refer to http://www-mobile.ecs.soton.ac.uk