Effectiveness of Hamming Single Error Correction Codes under Harsh Electromagnetic Disturbances

Jonas Van Waes*, Jonas Lannoo†, Jens Vankeirsbilck*, Andy Degraeve†, Joan Peuteman†, Dries Vanoost†,
Davy Pissoort† and Jeroen Boydens*
*Dept. of Computer Science
†Dept. of Electrical Engineering
KU Leuven, Bruges Campus
Spoorwegstraat 12, 8200 Brugge
Jonas.VanWaes@kuleuven.be

***Abstract*—Modern hi-tech systems rely heavily on communication networks while operating in increasingly harsher electromagnetic conditions. To protect transmitted data from corruption, Error Correcting Codes are widely used. In this paper, the effectiveness of the Hamming code is evaluated under harsh electromagnetic disturbances. Our simulations show that, under certain conditions, the impact of the introduced overhead cannot be compensated by the single error correcting capabilities of the Hamming codes. Moreover, for specific bit and disturbance frequencies and for larger data sets, the use of a Hamming code provides limited to no advantage.**

***Index Terms*—Hamming Code, Electromagnetic Interference, Error Correction Code, Embedded Systems, Fault Tolerance, Resilience**

I. Introduction

More and more electric, electronic and programmable electronic (E/E/PE) devices are being used in our everyday lives, ranging from smartphones and laptops to systems fulfilling mission- or safety-critical tasks. Applications such as autonomous systems are being developed and, very often, robust communication channels are crucial. Communication channels in general are increasingly affected by Electromagnetic Interference (EMI), most notably due to three trends:

1) Internal voltages of the E/E/PE systems are constantly lowered to reduce power consumption and heat dissipation;
2) The decrease of the minimum feature sizes in order to increase the transistor density, resulting in more processing power per area and reduced heat generation;
3) Harsher ElectroMagnetic (EM) environments due to evermore powerful and/or transmitting devices.

The combination of these trends leads to a lowered intrinsic immunity to EMI, which is often perceived as voltages induced on the channel, possibly leading to bit errors. To revert these corruptions, Error Detection Codes (EDCs) and Error Correction Codes (ECCs) have been used since the fifties [5]. By adding redundant information, error detection and/or recovery is performed at receiver's side. This methodology is known as Forward Error Correction (FEC). A FEC code example is the Hamming Code [5].

In previous work [10], the effectiveness of Cyclic Redundancy Checks (CRCs) was investigated. Residing as a detection mechanism, it would generate up to 50% false positives under the right conditions. False positives occur when the error detection scheme assumes the data to be error free, while the data differs from the originally sent data. This work considers equally harsh environments, but studies error correction instead of detection. Furthermore, a transition is made from plane-wave to reverberation room conditions.

Harsh and continuous electromagnetic disturbances can cause multiple upsets in the data. These disturbances are replicated in an in-house built framework for simulating reverberation room conditions. Different Hamming codes are considered, depending on the amount of data bits: 4, 11 or 26 bits. By comparison to an unprotected baseline, whereby the data bits are sent over the channel without any ECC, the effectiveness of the different Hamming codes is investigated. The reverberation chamber simulation framework is based on the Plane Wave Integral Representation for Reverberation Chambers [6].

The organization of this paper is as follows. The Hamming code is presented in Section II. The reverberation simulation framework is described in Section III with the results given in Section IV. Section V covers the conclusions of this work and the options for future work are presented in Section VI.

II. Hamming Code

The Hamming code is a linear, Single Error Correcting (SEC) Code named after its inventor R.W. Hamming [5]. In mathematical terms, the code parameters are described by Equation (1), where r is the number of redundancy (also called parity) bits, k the message (data bits) length and n the block (code word) length. For further notation, these codes are abbreviated as H$[n, k, r]$.

$$\begin{cases} r \geqslant 2 \\ k = 2^r - r - 1 \\ n = 2^r - 1 = k + r \end{cases} \tag{1}$$

In this work, three message length cases are considered: 4, 11 and 26 bits of data, resulting in H[7,4,3], H[15,11,4] and H[31,26,5], respectively. These are the smallest Hamming codes for $r \geq 3$. For $r = 2$ and $k = 1$, simple triplication is obtained. As running example for the code generation, H[7,4,3] is used, but each code is constructed in the same way. The codes have the layout as displayed in Equation (2). Dx are the data bits and Px are the parity bits.

$$[D3, D2, D1, P2, D0, P1, P0] \tag{2}$$

The Hamming code is constructed using the following three steps.

1) Each position in the code which is a power of two, is the position of a parity bit. Therefore, the parity bits are only dependent on the data bits. The data bits themselves fill the other positions.
2) Initialize the parity bits to zero.
3) The parity bits themselves are calculated by a specific algorithm which starts from the back (at $P0$). For the Pi parity bit: skip 2^i bits, check 2^i bits, skip 2^i bits, check 2^i and so on, until the end of the code word is reached. With this algorithm, the generated parity will automatically be even. Note: for the first skip operation, count one position outside of the coding array. In other words, for the three parity bits of H$[7, 4, 3]$, Equation (3) can be used. Px denotes a parity bit and Dx are data bits.

$$\begin{cases} P0 = P0 \oplus D0 \oplus D1 \oplus D3 \\ P1 = P1 \oplus D0 \oplus D2 \oplus D3 \\ P2 = P2 \oplus D1 \oplus D2 \oplus D3 \end{cases} \tag{3}$$

Since Hamming codes are linear, matrices can be used to encode and decode Hamming codes. The generator matrix $G_{(k \times n)}$ is used for encoding. The data vector is multiplied by G to generate the code word. In the correction and decoding process, three matrices are needed: the parity check matrix $H_{(r \times n)}$, the syndrome matrix $S_{(r \times 1)}$ and the decoding matrix $R_{(k \times n)}$. Multiplying H with the received code word yields S. When a correct code word was received, the syndrome S is an all-zero matrix. If an error is present in the code word, the syndrome S indicates the bit position of the error. After correction, the code word is decoded.

Note that if two or more errors reside in the code word, a non-zero syndrome is obtained as in previous paragraph. However, only one bit-error can be corrected. A successive multiplication between the code word and the parity check matrix, will yield an $S = 0$, indicating an error free code word. At this point, the code word is regarded as error free, but there are still uncorrected errors.

For all Hamming codes it is mandatory that $H \times G^T$ yields an all-zero matrix. The following matrices (Equations (4) to (6)) are used to yield the code presented in Equation (2).

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \tag{4}$$

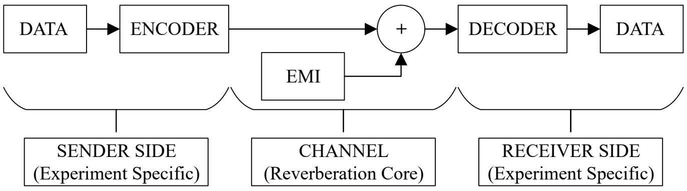


Figure 1: Framework Outline

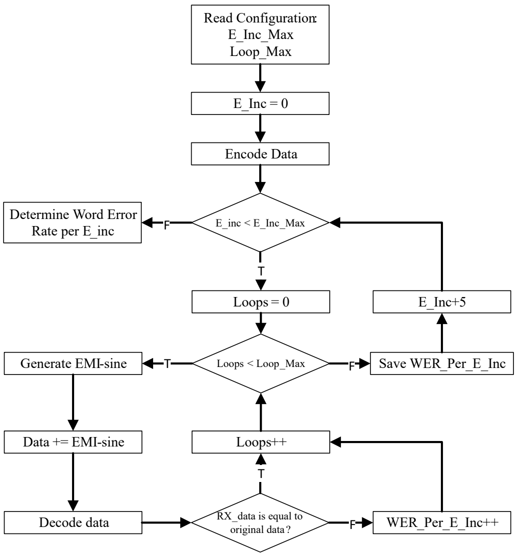


Figure 2: Framework Work Flow

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \tag{5}$$

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{6}$$

III. Reverberation Room Simulation

This section covers the in-house built reverberation room simulation framework used to test the effectiveness of Hamming code protection. Its global outline is depicted in Fig. 1.

The core of the framework is described first: the channel. Next, both sender and receiver sides are presented. These form the additions to the core. Finally, the specific parameters of the experiments are denoted.

A. Reverberation Core

The core reverberation simulation uses a numerical approach to simulate the reverberation room conditions while avoiding many full-wave simulations. The results within this simulation framework are based on the superposition of a sufficiently large set of randomly chosen plane waves to statistically represent the behavior a reverberation chamber [7]. The generic process of the reverberation chamber conditions is visualized in Fig. 2.

Three specific parameters should be known at the start: the maximum strength of the electromagnetic disturbance (E_{incMax}), the amount of repetitions to be made ($Loop_{max}$) and the considered code word (output of the encoder). The start value for E_{inc} was chosen as 0 V/m, simulating an EMI-free environment. For each of the considered strengths, $Loop_{max}$ repetitions are performed. In each repetition, an undisturbed code word is considered; there is no accumulation of EMI. After the generation of the disturbance, it is added to the encoded data. The disturbed data is then decoded and checked for errors. If an error is present, the error counter ($WER_{PerEinc}$) is incremented. In the end, this counter is divided by $Loop_{max}$ to yield the Word Error Rate (WER) per strength of E_{inc} as denoted in Equation (7). When all repetitions are completed, the next iteration is started with an increased E_{inc}. When E_{inc} equals E_{incMax}, the simulation is completed.

$$WER_{PerEinc} = \frac{Total\, Wrong\, Data\, Words}{Loop_{max}} \tag{7}$$

In order to calculate the induced voltages on the ports of a Device Under Test, many full-wave simulations are required: one simulation for each incident plane wave with a specific angle of incidence and polarization. As this would require a significant amount of simulating time, an improved algorithm was developed [9]. The improved algorithm only requires one full-wave simulation, and the induced voltages on the ports of the DUT are calculated via a reciprocity-based methodology. Since this paper considers reverberant conditions, a set of random plane waves is required to create this in accordance to the Plane Wave Integral Representation [6]. From that representation, we know that the resulting induced voltages and currents follow a Rayleigh distribution.. Since our DUT satisfies all requirements from [6], the parameter σ can be calculated, characterizing the Rayleigh distribution. From this distribution, the amplitude of the induced voltages can be calculated.

The amplitude A itself is based on an EMI strength of 1 V/m, so multiplication with the incoming strength E_{inc} must be performed. Furthermore, a random phase shift φ is added to account for the arrival times of the incident waves. The phase shift φ is uniformly distributed between $-\pi$ and π. Equation (8) represents the induced voltage $U(i)$ at the ports of the DUT.

$$\begin{aligned} U(i) = E_{inc} \times A \times \sin\left[2\pi \times F_{emi} \times T_i + \varphi\right]; \\ where\, T_i = \frac{i}{F_{bit}} + \frac{1}{2 \times F_{bit}} \\ and\, i \in [0, k + r - 1] \end{aligned} \tag{8}$$

$U(i)$ also depends on the considered disturbance frequency F_{emi} and bit frequency F_{bit}, in order to maintain the time dependencies between both frequencies. At receiver side, each bit is sampled according to F_{bit} in the middle of the bit, and not in the beginning of the bit. Therefore, the sample time T_i from Equation (8) considers half a bit period extra.

B. Coding Modules

In essence, the core in Section III-A only introduces noise (EMI) to the considered transmission line. The core was surrounded with data encoder and decoder modules. Two different encoder modules are used; one simply converting logic '0' and '1' to voltages (thus without any data protection schemes) and one with the Hamming code.

The voltages themselves (and thresholds) depend on the initial configuration of the framework. The encoding voltage V_s (at the source) is 1 V, i.e. Non-Return to Zero Level (NRZ-L) coding. Due to the assumption of matched transmission lines (the characteristic impedance of the transmission line equals both the source and load impedance), the voltage incurred at the end of the transmission line is half the original voltage. The voltage seen by the receiver is therefore $V_{rec} = V_s/2$. The core adds the reverberation EMI disturbance before the code word is decoded (see Fig 1).

Similarly, the thresholds for decoding are a function of V_s as well. Since a NRZ-L is used, the receiver assumes that anything above half the maximum theoretical receivable voltage is a '1', below it is considered '0'. The threshold lies thus at $V_{th} = V_s/4$.

All other parameters needed in this framework, are specific to the conducted experiments. These parameters are described in the following Section.

C. Experiment Specific Parameters

Six experiment specific parameters are needed: the data, the bit frequency F_{bit}, the disturbance frequency F_{emi}, the maximum number of repetitions $Loop_{max}$, the maximum considered EMI strength E_{incMax} and the difference between considered strengths ΔE_{inc}.

For the initial experiments presented here, a random 4-bit data sequence was generated. For the larger size experiments, the bit stream was expanded to the desired length (11 or 26 bits). Equation (9) provides the actual data used in the experiments. For further experiments, all data sets of the prescribed lengths are to be simulated.

$$\begin{cases} D(4) = \{1, 1, 0, 0\} \\ D(11) = \{1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1\} \\ D(26) = \{1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, \\ \qquad 1, 1, 1, 1, 1, 1, 0, 1, 0, \\ \qquad 1, 0, 0, 1, 0, 0\} \end{cases} \tag{9}$$

The values chosen in these experiments for F_{bit}, F_{emi} and E_{incMax} are based on previous work, published in [2] and [3]. The values used are depicted in Equation (10).

$$\begin{cases} F_{emi} \in [200\, MHz; 5\, GHz] \mid \Delta F_{emi} = 200\, MHz \\ F_{bit} = \{197; 200; 211; 1000\}\, MHz \\ E_{inc} \in [0; 10.000]\ V/m \mid \Delta E_{inc} = 5\ V/m \\ Loop_{max} = 100.000 \end{cases} \tag{10}$$

The values for ΔE_{inc} and $Loop_{max}$ were chosen at the beginning of the experiments.

IV. Results

In total, 300 figures were generated ($4\,F_{bit} \times 25\,F_{emi} \times 3\,DataSets$), therefore not all figures for the WER are presented in this paper; a selection was made. The results in the subsections are grouped according to the data-set length.

A. 4-bit data

The first result is shown in Fig. 3a. This is the expected outcome, where Hamming code is able to correct single bit errors and the actual WER is slightly lower. For the parameters specified in the caption of the Figure, the introduced overhead of Hamming results in a better WER. The difference in WER is the percentage of corrected single bit errors in the data. The addition of the Hamming parity bits, result in an increased ability to recover errors.

In contradiction, Fig. 3c show Hamming performing worse than sending the data without protection. This means that in a number of repetitions, the unprotected data is intact, but there are errors in the Hamming parity bits such that the correction mechanism is unable to correct. When the code word with multiple bit errors is processed by the decoder, the parity check matrix will identify one bit as false. That bit will be flipped, resulting in a correction, but an incorrect one. No double error detection is in place, thus this exposes the disadvantage of Hamming.

It should be noted that Fig. 3b includes both situations described in the previous paragraphs. For simulations with $E_{inc} < 2000\,V/m$, Hamming provides an advantage in terms of WER. This advantage however, is lost upon consideration of larger incident reverberation waves.

For the specific situation depicted in Fig. 3d, where F_{emi} is an integer multiple of F_{bit}, there is no difference in the word error rates at all. Depending on the generated wave amplitude of the Rayleigh distribution, the disruption is large enough to generate all-zero or all-one code words (which will result in wrong data) or a disruption not large enough to corrupt the data, therefore yielding correct data). Both code words are valid according to the Hamming decoding logic, but are different from the data originally sent. In essence, the same problem of false positives (as in [10] with CRC) arises here: data is seen as correct while it is not.

Figures for an F_{bit} of 1000 MHz are not provided. The results are similar to the ones provided in Fig. 3. Also note here that F_{emi} is an integer multiple of F_{bit} for specific frequencies, which yields the same graph as in Fig. 3d.

B. 11-bit data

This set of experiments shows a much more stable situation than in Section IV-A: none of the graphs shown in Fig. 4 depicts a Hamming code yielding worse WER than the unprotected data. The relative overhead of parity bits is lower than for 4 bits data. It is unlikely that the unprotected data is unaffected, while only the parity bits are corrupted. Furthermore, the presence of double (or more) bit errors (non-correctable by Hamming) are likely to have affected the original data bits as well. At this point, the disadvantage of the added parity bits being susceptible to corruption outweighs the possibility to correct a single bit error. The difference between the baseline and Hamming curves on Fig. 4a are regarded as the percentage of single bit errors in the data. The percentage below the Hamming curve, are multi-bit errors.

Fig. 4b shows again that, when F_{emi} is an integer multiple of F_{bit}, Hamming produces the same WER as unprotected data. As discussed in Section IV-A, this is due to the generation of all-zero or all-one code words. The Hamming decoder considers the code words as correct, but they differ from the sent data.

The mean ($\bar{d}$) and maximum deviation (d_{max}) between the two curves can be calculated and yield (Equation 11):

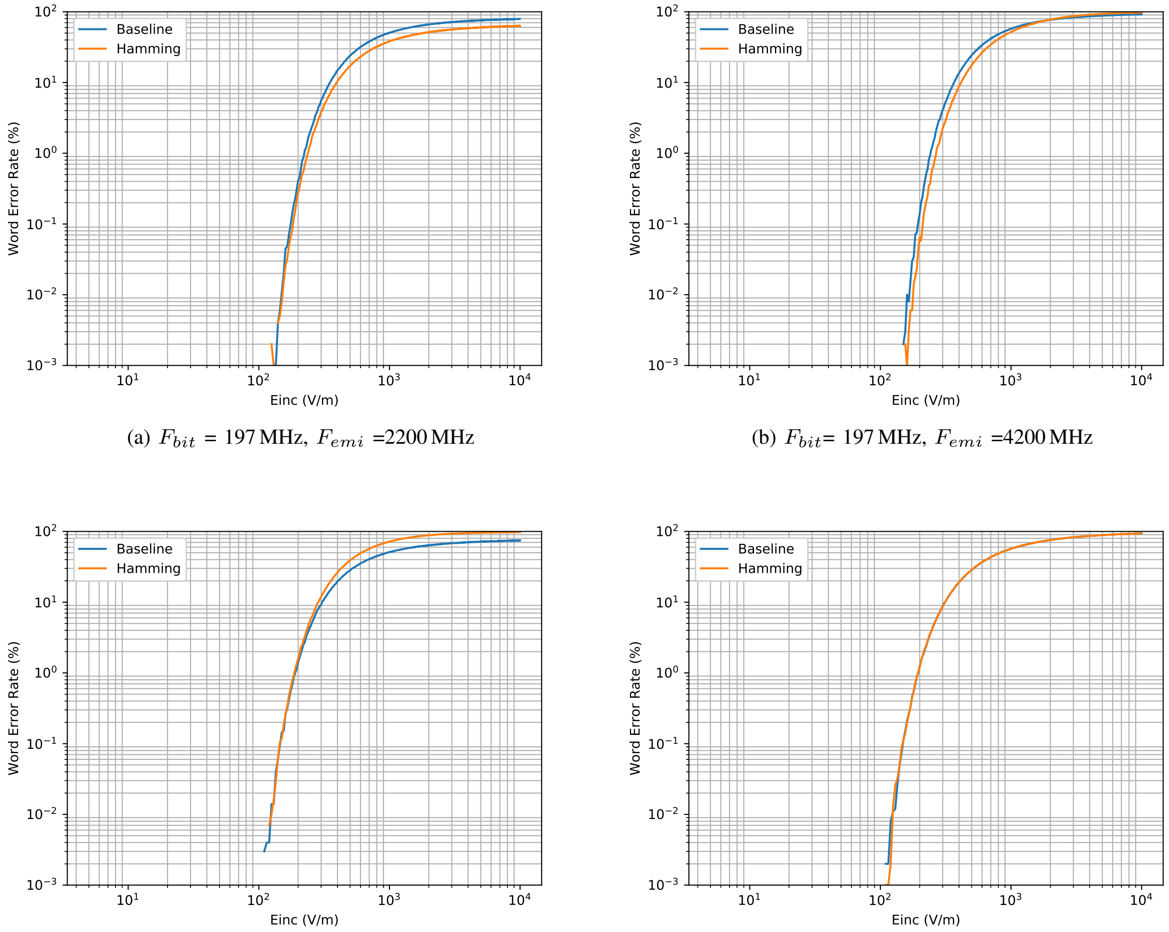
$$\begin{cases} \bar{d} = \dfrac{\sum \sqrt{(WER_{NP}(i) - WER_{H}(i))^2}}{2000} \\ d_{max} = Max(\sqrt{(WER_{NP}(i) - WER_{H}(i))^2} \end{cases} \tag{11}$$

where the i represents the value for each considered EMI strength at a specific frequency of F_{bit} and F_{emi}. There is a mean deviation of 0.1083% between the curves and a maximum of 0.554%. Given the nature of the simulation, this is identified as being equal. Note that the Hamming code is able to compensate the introduced overhead by correcting the received data blocks; the raw correct data throughput is higher.

C. 26-bit data

Now that 26 data bits are encoded with a H[31,26,5], Hamming is rapidly losing its advantages. Since a code word is 31 bits long, the probability of having only one bit-flip in the data has been reduced to almost zero for great strengths of E_{inc}. The WER curves for unprotected and Hamming encoded data are almost equal. Using the formula provided in Equation (11), the mean deviation is 0.223% and the maximum deviation is 2.12%. Note that the mean deviation is about double the percentage that was considered equal in Section IV-B. The overhead for a H[31,26,5] is about 19% (5 parity bits for each 26 bits of data) and the maximum deviation is 2.12%. Therefore, the actual correct data block rate of Hamming is lower. Performance-wise, sending the unprotected data is the best option. However, since only 18.88% of the received data blocks are correct, this is not acceptable in real-life applications unless there are other error detection schemes. The Hamming code considered here cannot distinguish a recoverable single error and an unrecoverable multi-bit error. To recover multi-bit errors, Hamming cannot be used: other types of code are required, such as Reed-Solomon Codes [4], Bose–Chaudhuri–Hocquenghem Codes [1], etc.

Extended Hamming codes are linear codes based on the original Hamming code, but do not follow the specific requirements set in Equation (1). An example is the code developed in [8], which can correct a single error, while detecting double and triple adjacent bit errors. To counter harsh EM environments, these codes might prove better and are to be considered in future experiments. It should be noted that these Extended Hamming Codes can provide multi-bit error detection, but are always Single Error Correction only.



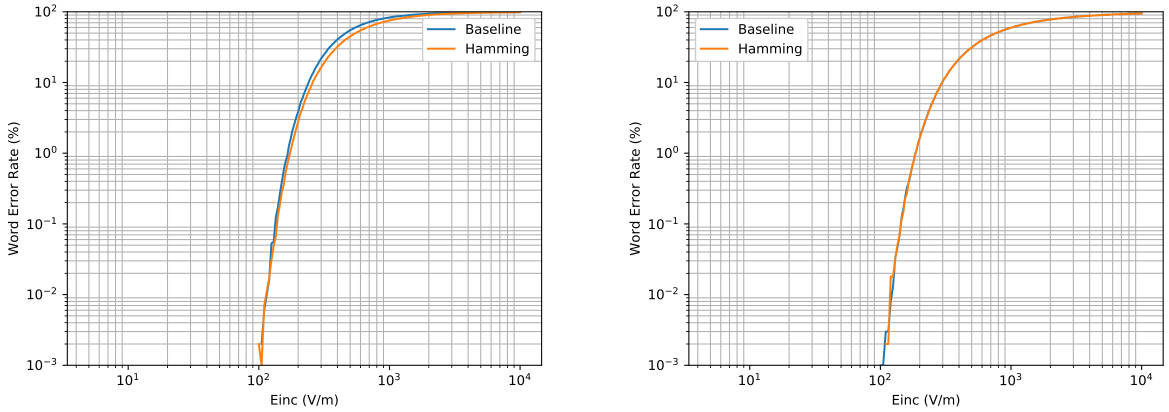
(a) F_{bit} = 197 MHz, F_{emi} =2200 MHz

(b) F_{bit}= 197 MHz, F_{emi} =4200 MHz

(c) F_{bit}= 211 MHz, F_{emi} =1000 MHz

(d) F_{bit}= 200 MHz, F_{emi} =1000 MHz

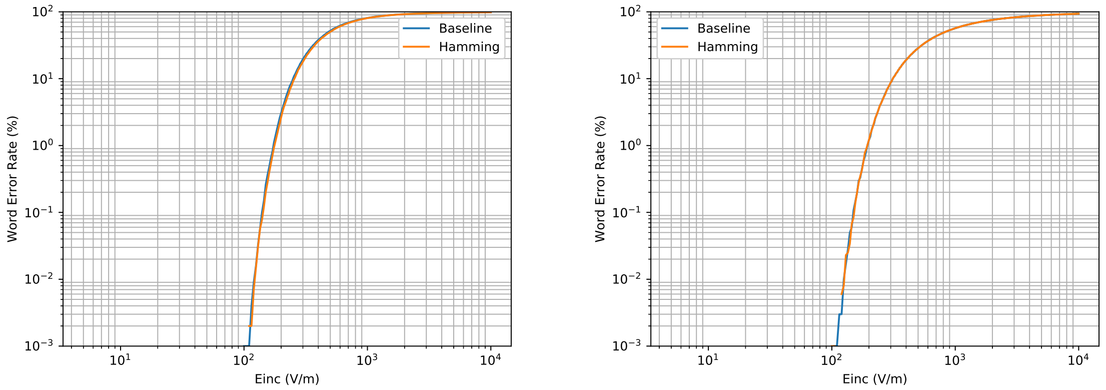
Figure 3: Results of the 4-bit experiments



(a) F_{bit}= 197 MHz, F_{emi} =3400 MHz

(b) F_{bit}= 200 MHz, F_{emi} =3400 MHz

Figure 4: Results of the 11-bit experiments



(a) F_{bit}= 211 MHz, F_{emi} =1800 MHz

(b) F_{bit}= 1000 MHz, F_{emi} =1000 MHz

Figure 5: Results of the 26-bit experiments

Furthermore, for continuous EMI, other strategies such as interleaving might prove useful. However, when the EMI persists beyond the interleaving time interval, other measure must be taken as well.

In essence, the used technique must be able to overcome the induced overhead in terms of correction before usefulness can be claimed. However, each code has its limits and needs careful consideration.

V. Conclusions

This paper presented the effectiveness of Hamming codes operating in harsh EM environments. Resulting from our in-house built framework, multiple conclusions could be drawn. In terms of WER, it is possible to achieve a slightly increased performance. The results show a difference in WER of the Hamming code compared to unprotected transmissions. The results showed that special consideration is in order when the disturbance frequency is an integer multiple of the bit frequency. In practice this can arise as harmonics of the sending frequency of another, similar device. At this point, for large disturbances, the code words are easily transformed into all-one or all-zero data words. Those code words are then regarded as correct by the receiver. The effectiveness of the implemented Hamming code significantly drops under these conditions. Moreover, those harsh disturbances introduce multiple bit errors, to which a Hamming code is not the best solution. Harsh electromagnetic environments clearly show the limits (and operational conditions) of the Hamming code. In operational systems, increased efforts are needed to protect the transmitted data from corruption. Examples thereof are Double or Triple Error Correction (DEC/TEC) Codes, possibly supplemented with interleaving of data.

VI. Future Work

This work only considered one specific data stream as noted in Equation (9). In future work, the same comparisons will need to be made to verify the claims are applicable to all data sets.

Secondly, more ranges for F_{emi} and F_{bit} will be evaluated. Building on the increased research of F_{emi} and F_{bit}, WER-curve characterization will be attempted.

As a third option, more ECCs can be simulated and verified. Migrating beyond Hamming codes, multi-bit ECCs are to be simulated. Line coding is another variable to consider. Besides NRZ-L, a multitude of encoding schemes exist, for example Manchester-encoding.

References

[1] R.C. Bose and D.K. Ray-Chaudhuri. On a class of error correcting binary group codes. *Information and Control, Elsevier*, 3(1):68–79, March 1960.

[2] A. Degraeve and D. Pissoort. Study of the effectiveness of spatially em-diverse redundant systems under plane-wave illumination. In *Proc. Asia-Pacific Int. Symp. Electromagnetic Compatibility (APEMC)*, volume 01, pages 211–213, May 2016.

[3] A. Degraeve and D. Pissoort. Study of the effectiveness of spatially em-diverse redundant systems under reverberation room conditions. In *Proc. IEEE Int. Symp. Electromagnetic Compatibility (EMC)*, pages 374–378, July 2016.

[4] P. M. Ebert and S. Y. Tong. Convolutional reed-solomon codes. *The Bell System Technical Journal*, 48(3):729–742, March 1969.

[5] R. W. Hamming. Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2):147–160, April 1950.

[6] D. A. Hill. Plane wave integral representation for fields in reverberation chambers. *IEEE Transactions on Electromagnetic Compatibility*, 40(3):209–217, August 1998.

[7] Kent Rosengren and Per-Simon Kildal. Study of distributions of modes and plane waves in reverberation chambers for the characterization of antennas in a multipath environment. *Microwave and Optical Technology Letters*, 30(6):386–391, 2001.

[8] A. Sanchez-Macian, P. Reviriego, and J. A. Maestro. Hamming sec-daed and extended Hamming sec-ded-taed codes through selective shortening and bit placement. *IEEE Transactions on Device and Materials Reliability*, 14(1):574–576, March 2014.

[9] F. Vanhee, D. Pissoort, J. Catrysse, G. A. E. Vandenbosch, and G. G. E. Gielen. Efficient reciprocity-based algorithm to predict worst case induced disturbances on multiconductor transmission lines due to incoming plane waves. *IEEE Transactions on Electromagnetic Compatibility*, 55(1):208–216, February 2013.

[10] J. Van Waes, J. Lannoo, A. Degraeve, D. Vanoost, D. Pissoort, and J. Boydens. Effectiveness of cyclic redundancy checks under harsh electromagnetic disturbances. In *Proc. Int. Symp. Electromagnetic Compatibility - EMC EUROPE*, pages 1–6, September 2017.