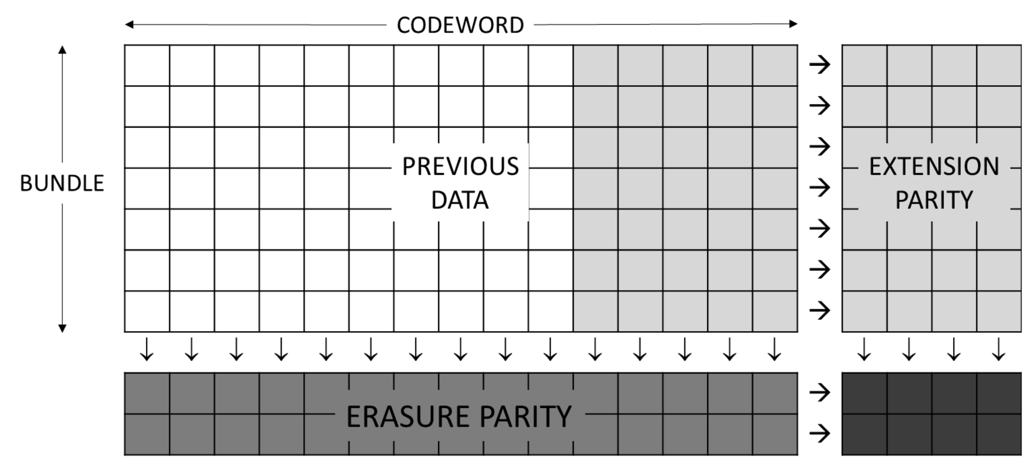
Feedback-Assisted Incremental Redundancy in mmWave Communications

Borja Peleato

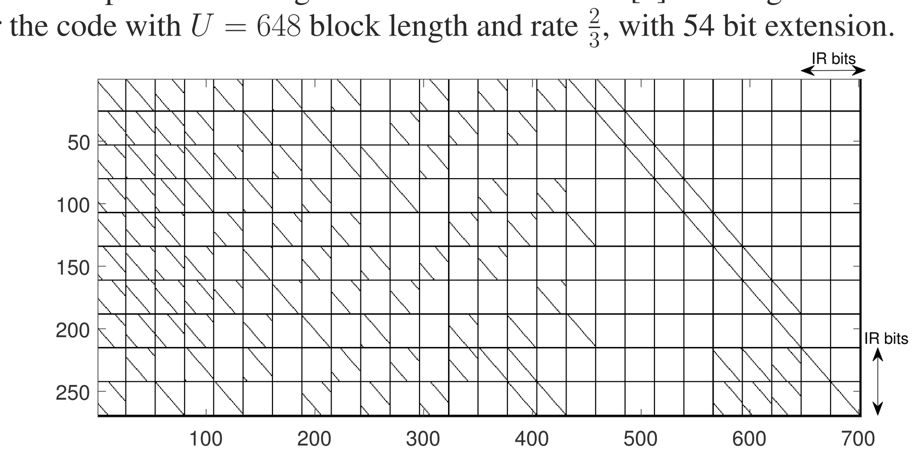
Purdue University

Background

- Acknowledgement bundling: Instead of reporting success or failure for each individual block (codeword), we use a single message to provide feedback about multiple codewords [4].
- Incremental redundancy: Can consist of extension bits for each individual codeword or the parity of an erasure code over a bundle of codewords



- QC-LDPC code extension: We will use the quasi-cyclic LDPC codes proposed for 802.11n in [2] and our own extensions optimized through numerical simulations [3]. The figure below shows the parity-check (H) for the code with $U = 648$ block length and rate $\frac{2}{3}$, with 54 bit extension.



- Min-Sum decoding for binary codes (iterative message passing on Tanner graph)

$$L_{vc,i} = L_{ch} + \sum_{j=1\ j\neq i}^{d_v} L_{cv,j} \qquad L_{cv,i} = \alpha \cdot \min_{j\neq i} |L_{vc,j}| \cdot \prod_{j=1\ j\neq i}^{d_c} \text{sign}(L_{vc,j})$$

- Modulation: The bits in a modulation symbol will be mapped to different codewords, grouped according to their probability of error. Binary LDPC decoders cannot deal with multivariate transition probabilities, so this approach provides better throughput.

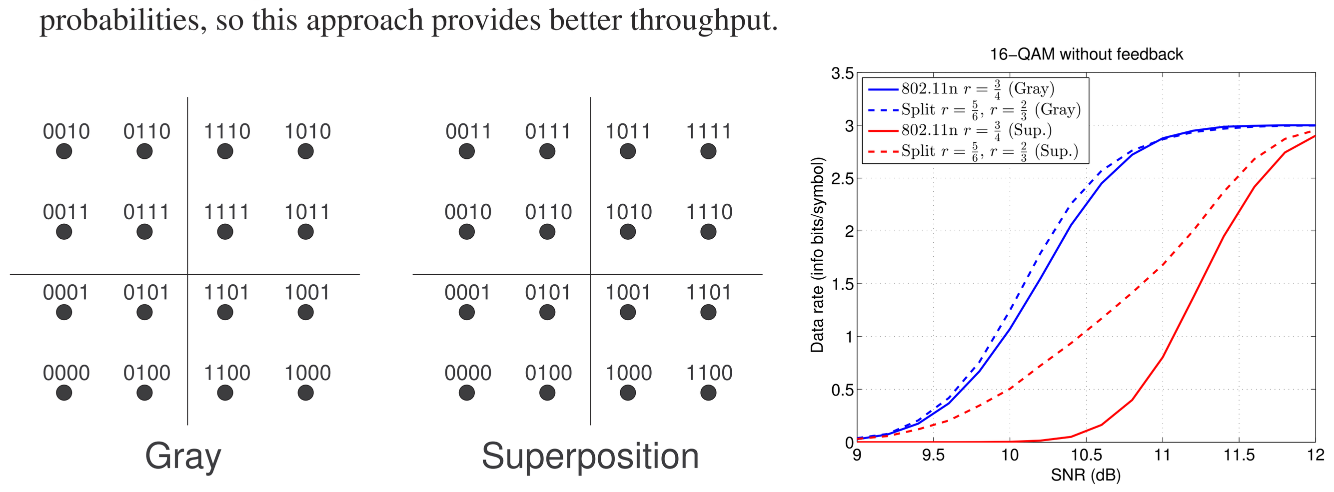


Figure 1: Left: Different mappings of bits to 16-QAM constellation symbols. Right: Average number of information bits recovered per 16-QAM symbol sent.

- Channel: It will be modeled as an AWGN channel. The noise will be iid for each bit in a codeword, but the SNR could change between codewords in a bundle.
- Performance functions:

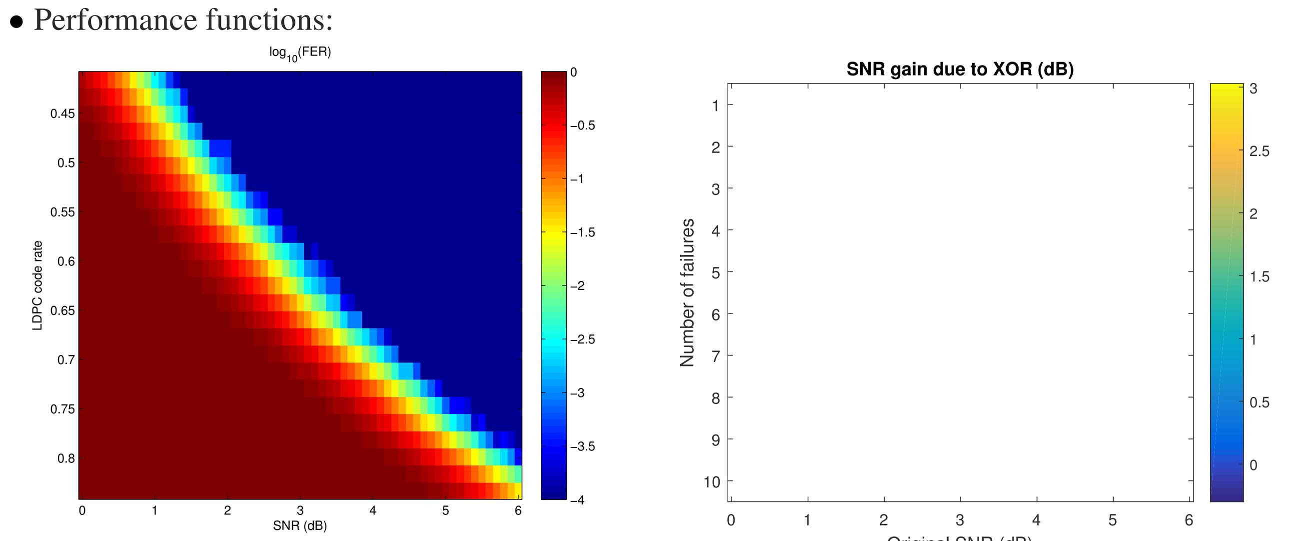


Figure 2: Left: Experimental probability of decoding failure as a function of the SNR and code rate, denoted $P_{\text{fail}}(SNR, R)$. Right: Effective SNR increase provided by the XOR, as a function of the number of failures in the XOR and their original SNR, denoted $\Delta(f, SNR)$.

Abstract

The high speeds and low latency expected in mmWave communications will require strong error correcting codes and efficient decoders. Channel conditions can change widely due to factors such as beam alignment, blockage, and interference; so it will be necessary to adjust the modulation size, coding rate and other transmission parameters frequently. It is expected that mmWave networks will use incremental redundancy (IR) schemes, where additional parity bits are transmitted whenever a codeword fails to decode successfully.

This work studies some of the trade-offs that arise in mmWave scenarios with IR, proposes a method to optimize the number and composition of the incremental bits, and provides guidelines on how to analyze the performance of different policies. It addresses the topic of acknowledgement bundling and shows that it can provide significant gains in terms of throughput.

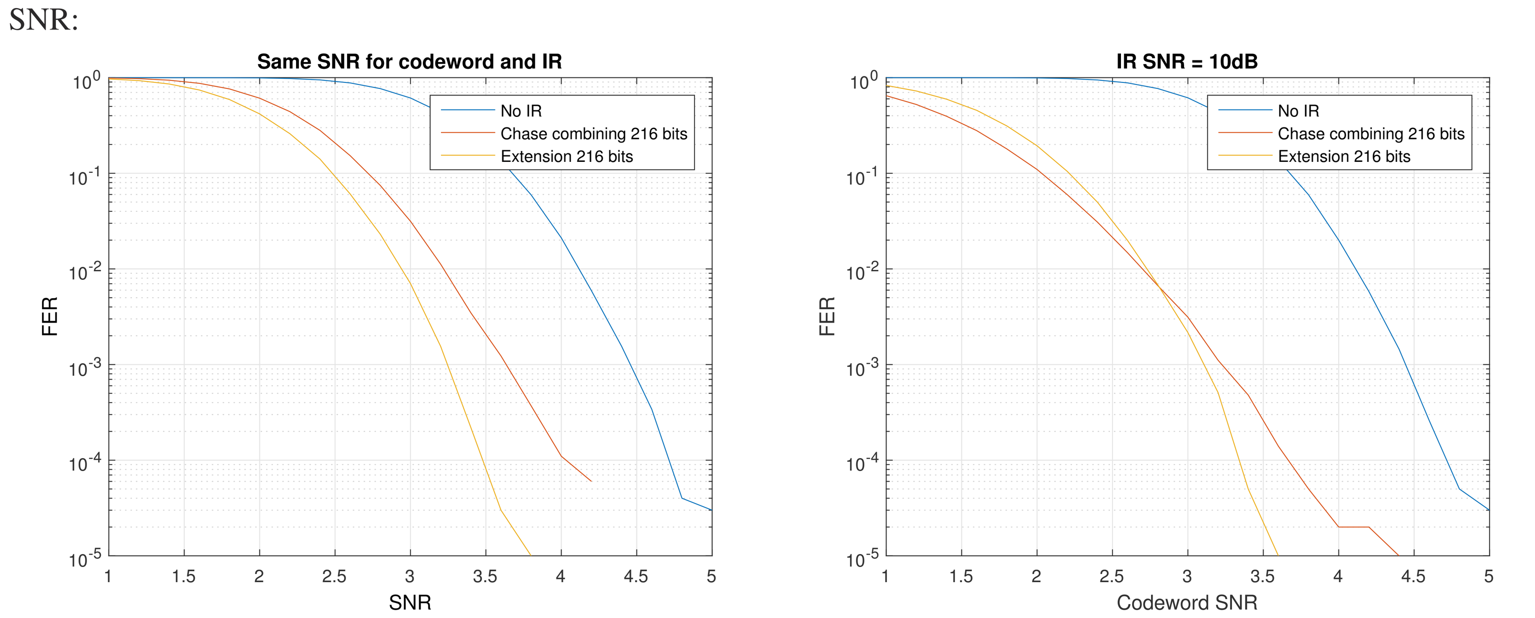
Performance analysis

Types of incremental redundancy:

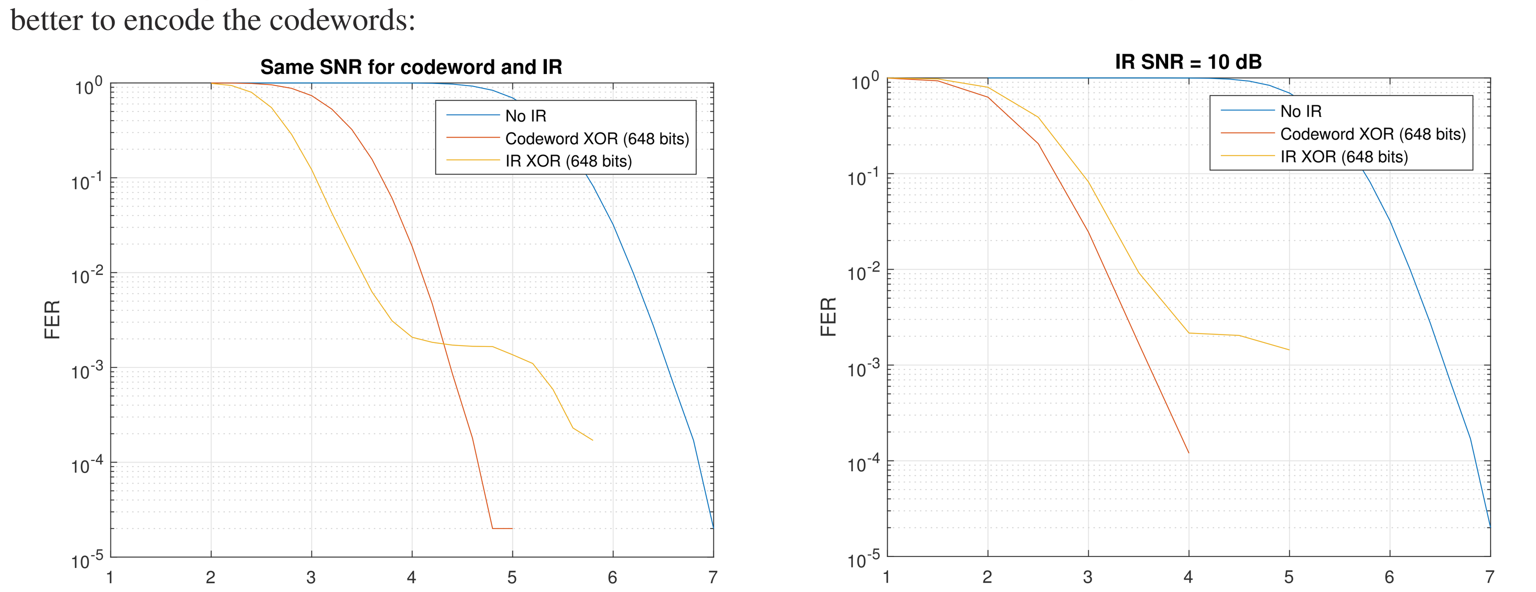
1. Chase combining: Retransmit some of the bits previously sent. The receiver can use maximal ratio combining to improve their effective SNR. It has the advantage of simplicity, since there is no need to change the decoder, but it offers suboptimal performance.
2. Extension: generate new parity bits by extending H with a new combination of bits not previously used. There are ways to avoid complicating the decoder, but it cannot take advantage of correlation between codewords.
3. Erasure codes: construct a bitwise erasure code across multiple codewords [1]. The decoding is slightly more complicated, but it couples multiple codewords together, making it possible to exploit their correlation and leverage their information.

Typically, the performance is evaluated with FER-SNR curves, implicitly assuming that both the codeword and the IR experience the same SNR. However, this is usually not the case in practice. Codewords usually fail because they experienced a lower than expected SNR, and the IR is generally transmitted with higher SNR by adjusting the modulation order, subcarrier power, etc. The analysis when the IR has higher SNR than the codeword can lead to different results.

Code extension is generally regarded as a better solution than chase combining, but this does not always hold with different SNR:



Another alternative worth considering is constructing an erasure code over the extensions and transmit its parity instead of the actual extensions. Then we can alternate the decoding of the codewords and the XOR until convergence. In the case with constant SNR this can provide better performance than constructing the erasure code over the actual codewords, albeit with a significant error floor due to miscorrection. However, when the IR has higher SNR than the codeword, it is better to encode the codewords:



References

[1] Thomas A Courtade and Richard D Wesel. Optimal allocation of redundancy between packet-level erasure coding and physical-layer channel coding in fading channels. *IEEE Transactions on Communications*, 59(8):2101–2109, 2011.

[2] IEEE. Ieee 802.11n wireless lan medium access control mac and physical layer phy specifications., 2006.

[3] Alcatel-Lucent Shanghai Bell Nokia. Ir harq schemes for ldpc. *3GPP TSG-RAN WG1 #86bis, Lisbon, Portugal*, 2016.

[4] Sangjoon Park. Multiple retransmission strategy-based HARQ for MIMO systems with acknowledgment bundling. *Electronics Letters*, 50(8):637–639, 2014.

Incremental redundancy

Figure 2 provides the probability of decoding failure for a codeword as a function of its rate R and SNR, $P_{\text{fail}}(SNR, R)$, as well as the effective SNR gain obtained from erasure coding a bundle with f failures.

Let b denote the number of codewords in a bundle and k the number of information bits per codeword. Incremental redundancy is expected to be received with SNR_{IR}

Optimization framework:

- State: $s = (f, SNR, R)$, where f represents the number of failed codewords, SNR their average signal to noise ratio, and R their coding rate. If the failed codewords have different SNR or R, we take the one with highest $P_{\text{fail}}(SNR, R)$.
- Action: a = (α,β), where α represents the number of extension bits requested (for each codeword in the bundle) and β represents the number of erasure parity bits requested (coupling all codewords in the bundle).
- Cost: $C = \sum_{i=1}^{T} b \cdot \alpha_i + \beta_i + C_{\text{drop}} I_{\{\text{fail}\}}$.

Action (α, β) will reduce the R and increase the SNR, taking the system from state $s_1 = (f_1, SNR_1, R_1)$ to $s_2 = (f_2, SNR_2, R_2)$, with $f_2 \leq f_1$, $SNR_2 = \frac{1}{n+\alpha}\left(\frac{k}{R_1} SNR_1 + \beta\Delta(f_1, SNR_1) + \alpha SNR_{IR}\right)$ and $R_2 = \frac{k}{k+\alpha R_1} R_1$. The number of failures f_2 is not deterministic, its pmf is

$$P(f_2 = k) = \binom{f_1}{f_2} p^{f_2}(1-p)^{f_1-f_2},$$

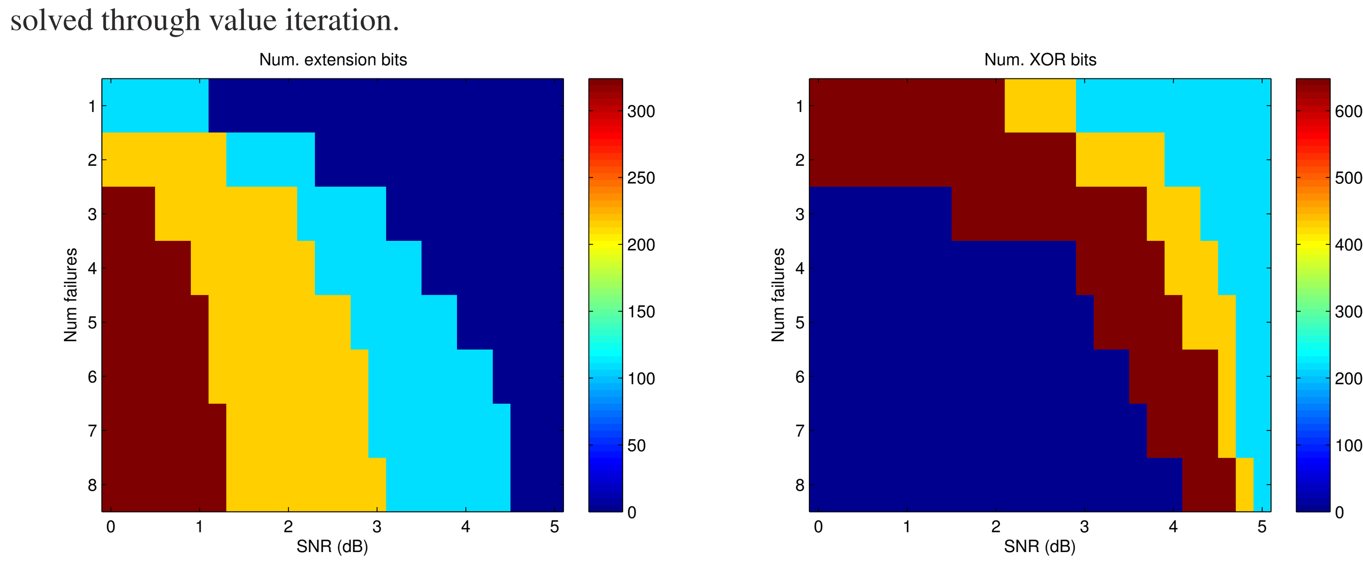
where p represents the probability of decoding failure with the IR. Specifically,

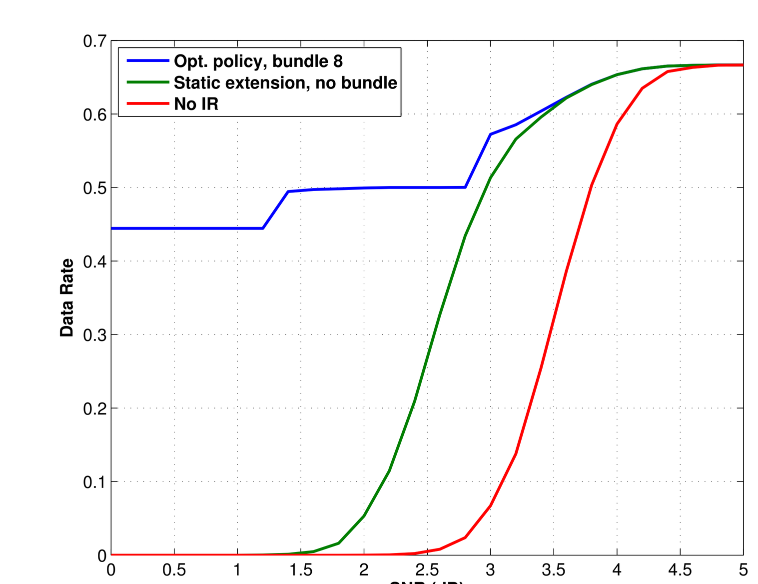
$$p = \frac{P_{\text{fail}}(SNR_2, R_2)}{P_{\text{fail}}(SNR_1, R_1)}.$$

The optimal number of bits to be transmitted in a given state s can be found as

$$(\alpha^\star, \beta^\star) = \arg\min_{\alpha,\beta} E\{Cost|s\}$$

If multiple rounds of IR are possible, the problem can be formulated as a Markov decision process, and solved through value iteration.





Conclusions

- Interleaving provides stability, but it reduces the information-theoretic capacity.
- For binary decoders, the bits within a modulation symbol should be mapped to different codewords, grouped by their marginal SNR.
- Chase combining IR can outperform extension IR when the SNR for the initial codeword is significantly lower than that for the IR bits.
- Erasure coding of the codeword extensions can be useful in the low SNR regime, but when the IR has high SNR, it is preferable to encode the actual codewords.
- Acknowledgement bundling, coupled with a dynamic programming scheme for optimizing the IR bits can significantly increase throughput, specially in the low SNR regime.

Acknowledgment

This work was supported in part by Nokia Networks