

# Redução de PAPR baseada em *Deep-Q-Learning* para sistemas de comunicação móvel

**Alunos:** Bianca S. de C. da Silva e Anderson R. Reis

**Matrículas:** 58 e 60

Setembro 2025

## 1 Introdução

O avanço dos sistemas de comunicação móvel aumentou a demanda por larguras de banda mais amplas, incentivando a utilização de espectros subutilizados, como o *TV White Space* (TVWS)[1]. No Brasil, as regulamentações permitem o uso de canais de TV ociosos nas faixas *Very High Frequency* (VHF) e *Ultra High Frequency* (UHF), desde que não haja interferência nos usuários primários[2]. Entretanto, a potência de transmissão no TVWS é limitada a 1W, o que pode restringir a cobertura e comprometer a eficiência da rede — especialmente em sistemas baseados em *Orthogonal Frequency Division Multiplexing* (OFDM), que apresentam elevados níveis de *Peak-to-Average Power Ratio* (PAPR)[3]. No próximo tópico, será apresentado em detalhes o que origina a ocorrência de PAPR.

### 1.1 PAPR no sistema OFDM

A PAPR é definido como a razão entre a potência máxima e média do sinal OFDM no domínio do tempo [4]

$$\mathcal{P} = \frac{\max(|\mathbf{x}|^2)}{\mathbb{E}[|\mathbf{x}|^2]}, \quad (1)$$

onde  $\mathbb{E}[\cdot]$  denota o valor esperado. A amplitude do sinal OFDM apresenta uma distribuição aproximadamente gaussiana, em razão da combinação de múltiplos senos e cossenos que podem se alinhar em fase ou em oposição, resultando em variações expressivas e picos de alta potência. Esses picos podem levar o amplificador de potência à saturação, provocando cortes no sinal que geram *Intercarrier Interference* (ICI) e interferência em canais adjacentes, além de aumentar a *Out-of-Band Emissions* (OOBE) [5]. Uma abordagem para mitigar o PAPR consiste na aplicação de *high back-off*, técnica em que a potência média do sinal é mantida significativamente abaixo do limite máximo de operação

linear do amplificador, de modo a evitar a saturação e, conseqüentemente, as distorções. No entanto, essa estratégia compromete a eficiência energética do amplificador.

Diversas abordagens têm sido propostas para a redução do PAPR; no entanto, a complexidade computacional ainda representa um desafio relevante. Métodos baseados em *Machine Learning* (ML), como as *Neural Networks* (NNs), têm se destacado por otimizar a configuração de pilotos com menor custo computacional. Além disso, técnicas como autoencoders, *deep unfolding* e modulação adaptativa [6, 7, 8] vêm sendo amplamente exploradas.

O principal objetivo deste trabalho é reduzir a PAPR sem a necessidade de grandes bancos de dados ou do uso direto de uma NN, utilizando, para isso, técnicas de Aprendizado por Reforço. Especificamente, busca-se otimizar a alocação das subportadoras piloto nos símbolos OFDM, de forma que, a cada nova configuração, se a PAPR resultante for inferior à anterior, o agente receba uma recompensa positiva; caso contrário, uma recompensa negativa, desencadeando a transição para um novo estado. A proposta inicial considera o símbolo OFDM formado apenas pelos  $N - N_p$  dados úteis (sem as subportadoras piloto), sendo  $N$  o número total de subportadoras, como os estados e as ações correspondendo à escolha das subportadoras piloto. Inicialmente, optou-se pelo uso do Q-Learning, por se tratar de um algoritmo simples, baseado em estados e ações discretas.

Entretanto, o Q-Learning apresenta uma limitação fundamental: sua incapacidade de lidar com espaços contínuos de estados e ações. Isso leva a soluções subótimas, já que o espaço amostral de símbolos OFDM é extremamente vasto e o método não consegue generalizar adequadamente para estados não explorados durante o treinamento.

Diante desse cenário, o *Deep-Q-Learning* (DQL) surge como alternativa promissora, pois substitui a tabela de valores por uma rede neural, possibilitando melhor generalização e maior capacidade de representação. Dessa forma, espera-se alcançar resultados mais satisfatórios, com maior capacidade de adaptação a símbolos não vistos, aproximando-se de uma solução ótima.

A Figura 1 ilustra o funcionamento do processo proposto. Nela, observa-se o vetor de bits de entrada,  $\mathbf{b}$ , que, após a modulação, gera o vetor de dados úteis,  $\mathbf{s}$ . Antes da aplicação da *Inverse Fast Fourier Transform* (IFFT),  $N_p$  subportadoras piloto são alocadas (representadas pelas linhas tracejadas). Em seguida, obtém-se o sinal OFDM no domínio do tempo, ao qual é adicionado o *Cyclic Prefix* (CP), e a PAPR é medida. Dentro do bloco DQL, se a PAPR atual for menor que a anterior, é atribuída uma recompensa positiva; caso contrário, transita-se para um novo estado e uma nova ação (escolha de piloto) é selecionada. Esse ciclo se repete até a conclusão do treinamento. No próximo tópico será detalhado o funcionamento do DQL.

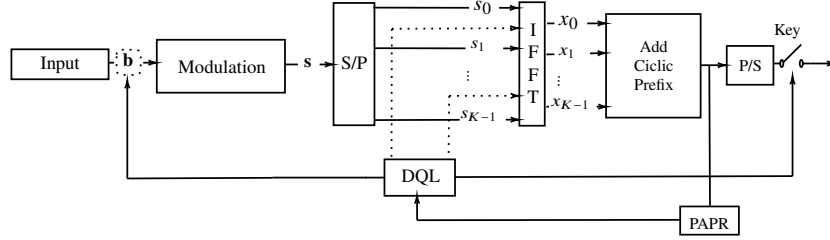


Figure 1: Transmissor OFDM com a adio do DQL e as pilotos.

## 2 Deep-Q-Learning

O DQL expande o Q-Learning tradicional ao substituir a tabela de valores  $Q(s, a)$  por uma rede neural profunda capaz de aproximar a funo de valor. Dessa forma,  possvel lidar com espaos contnuos de estados e aes, que no caso do sistema OFDM so extremamente amplos devido s diferentes combinaes possveis de subportadoras piloto.

A arquitetura do DQL segue o paradigma de aprendizado por reforo: um agente interage com o ambiente, observando estados e selecionando aes que buscam maximizar a recompensa acumulada. Nesse contexto, o ambiente  representado pelo sistema OFDM com alocao dinmica de pilotos, e a funo de recompensa  definida em termos da PAPR. Assim, aes que reduzem a PAPR em relao ao estado anterior recebem reforo positivo, enquanto aes que elevam a PAPR recebem reforo negativo.

Alm disso, o DQL incorpora tcnicas como *experience replay*, que armazena transies em uma memria e as amostra de forma aleatria, reduzindo a correlao entre os dados de treinamento, e a rede-alvo (*target network*), que estabiliza a atualizao dos parmetros da rede principal. Esses mecanismos tornam o treinamento mais robusto e evitam instabilidades comuns em problemas de RL com alta dimensionalidade.

Portanto, ao contrrio do Q-Learning, que resulta em solues discretizadas e por vezes subtimas, o DQL possui maior capacidade de generalizao e adaptao a estados no vistos, tornando-o mais adequado para problemas de alocao de pilotos em sistemas OFDM.

## 3 Resultado Esperado

Espera-se que o uso do DQL proporcione uma reduo mais significativa e estvel da PAPR em comparao ao Q-Learning tradicional, uma vez que a rede neural  capaz de explorar padres complexos entre os estados e aes. O agente deve aprender a selecionar configuraes de subportadoras piloto que, mesmo sem conhecimento explcito da estrutura do sinal, resultem em valores de PAPR mais prximos de solues timas.

Outro resultado esperado  a melhor generalizao: ao contrrio da abor-

dagem tabular, o DQL deve ser capaz de manter um bom desempenho mesmo para símbolos OFDM que não foram observados durante o treinamento. Isso se traduz em maior aplicabilidade prática do método, especialmente em cenários dinâmicos onde as condições do canal e da rede variam ao longo do tempo.

Adicionalmente, espera-se observar uma relação de compromisso (*trade-off*) entre a eficiência da redução de PAPR e a complexidade computacional. Embora o uso de uma rede neural introduza maior custo de processamento em comparação ao Q-Learning, esse custo tende a ser compensado pela eliminação da necessidade de tabelas extensas e pelo ganho em desempenho e escalabilidade.

## References

- [1] N. Rajatheva, I. Atzeni, E. Björnson, A. Bourdoux, S. Buzzi, J.-B. Doré, S. Erkucuk, M. Fuentes, K. Guan, Y. Hu, X. Huang, J. Hultkonen, J. M. Jornet, M. Katz, R. Nilsson, E. Panayirci, K. Rabie, N. Rajapaksha, M. Salehi, and W. Xu, “White Paper on Broadband Connectivity in 6G,” *6G Research Visions*, No. 10. University of Oulu, 2020.
- [2] Anatel, “Resolução nº 190,” 2021.
- [3] M. Khalil, J. Qadir, O. Onireti, M. A. Imran, and S. Younis, “Feasibility, architecture and cost considerations of using tvws for rural internet access in 5g,” in *2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN)*, pp. 23–30, 2017.
- [4] Y. A. Jawhar, L. Audah, M. A. Taher, K. N. Ramli, N. S. M. Shah, M. Musa, and M. S. Ahmed, “A Review of Partial Transmit Sequence for PAPR Reduction in the OFDM Systems,” *IEEE Access*, pp. 18021–18041, 2019.
- [5] B. Tang, K. Qin, X. Zhang, and C. Chen, “A clipping-noise compression method to reduce papr of ofdm signals,” *IEEE Communications Letters*, pp. 1389–1392, 2019.
- [6] S. Chakravarty and A. Kumar, “PAPR Reduction of GFDM Signals Using Encoder-Decoder Neural Network (Autoencoder),” *National Academy Science Letters*, pp. 213–217, 2023.
- [7] Z. Liu, X. Hu, K. Han, S. Zhang, L. Sun, L. Xu, W. Wang, and F. M. Ghannouchi, “Low-complexity PAPR reduction method for OFDM systems based on real-valued neural networks,” *IEEE Wireless Communications Letters*, pp. 1840–1844, 2020.
- [8] S. Gökceli, T. Riihonen, T. Levanen, and M. Valkama, “Machine learning based tuner for frequency-selective PAPR reduction,” *IEEE Transactions on Vehicular Technology*, pp. 5378–5383, 2022.