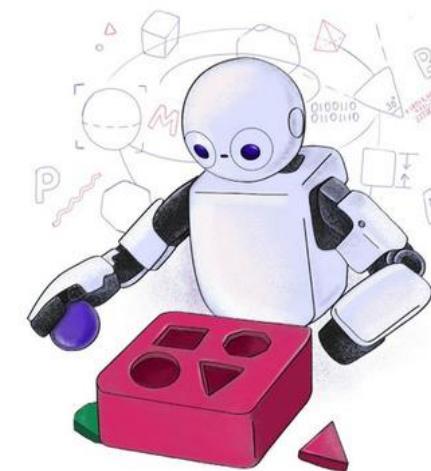


# TP558 - Tópicos avançados em Machine Learning:

## Deep Q-Network to Reduce PAPR in Communication Systems



**Inatel**

Anderson R. R. Marins  
Bianca Sabrina de C. da Silva  
[anderson@inatel.br](mailto:anderson@inatel.br)  
[bianca.sabrina@dtel.inatel.br](mailto:bianca.sabrina@dtel.inatel.br)

# Introdução

- ❑ A crescente demanda por comunicação móvel **exige cada vez mais taxa de dados, o que exige mais largura de banda.**
- ❑ O espectro subutilizado, como os **Espaços em Branco de TV (TVWS)**, é uma alternativa promissora.

# Introdução

- ❑ No Brasil, é permitido o uso de canais de TV ociosos nas faixas de VHF e UHF, **sobre algumas restrições e a principal delas é que não haja interferência com os usuários primários.**
- ❑ No entanto, outra restrição importante é que a **potência de transmissão é limitada a 1W de pico**, o que torna valores elevados de PAPR particularmente críticos, pois podem facilmente levar o transmissor à região não linear. Causando ICI, OOB, etc.

# PAPR no Sistema OFDM

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

- Os sinais OFDM têm uma distribuição de amplitude aproximadamente gaussiana (**soma de muitas subportadoras**).
- Picos altos ocorrem quando as subportadoras se **combinam em fase**.
- Esses picos podem **levar os amplificadores à saturação**, causando:
  - ✓ Distorção por recorte (clipping)
  - ✓ Interferência entre portadoras (ICI)
  - ✓ Emissões fora da banda (OOBE)

# Solução Proposta

❑ **Objetivo:** Reduzir a PAPR em sistemas OFDM usando algoritmos de **Aprendizado por Reforço**.

❑ **Abordagens:**

- ✓ Q-Learning (tabular e simples)
- ✓ Deep Q-Network (DQN) (aproximação com rede neural)

# Q-Learning

## O que é Q-Learning?

- ✓ Q-Learning é um algoritmo **value-based**.
- ✓ Ele aprende uma função  $Q(s,a)$  que estima o valor de realizar uma ação  $a$  em um estado  $s$ .

# Q-Learning

## O que é Q-Learning?

- ✓ Atualização baseada na **Equação de Bellman**:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[ r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right]$$

- ❑  $\alpha$  (learning rate): quanto rápido aprendemos.
- ❑  $\gamma$  (discount factor): equilíbrio entre recompensa imediata e futura.
- ❑  $r$ : recompensa recebida ao executar a ação.
- ❑  $s'$ : próximo estado.

# Q-Learning

## Limitações do Q-Learning?

- ❑ Funciona bem para **espaços discretos e pequenos.**
- ✓ Fica inviável quando:
  - ❑ há muitos estados;
  - ❑ há muitas ações;
  - ❑ ou o estado é contínuo/complexo;

# Deep Q-Network (DQN)

## Motivação

- Q-Learning tabular não escala para problemas maiores.
- Precisamos **aproximar a função  $Q(s,a)$  com uma rede neural.**

# Deep Q-Network (DQN)

DQN usa **mecanismos fundamentais**:

- ✓ Experience Replay
- ✓ Target Network
- ✓  $\epsilon$ -greedy exploration

# Deep Q-Network (DQN)

DQN usa **mecanismos fundamentais**:

✓ **Experience Replay**

- ❑ Armazena transições em um buffer.
- ❑ Durante o treinamento, amostra **aleatoriamente** essas experiências.
- ❑ Isso **quebra a correlação** entre amostras e estabiliza o gradiente.

# Deep Q-Network (DQN)

DQN usa **mecanismos fundamentais**:

✓ **Target Network**

❑ Mantém duas redes:

- Rede online (atualizada a cada passo)
- Rede-alvo (mais estável)

❑ A rede alvo produz o valor alvo e só é atualizada **periodicamente**, evitando instabilidade.

# Deep Q-Network (DQN)

DQN usa **mecanismos fundamentais**:

## ✓ Política $\epsilon$ -Greedy

- ❑ Controla a exploração:
  - Com probabilidade  $\epsilon$  → ação aleatória
  - Com  $1-\epsilon$  → melhor ação segundo a rede
- ❑ Começa explorando bastante e reduz gradualmente.

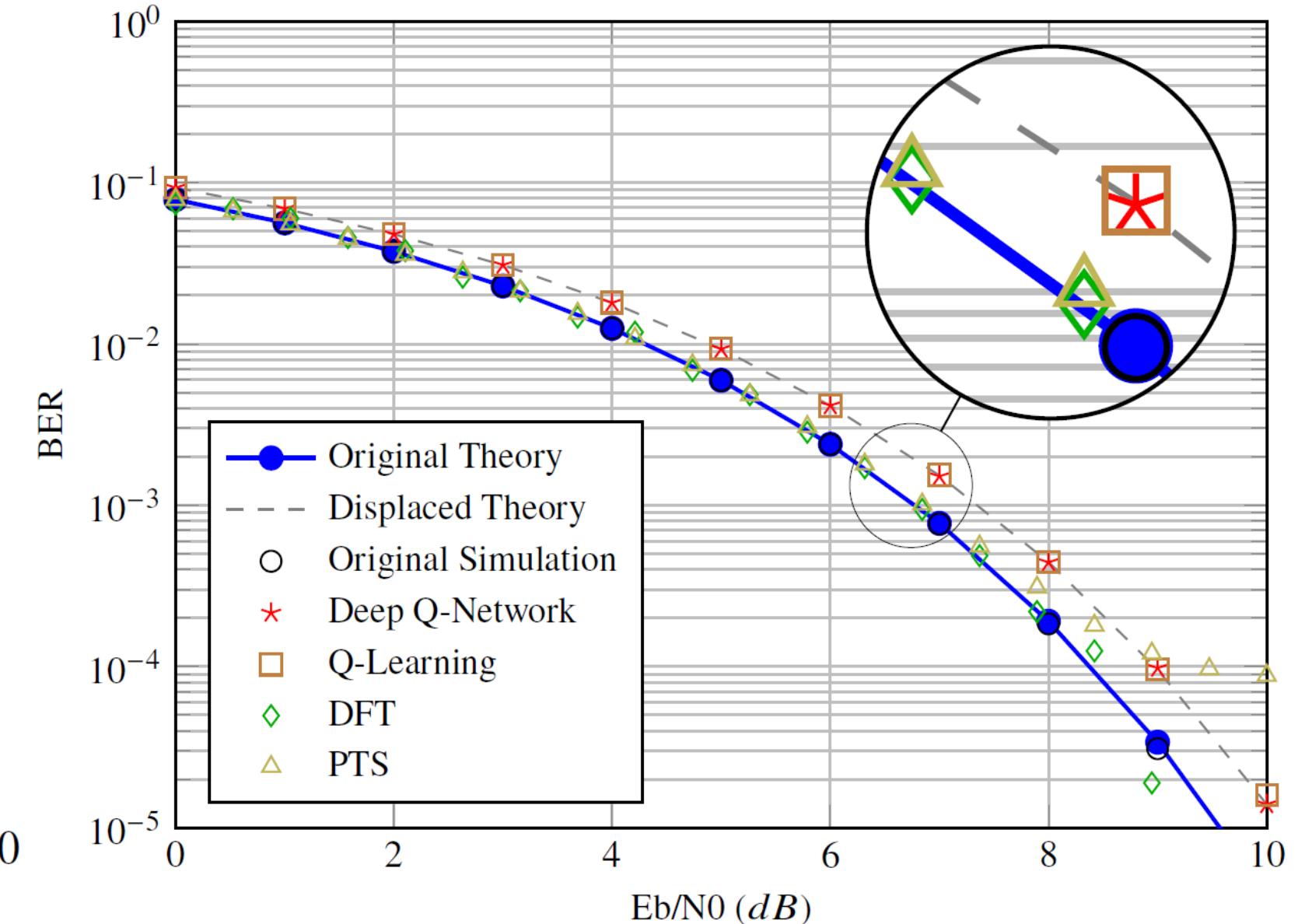
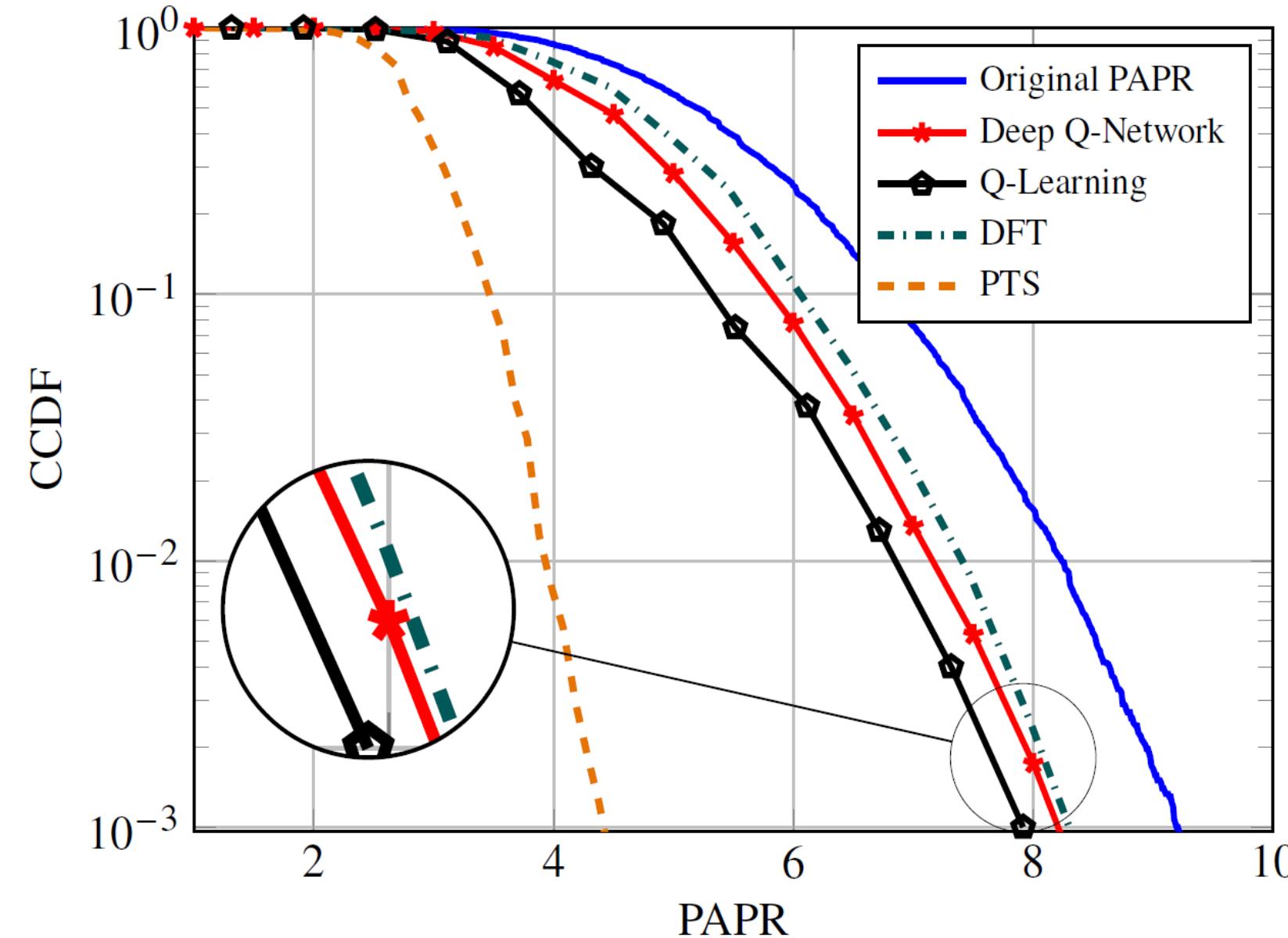
# Parâmetros de Simulação

Parâmetro	Q-Learning	DQN
Modulação	QPSK	QPSK
Subportadoras (N)	16	16
Pilotos ( $K_p$ )	2	2
Nº de estados	1.000	30.000
Aproximação Q	Tabela Q	MLP (256 neurônios)

# Parâmetros de Simulação

Parâmetro	Q-Learning	DQN
Replay buffer	—	1.000.000
Batch size	—	128
Learning rate	$\alpha$	0.0001
Discount factor ( $\gamma$ )	$\gamma$	0.9252
Exploração	$\varepsilon$ -greedy	$\varepsilon$ -greedy (1.0 → 0.05)
Target network	—	Atualização a cada 100 steps

# Resultados



# Conclusão

- Apresentamos uma nova abordagem para **redução de PAPR em sistemas OFDM** usando **DQN**.
- O método utiliza **duas subportadoras piloto** e alcançou **redução significativa de PAPR** em relação a técnicas tradicionais (DFT e PTS).
- Entre os métodos avaliados:
  - ✓ **Q-Learning** teve desempenho ligeiramente superior;
  - ✓ **DQN** apresentou melhor **generalização** e maior capacidade de **mitigação de picos de potência**.
- Houve uma **pequena degradação na BER** devido aos pilotos, mas o impacto é **mínimo** frente aos ganhos em PAPR.
- Trabalhos futuros incluem estender a abordagem para **cenários com mais subportadoras**, avaliando **escalabilidade e robustez**.

Obrigado!