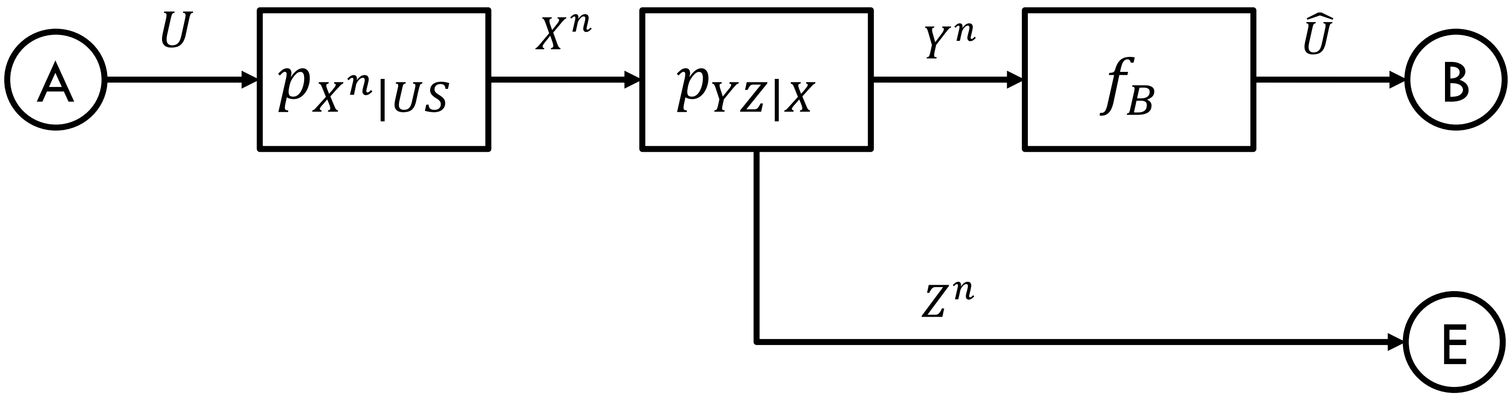
# Adversarial Networks for Secure Wireless Communications

Thomas Marchioro
Nicola Laurenti
Deniz Gündüz

# Physical layer secrecy

Wiretap channel: A wants to transmit $U$ to B, E has access to the channel, but with additional distortion




**Secrecy capacity** $C_s$ = maximum rate satisfying

1. Reliability: $\lim_{n\to\infty} P\left[\hat{U} \neq U\right] = 0$

2. Secrecy: $\lim_{n\to\infty} I(U, Z^n) = 0$

# A less stringent formulation

The condition

$$\lim_{n \to \infty} I(U, Z^n) = 0$$

might be too strict in some cases.

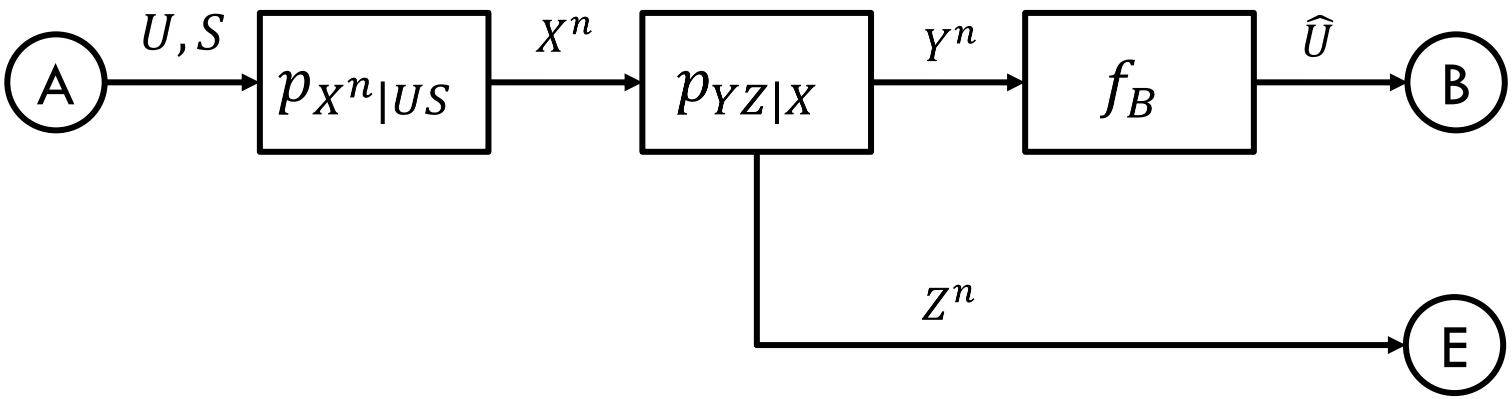
Example: A wants to transmit an image ($U$) representing a car to B but doesn't want E to know that it represents a car ($S$).

The image contains a lot of information, and not all that information is useful for classification.

# A less stringent formulation

Useful information $U$ to be transmitted to B

Sensitive information $S$ to be kept secret from E




**Physical layer secrecy**

- ❑ Reliability: $\lim_{n\to\infty} P[\hat{U} \neq U] = 0$
- ❑ Secrecy: $\lim_{n\to\infty} I(U; Z^n) = 0$

**Our problem**

- ❑ Quality: $\mathbb{E}[d(U, \hat{U})] \leq \varepsilon_n$
- ❑ Privacy: $I(S; Z^n) \leq \delta_n$

# Optimization problem

- Quality: $\mathbb{E}[d(U,\widehat{U})] \leq \varepsilon_n$
- Privacy: $I(S;Z^n) \leq \delta_n$

$$\min_{p_{X^n|US}, f_B} \mathbb{E}[d(U,\widehat{U})] + \alpha I(S;Z^n)$$

⇧ Tradeoff parameter

# Lower bound on $I(S;Z^n)$

Mutual information between $S$ and $Z^n$:

$$I(S;Z^n) = \sum_{s} p_S(s) \boxed{\sum_{z^n} p_{Z^n|S}(z^n|s) \log \frac{p_{Z^n|S}(z^n|s)}{p_{Z^n}(z^n)}}$$

Requires to estimate
conditional distributions

Alternative formulation:

$$\min_{p_{X^n|US}, f_B} \mathbb{E}[d(U,\widehat{U})] + \alpha \boxed{\left(\max_{Q_{S|Z^n}} \left(-H(e_S, q)\right)\right)}$$

Variational lower bound

$e_S$ = one hot encoding of $S$
$q$ = adversary likelihood estimation

# Minimax cross-entropy game

The problem

$$\min_{p_{X^n|US}, f_B} \left\{ \mathbb{E}\left[ d\left(U, \widehat{U}\right)\right] + \alpha \max_{Q_{S|Z^n}} \left(-H(e_S, q)\right)\right\}$$

can be interpreted as a minimax game.

- (A,B) needs to minimize

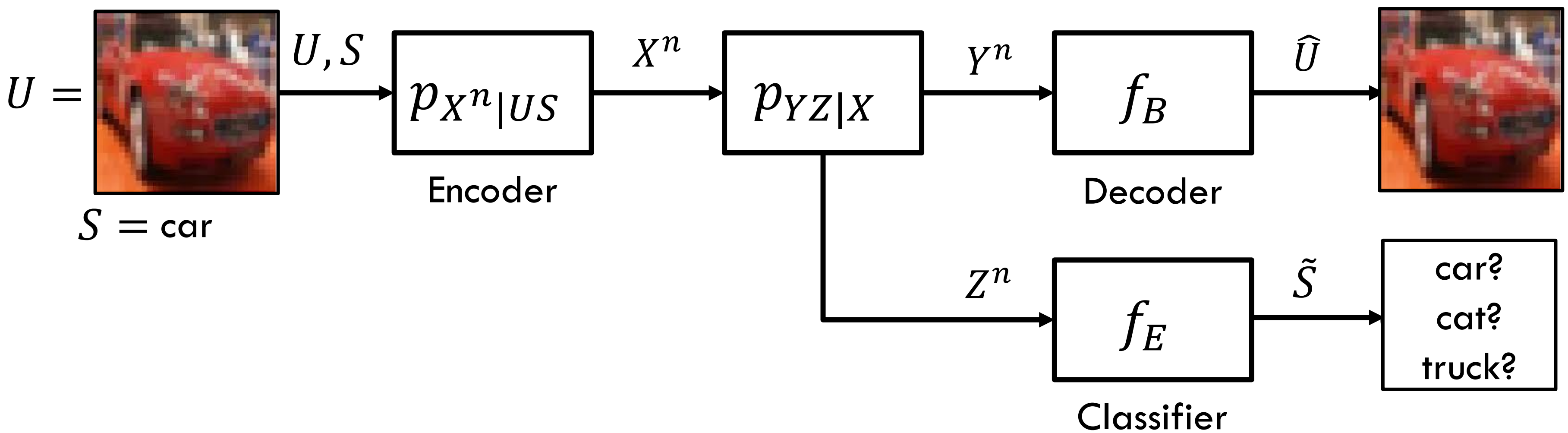
$$\mathcal{L}_{AB} = \mathbb{E}[d(U, \hat{U})] - \alpha H(e_S, q)$$

- E needs to minimize

$$\mathcal{L}_M = H(e_S, q)$$

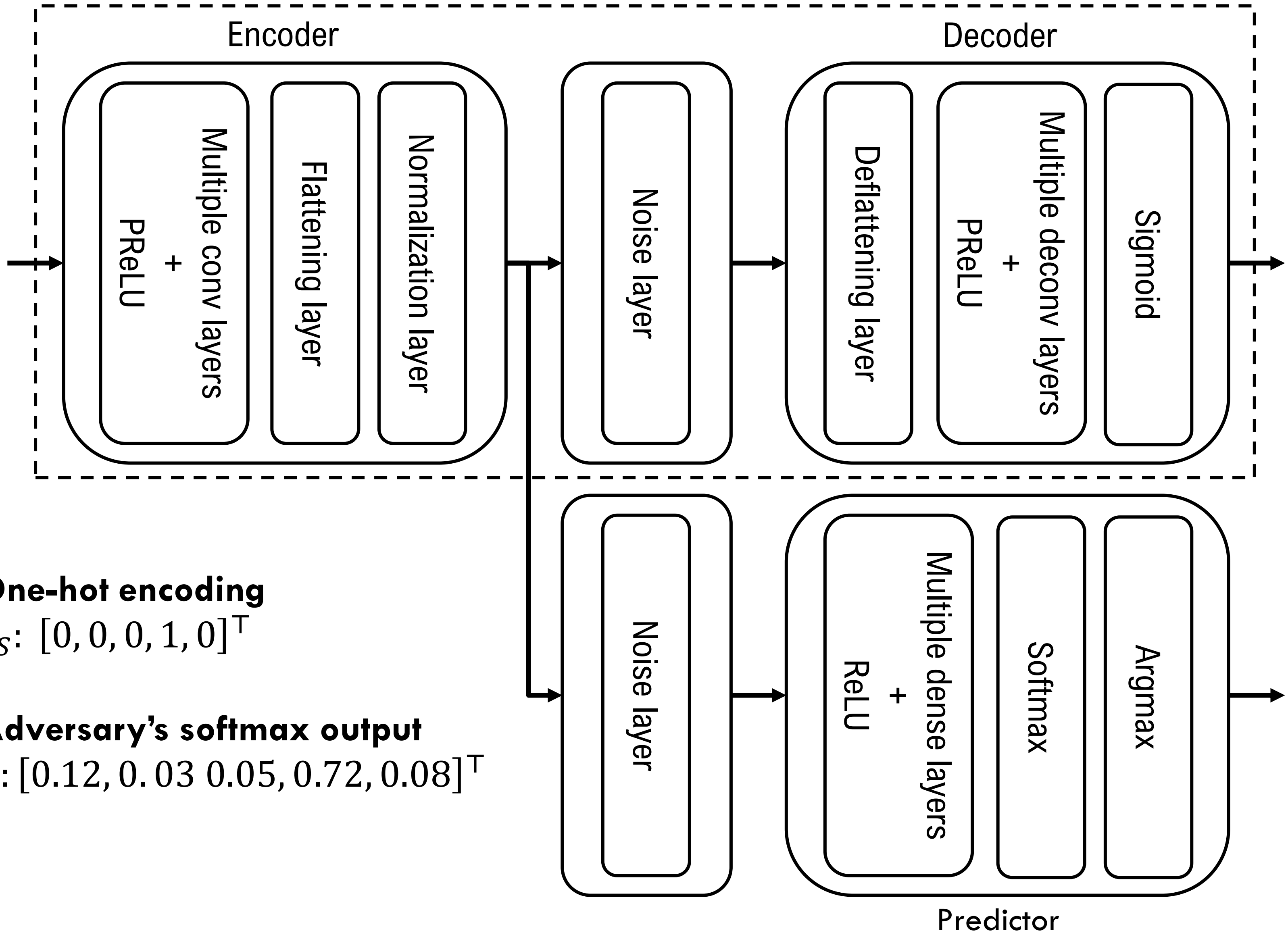
# Secure image transmission

Application: transmitting images while preventing the eavesdropper from correctly classifying the class.




$$\min_{p_{X^n|US}, f_B} \left\{ \mathrm{MSE}\left[d\left(U, \hat{U}\right)\right] + \alpha \max_{Q_{S|Z^n}} \left(-H\left(e_S, q\right)\right) \right\}$$

# Adversarial network model




**One-hot encoding**

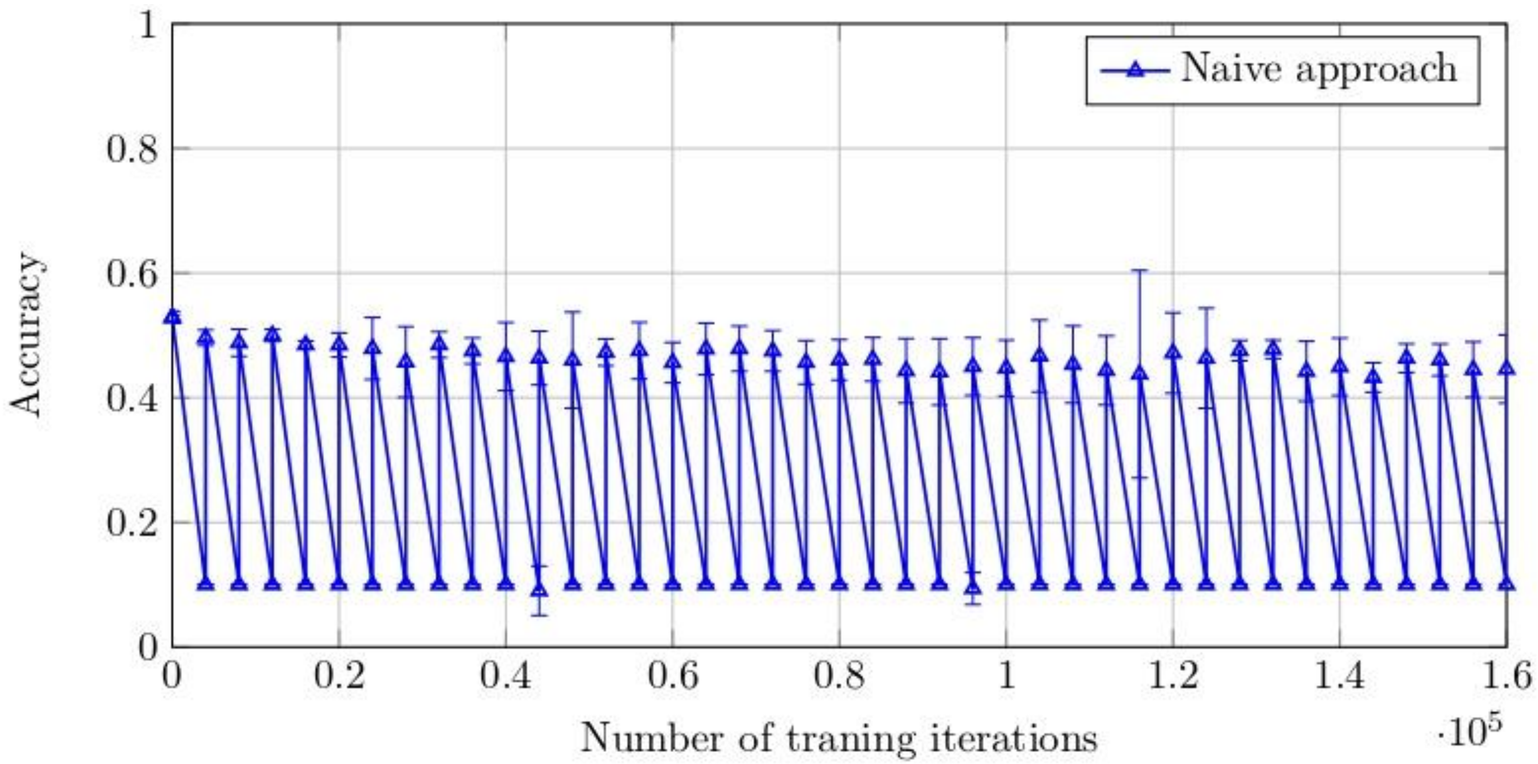
$e_S$: $[0, 0, 0, 1, 0]^\top$

**Adversary's softmax output**

$q$: $[0.12, 0.03\ 0.05, 0.72, 0.08]^\top$

# Stability issues

At each training cycle E's estimation is brought to be independent of $S$ after training (A,B), then the subsequent training of E partly recovers the missing information.



# Softmax equalization

Main idea: rather than maximizing the cross-entropy between the one-hot encoding and the softmax, minimize the cross-entropy between the distribution $\bar{p}$ and the softmax, where

$$\bar{p} = \left[\frac{1}{\ell}, \frac{1}{\ell}, \dots \frac{1}{\ell}\right]^{\mathsf{T}}, \qquad \ell = \#\ \text{of classes}$$

- ❑ (A,B) needs to minimize

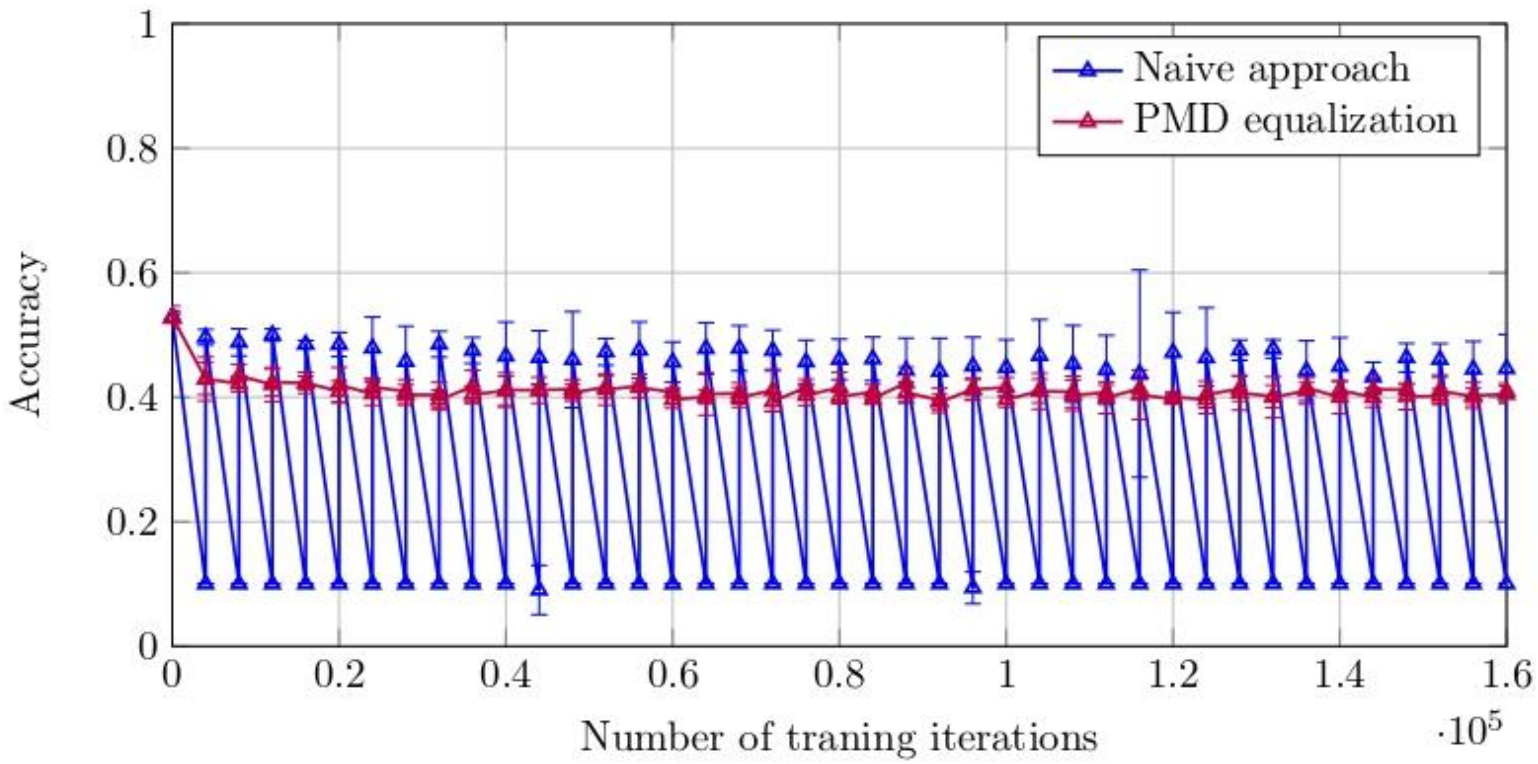
$$\mathcal{L}_{AB} = \mathbb{E}[d(U, \hat{U})] + \alpha H(\bar{p}, q)$$

- ❑ E needs to minimize

$$\mathcal{L}_{M} = H(e_S, q)$$

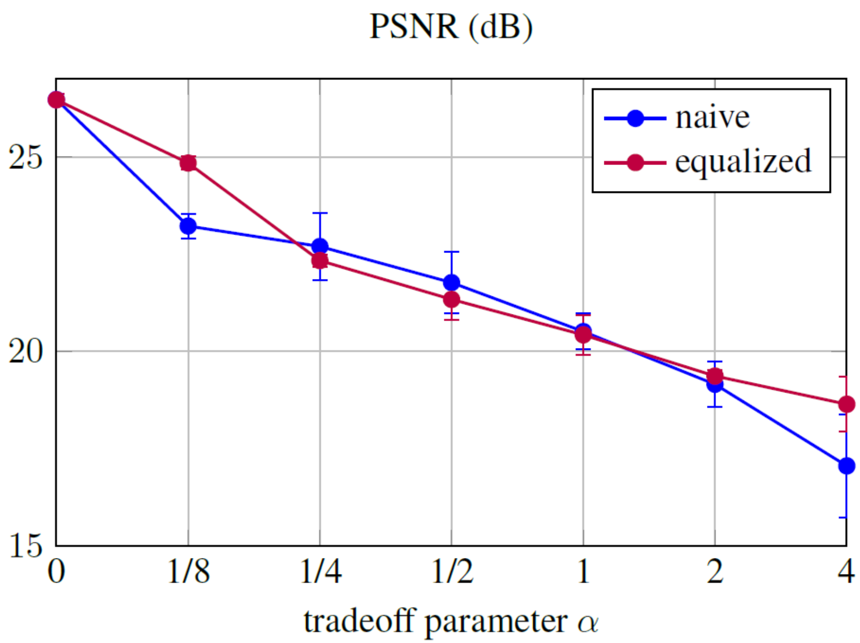
# Training results

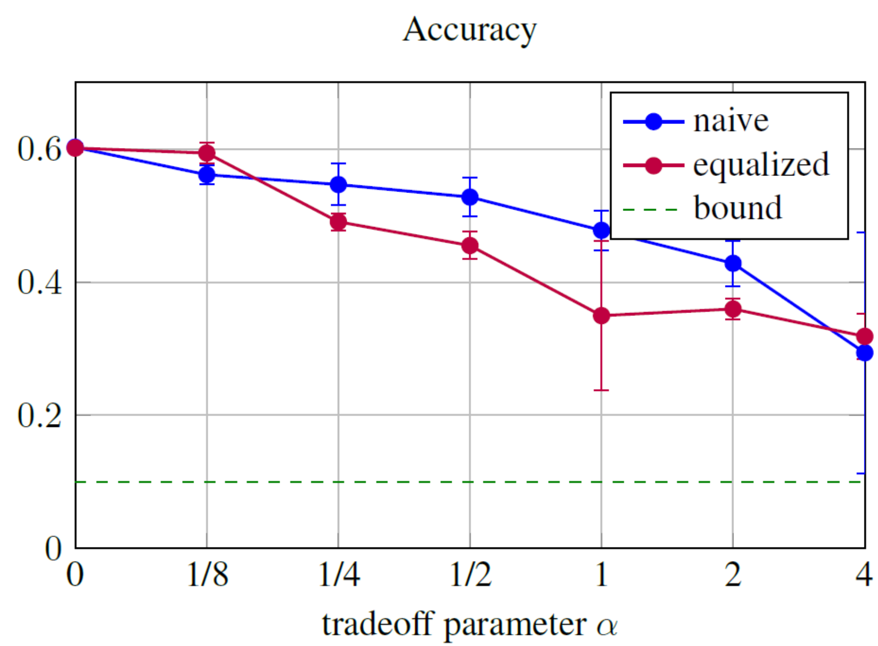
Softmax equalization is more stable and the results are subject to less variance.



# Test results

Main parameters: quality-privacy tradeoff α, SNR of E.
SNR of (A, B) = 10 dB.




⇧

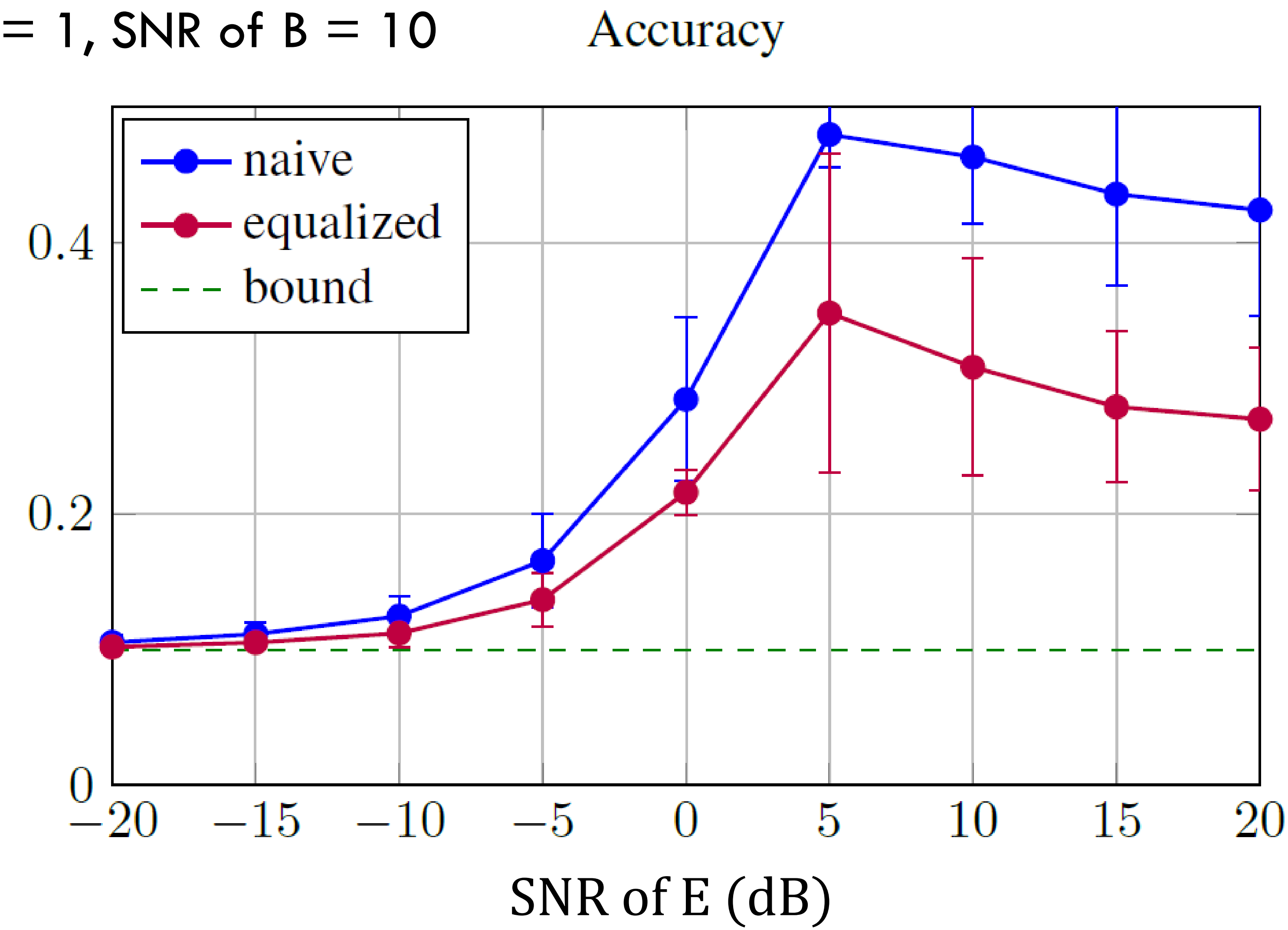
Quality measure:
the higher the better

⇧

Privacy measure:
the lower the better

# Test results

α = 1, SNR of B = 10



# Conclusions

We have:

- introduced a relaxed privacy condition with respect to physical layer secrecy to protect sensitive information only
- proposed a general formulation of the corresponding minimax problem
- applied this formulation to secure image transmission employing **adversarial neural networks**
- shown that it is possible to regulate the tradeoff between quality and privacy and to exploit the channel advantage to achieve better secrecy.

# Future work

- Train the model with fading channels to improve the scalability for SNR variations
- Introduce a stochastic encoder to improve the quality-privacy tradeoff

THANK YOU!