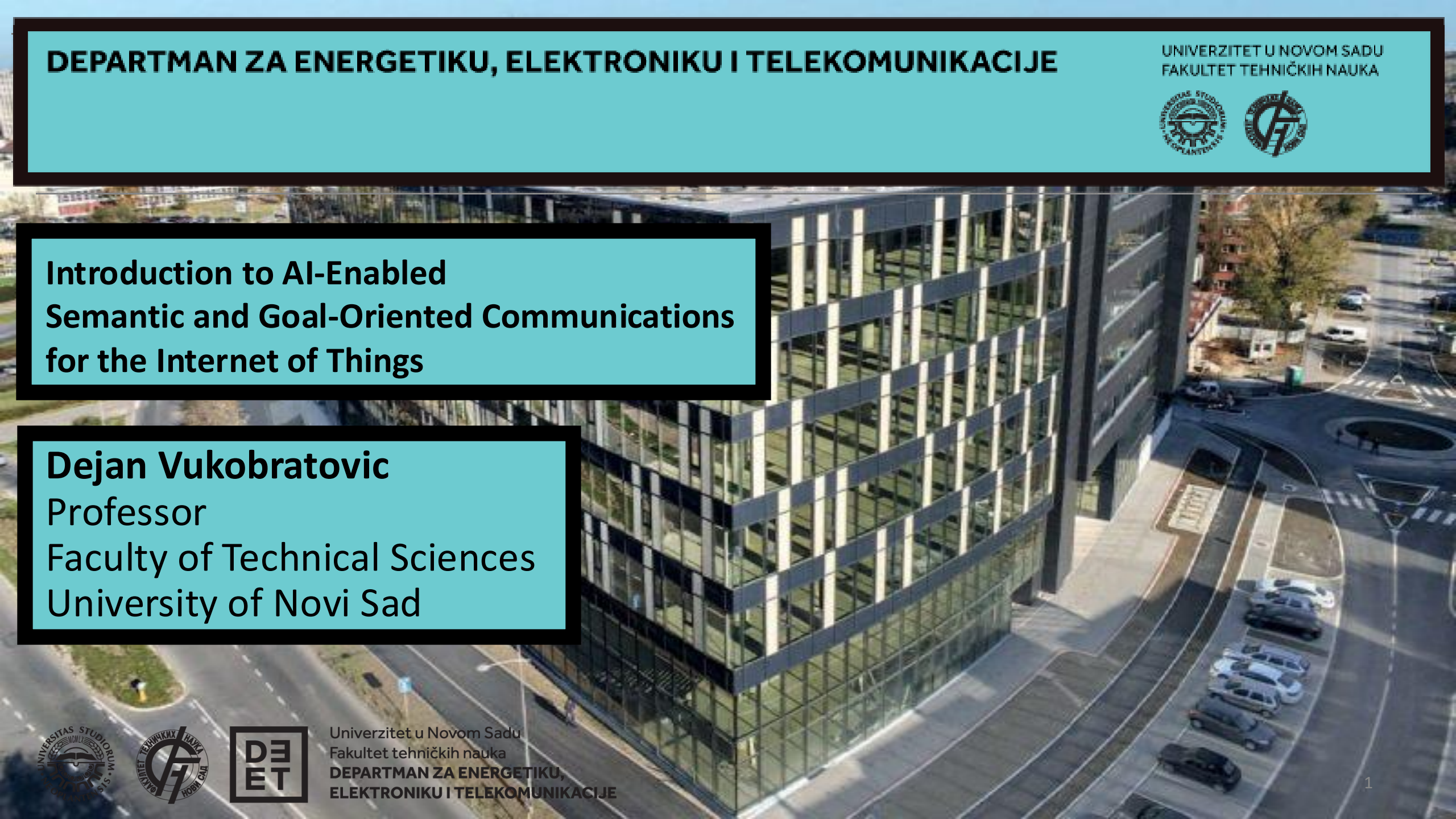
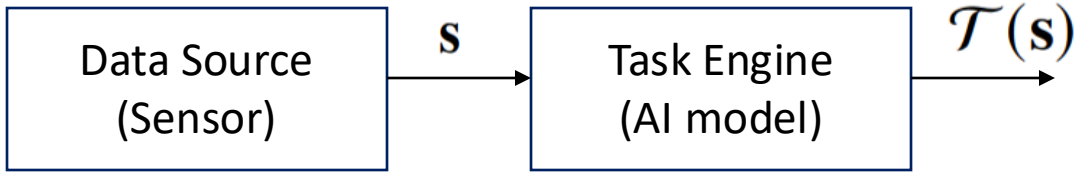


Introduction to AI-Enabled Semantic and Goal-Oriented Communications for the Internet of Things

Dejan Vukobratovic
Professor
Faculty of Technical Sciences
University of Novi Sad

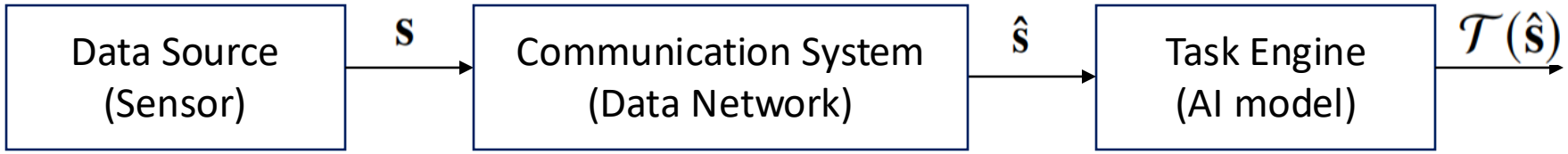
Goal-Oriented Communications

Data source and task engine collocated



Task engine trained to minimize task-specific expected loss function

Data source and task engine separated



Task-oriented (goal-oriented) communications

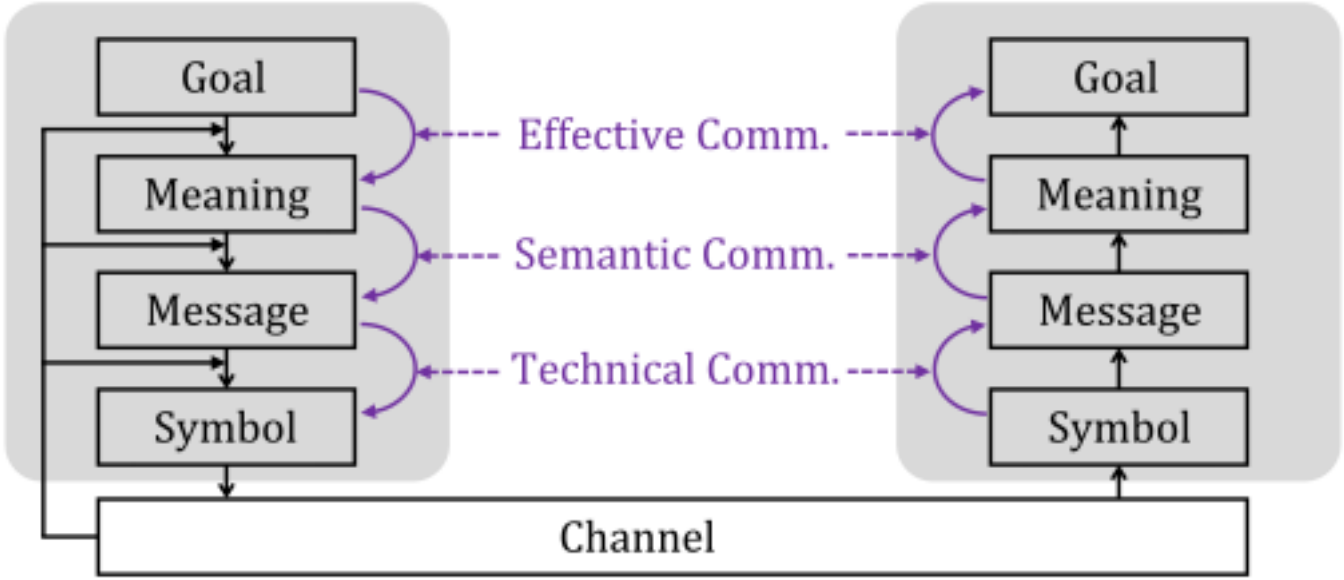
Design Communication System with the goal of $\mathcal{T}(\hat{\mathbf{s}})$ being equal to $\mathcal{T}(\mathbf{s})$

Traditional Shannon-theory communications

Design Communication System with the goal of $\hat{\mathbf{s}}$ being equal to $\mathbf{s}$

C. E. Shannon and W. Weaver, The Mathematical Theory of Communication. Urbana, IL, USA: Univ. Illinois Press, 1949.

Goal-Oriented Communications



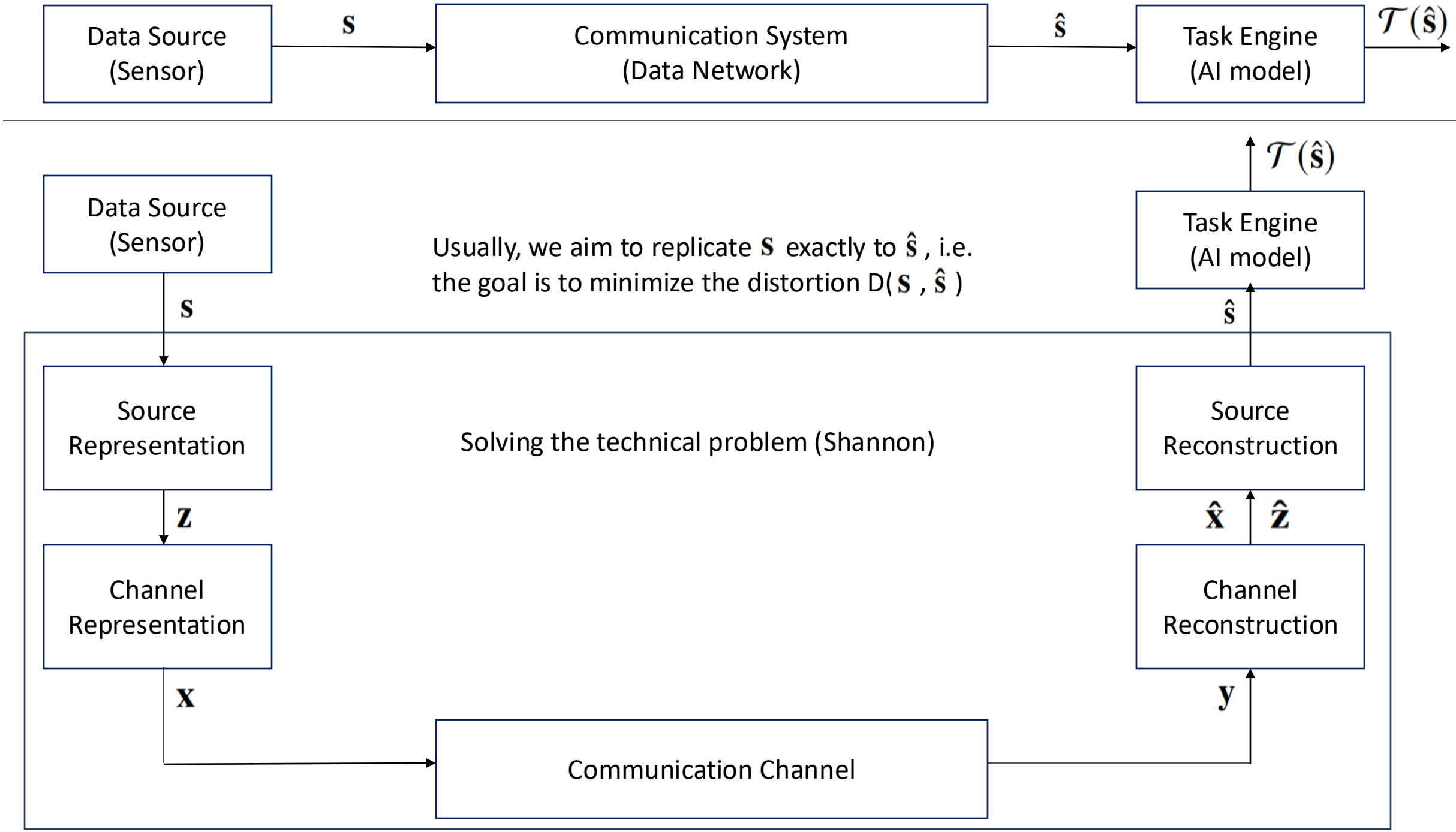
The effectiveness problem: How effectively does the received meaning affect desired conduct?

The semantic problem: How precisely do the transmitted symbols convey the desired meaning?

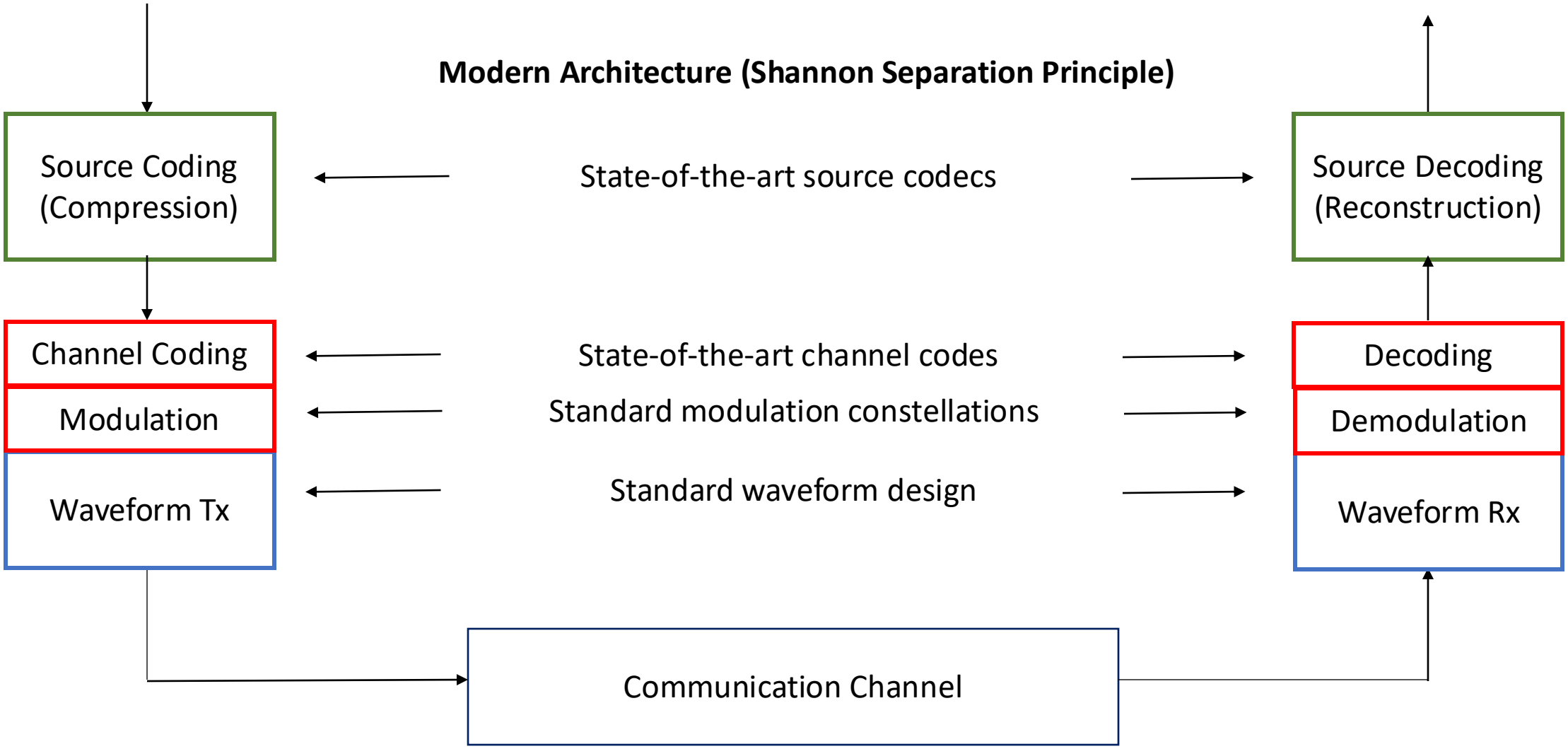
The technical problem: How accurately can the symbols of communication be transmitted?

Traditional Shannon-theory communications solves only the technical problem of communications.

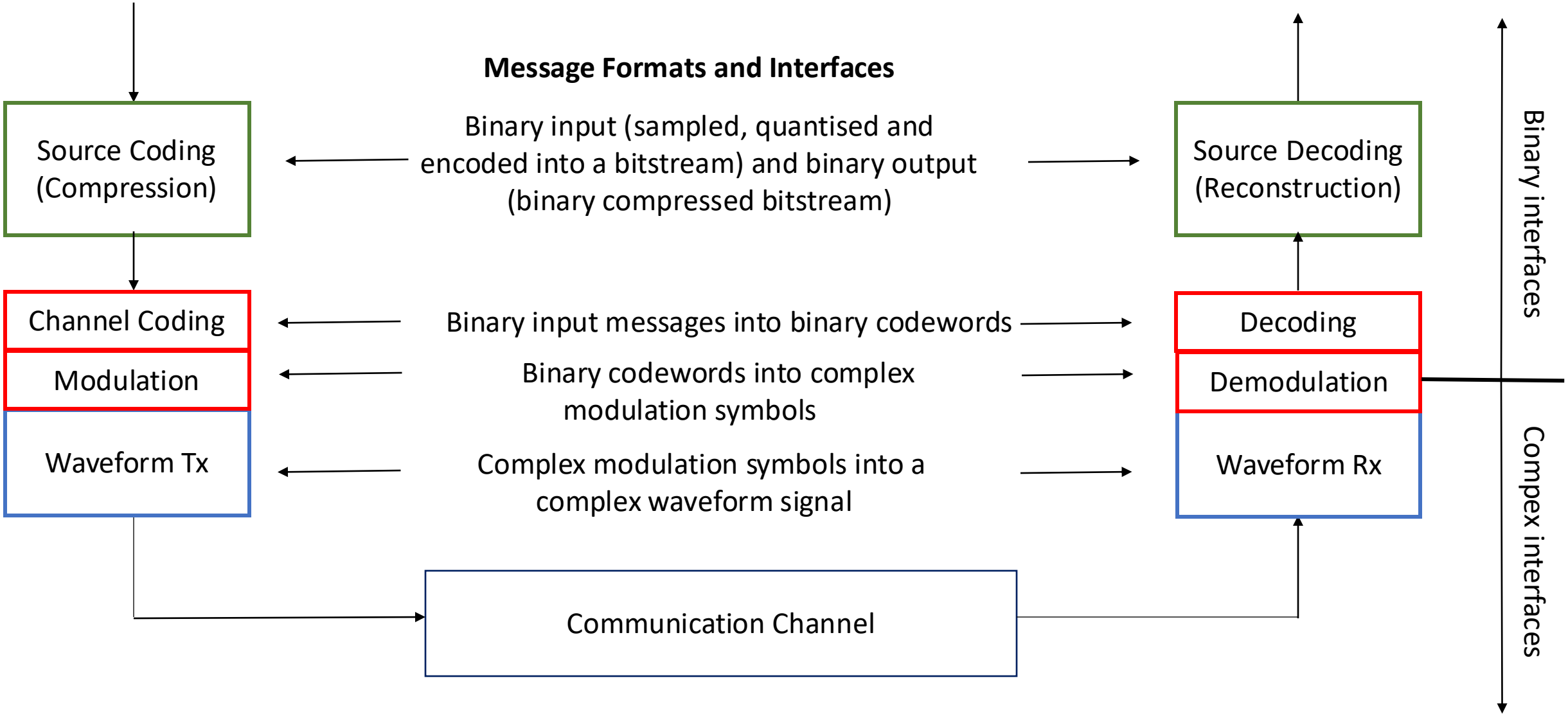
C. E. Shannon and W. Weaver, The Mathematical Theory of Communication. Urbana, IL, USA: Univ. Illinois Press, 1949.


Data Source
(Sensor)
s
Communication System
(Data Network)
ŝ
Task Engine
(AI model)
𝒯(ŝ)
Data Source
(Sensor)
s
Usually, we aim to replicate s exactly to ŝ , i.e.
the goal is to minimize the distortion D(s , ŝ)
Solving the technical problem (Shannon)
Source
Representation
z
Channel
Representation
x
Communication Channel
y
Channel
Reconstruction
x̂
ẑ
Source
Reconstruction
ŝ
Task Engine
(AI model)
𝒯(ŝ)

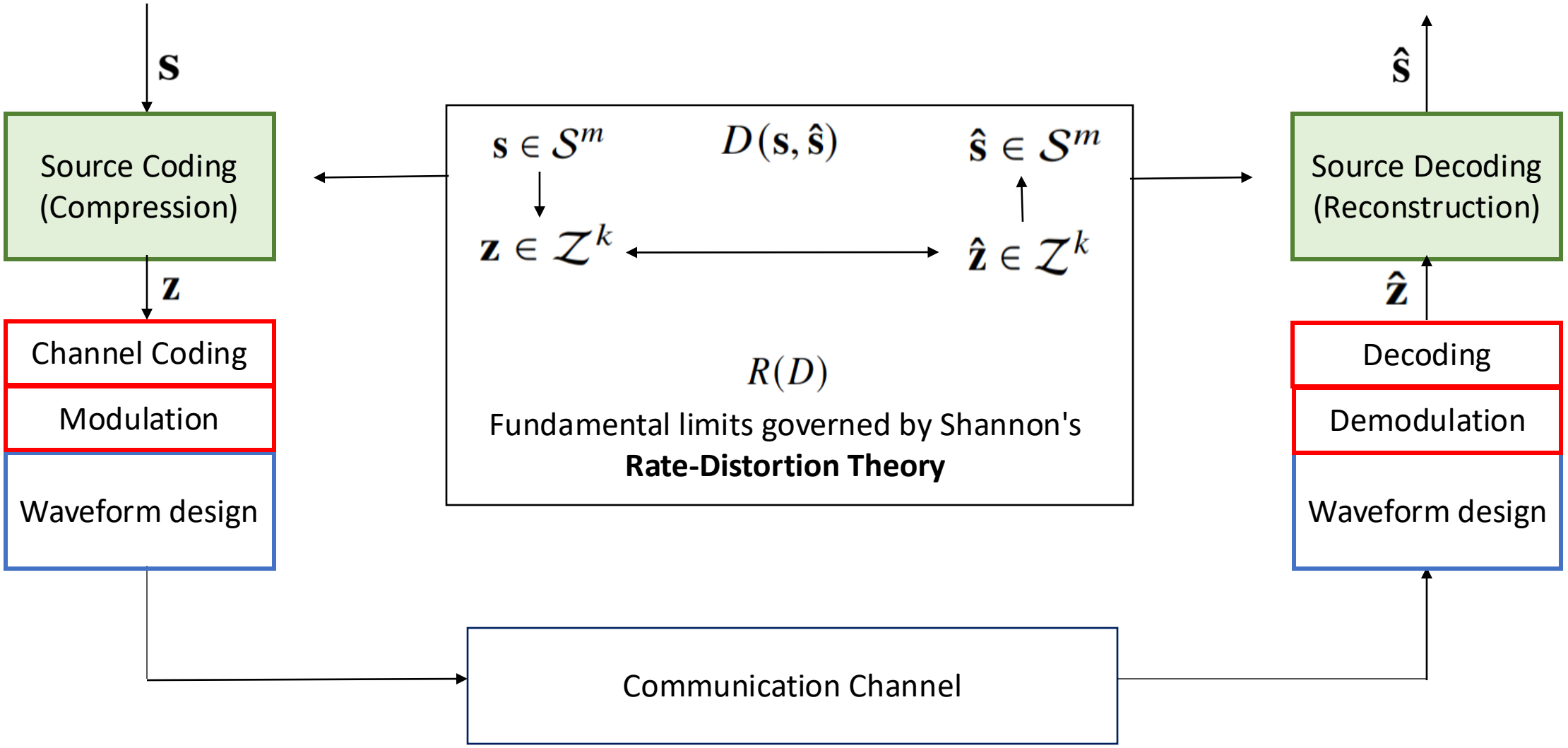
Solving Classical Shannon Technical Problem



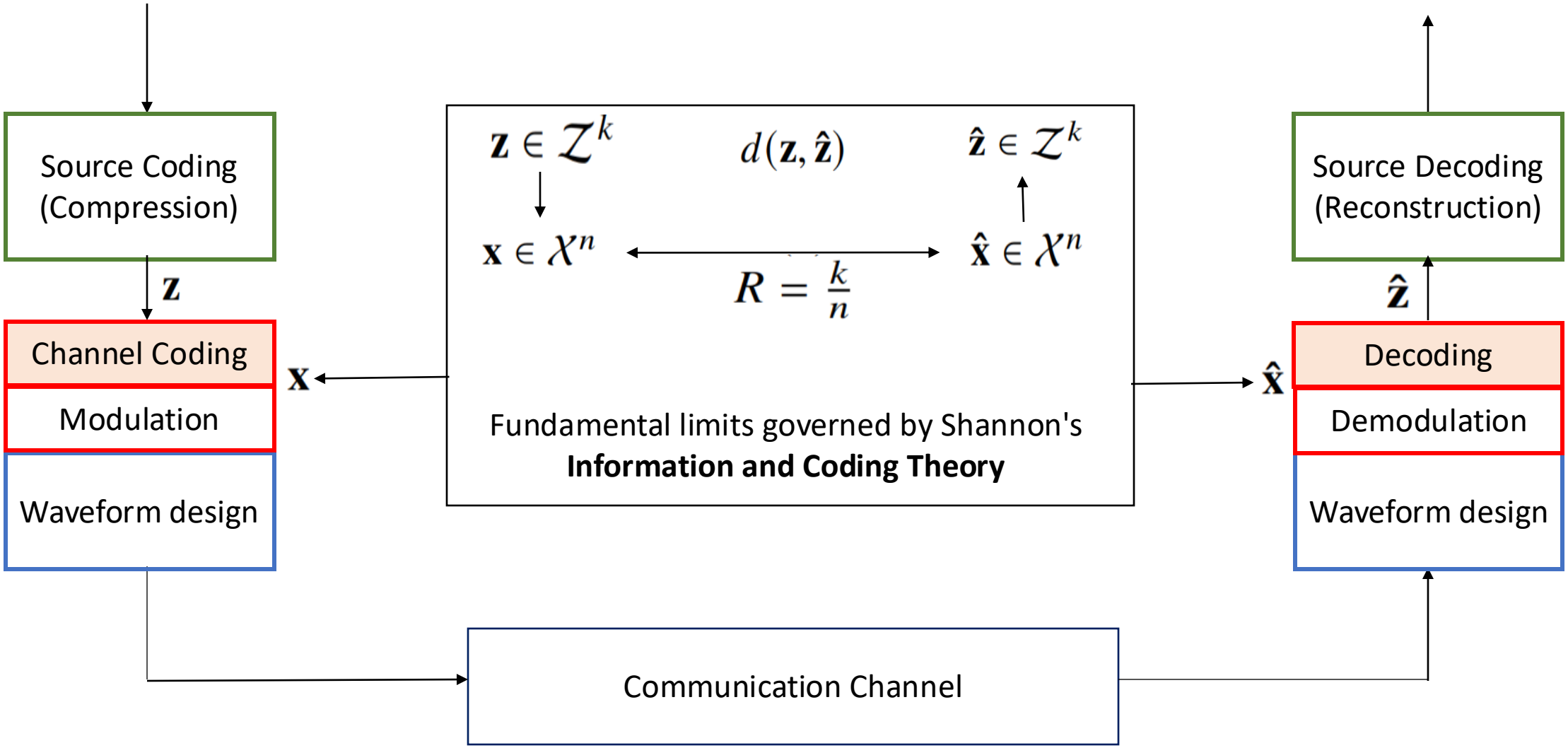
Solving Classical Shannon Technical Problem



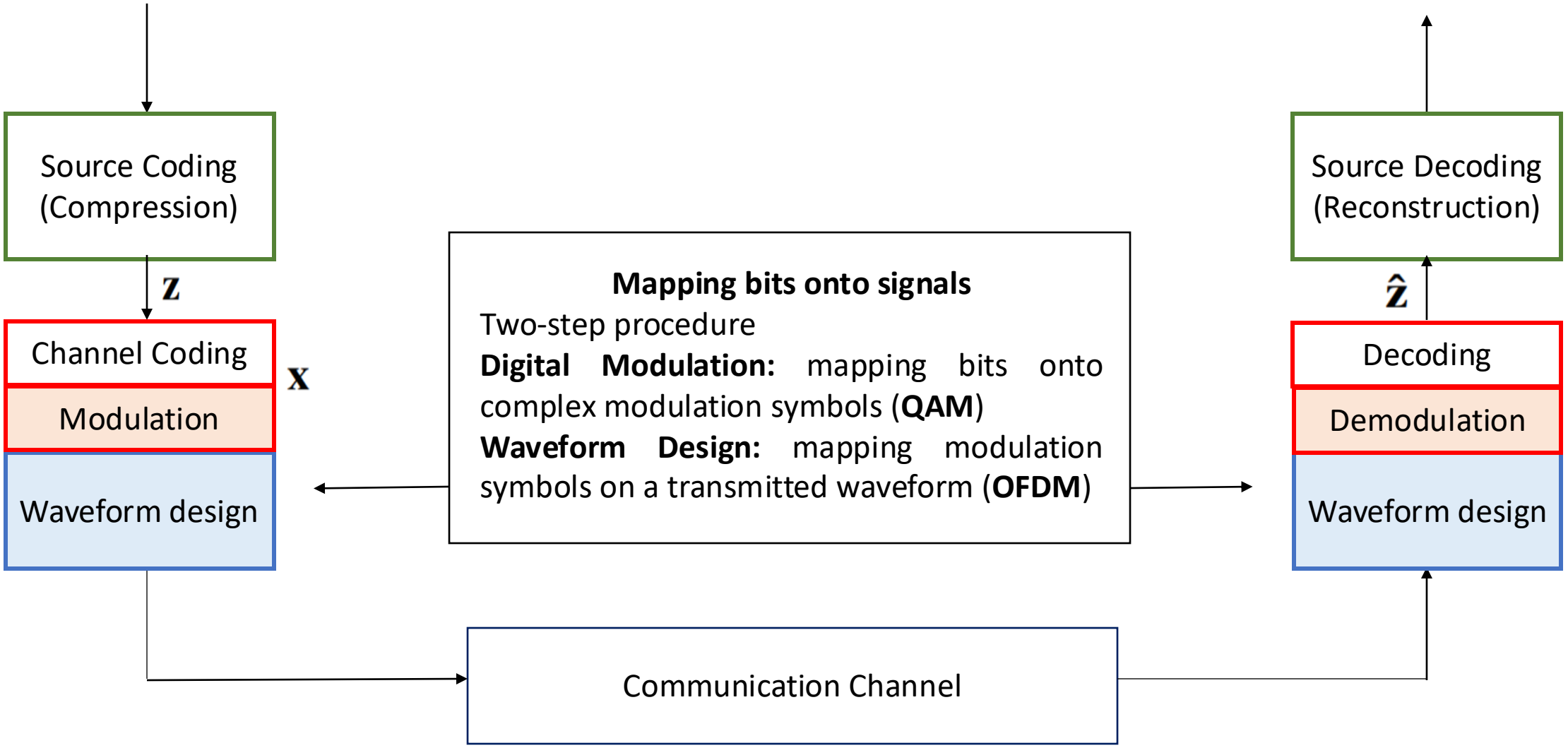
Solving Classical Shannon Technical Problem



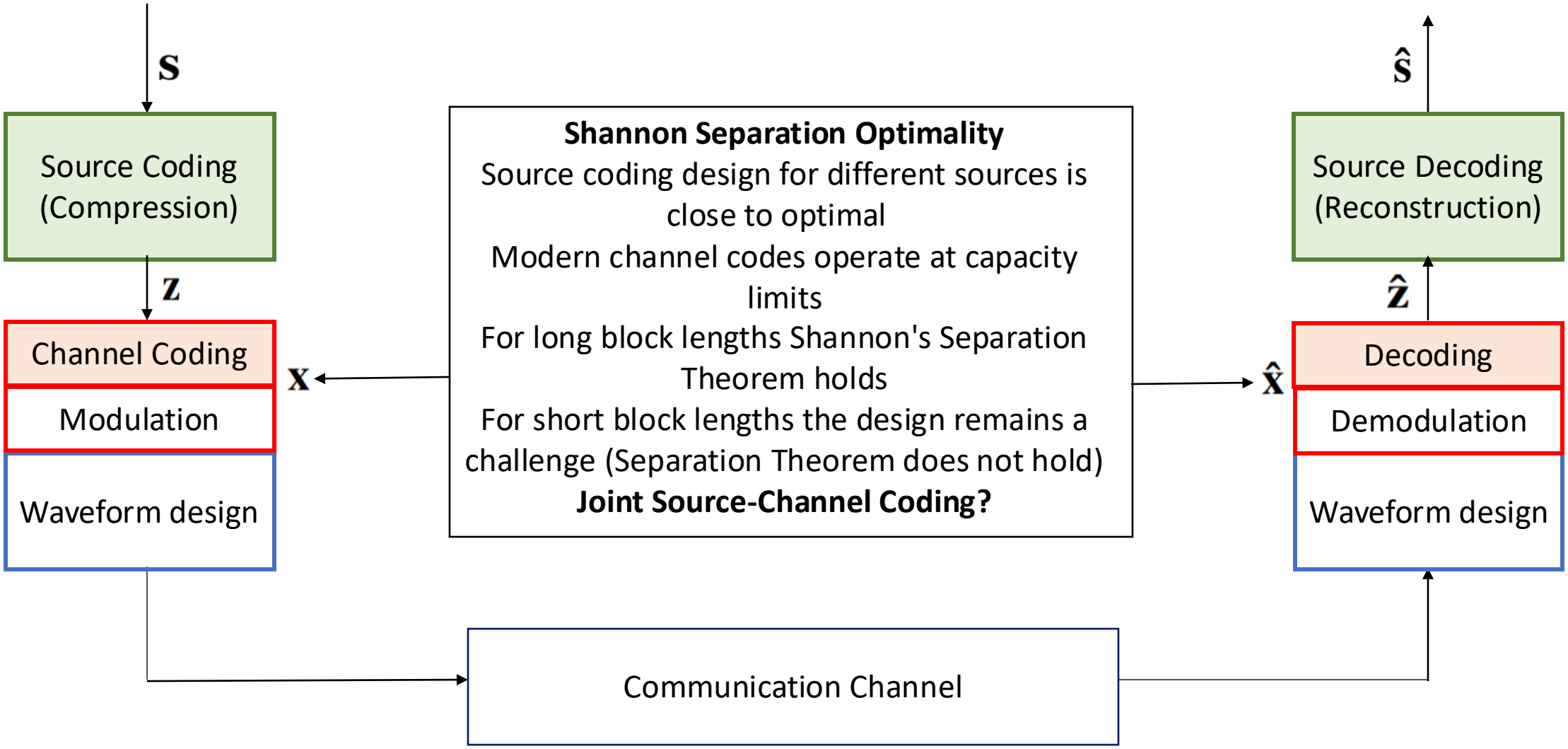
Solving Classical Shannon Technical Problem

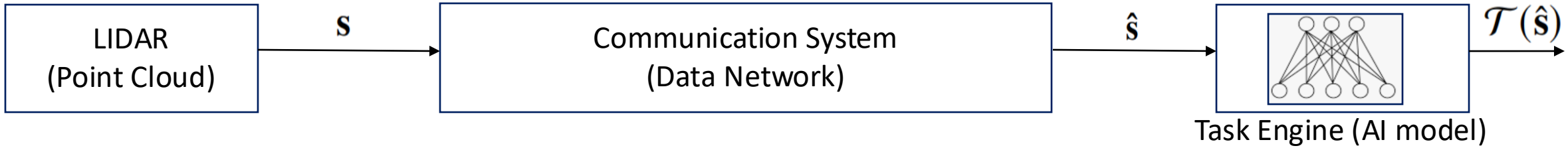


Solving Classical Shannon Technical Problem



Solving Classical Shannon Technical Problem

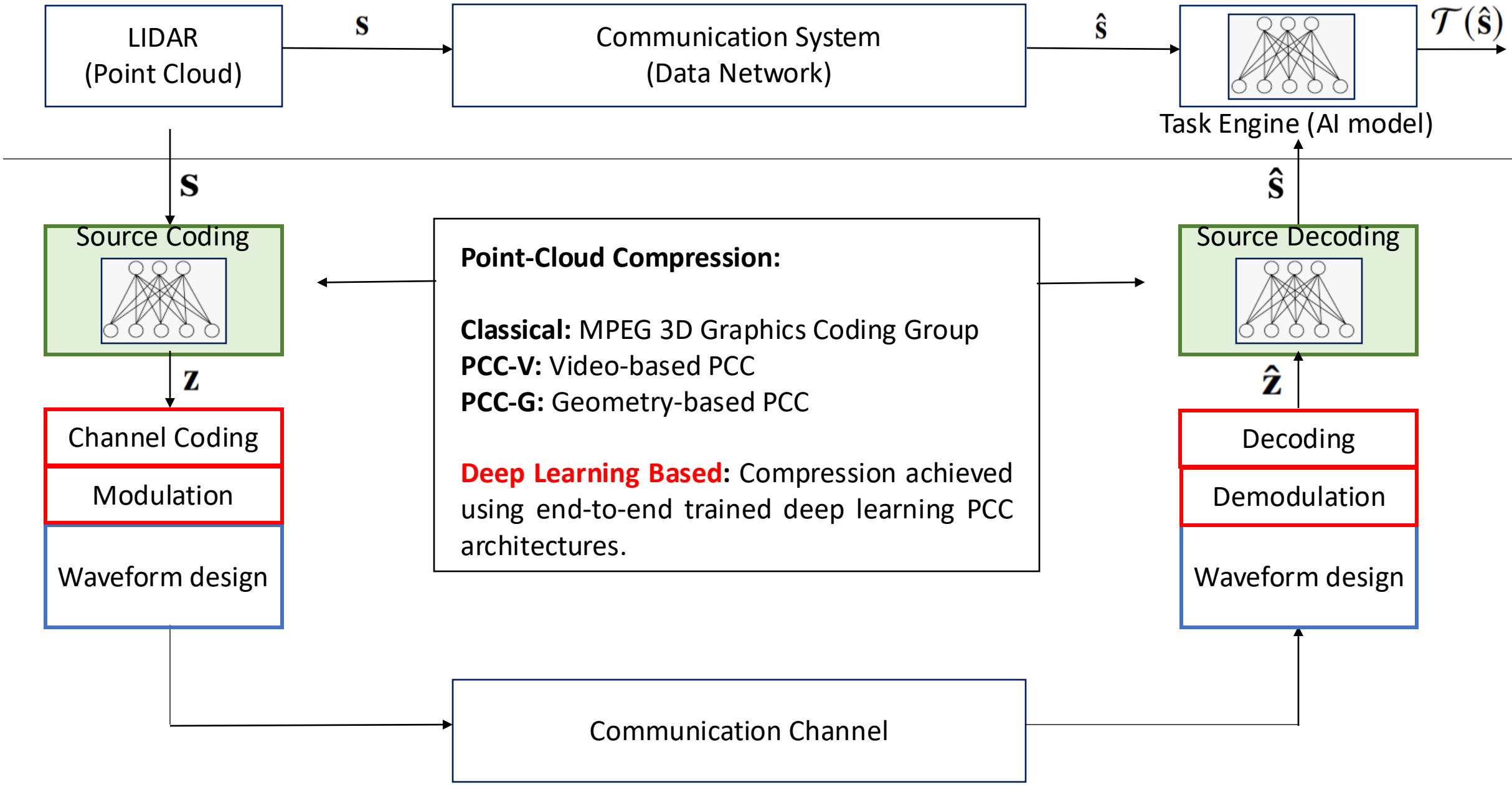


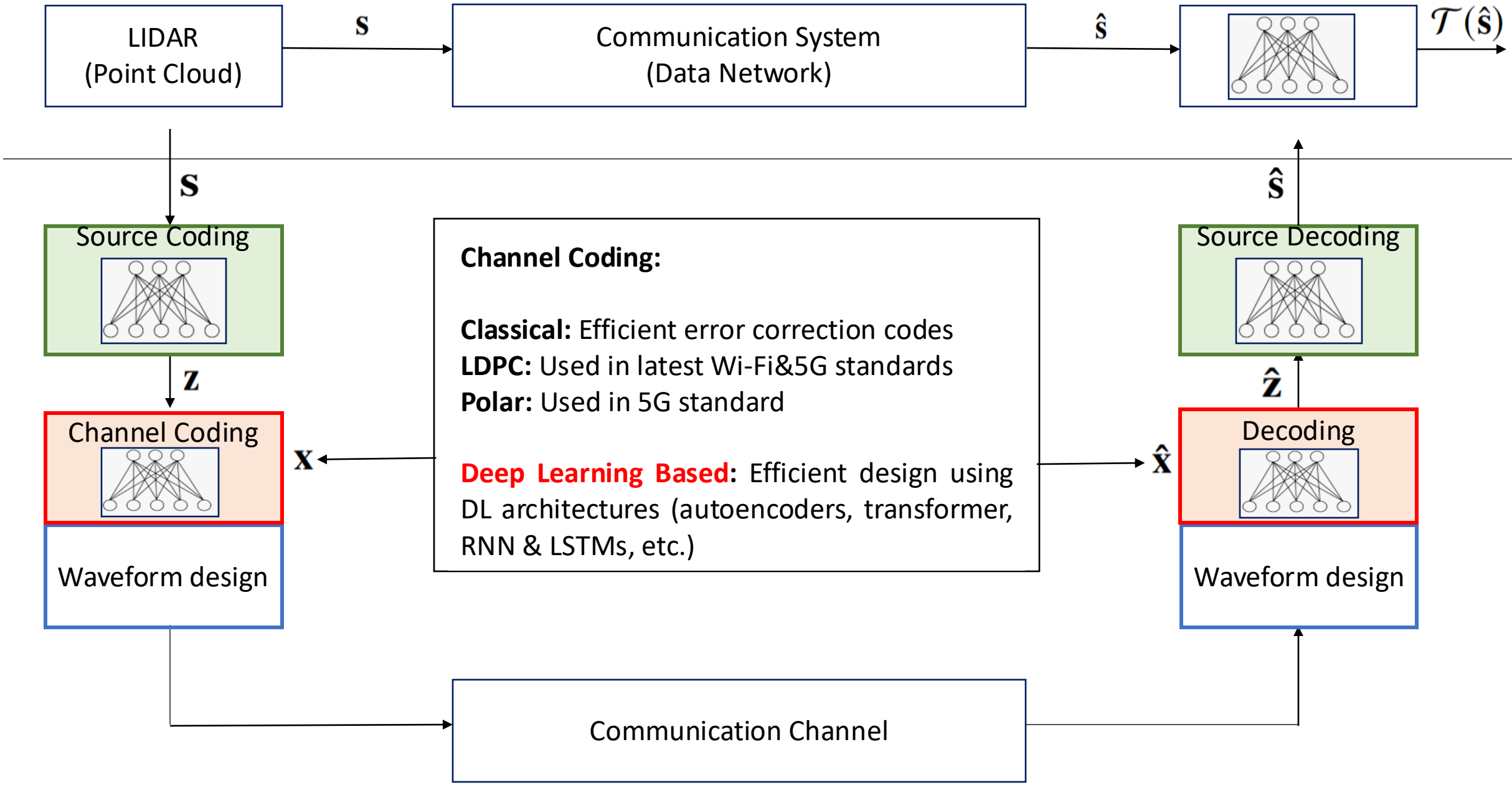


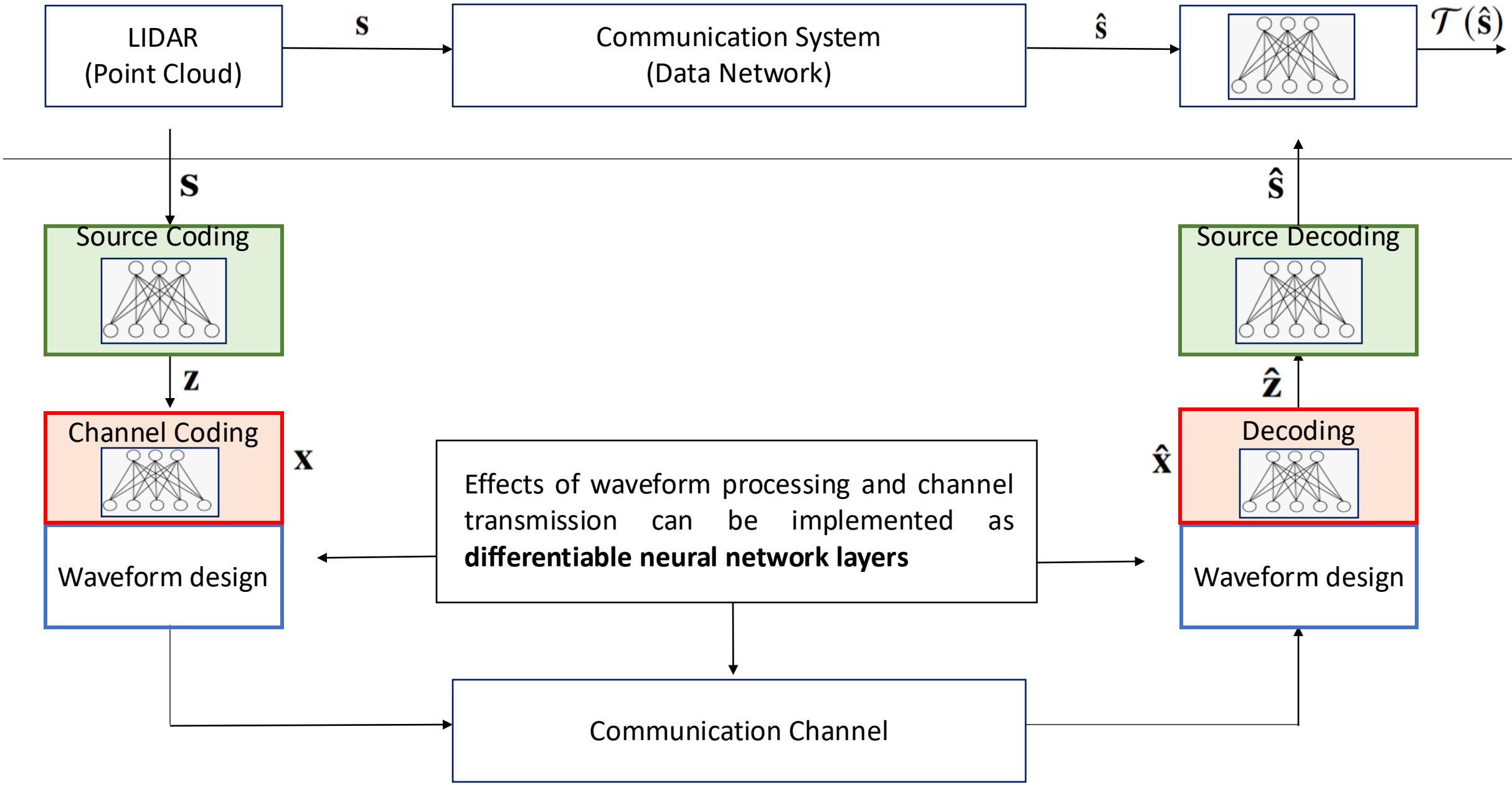
Key Question: How to represent source information $\mathbf{s}$ efficiently in order to achieve the same (or as close) task performance on remotely received data $\hat{\mathbf{s}}$ as it would be achieved if the task is executed on the original data?

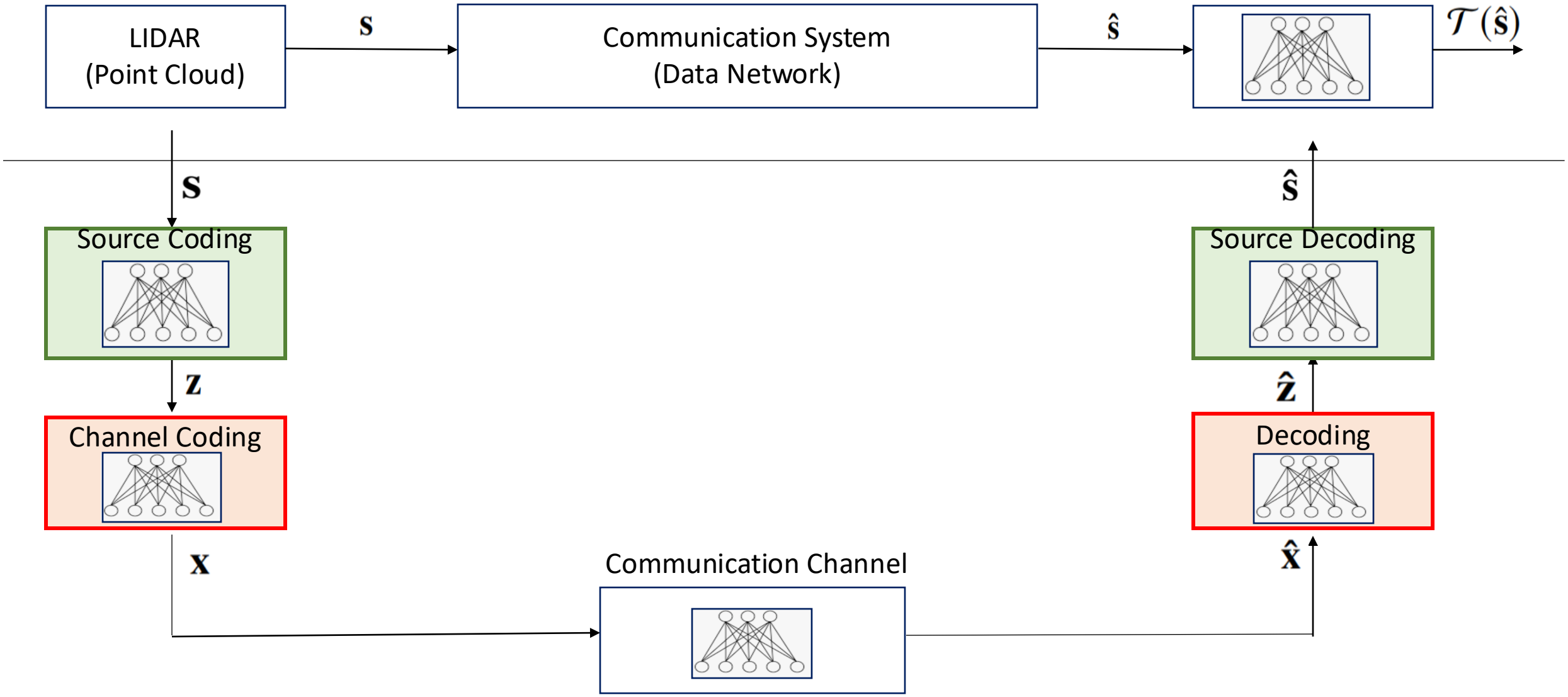
Question? Is attempting to replicate source data at the receiver side (Shannon technical communication problem) a good strategy?

Comment: Efficient source data semantic representations could be task-dependent!


LIDAR
(Point Cloud)
s
Communication System
(Data Network)
ŝ
Task Engine (AI model)
T(ŝ)
s
Source Coding
z
Channel Coding
Modulation
Waveform design
Point-Cloud Compression:
Classical: MPEG 3D Graphics Coding Group
PCC-V: Video-based PCC
PCC-G: Geometry-based PCC
Deep Learning Based: Compression achieved using end-to-end trained deep learning PCC architectures.
ŝ
Source Decoding
ẑ
Decoding
Demodulation
Waveform design
Communication Channel

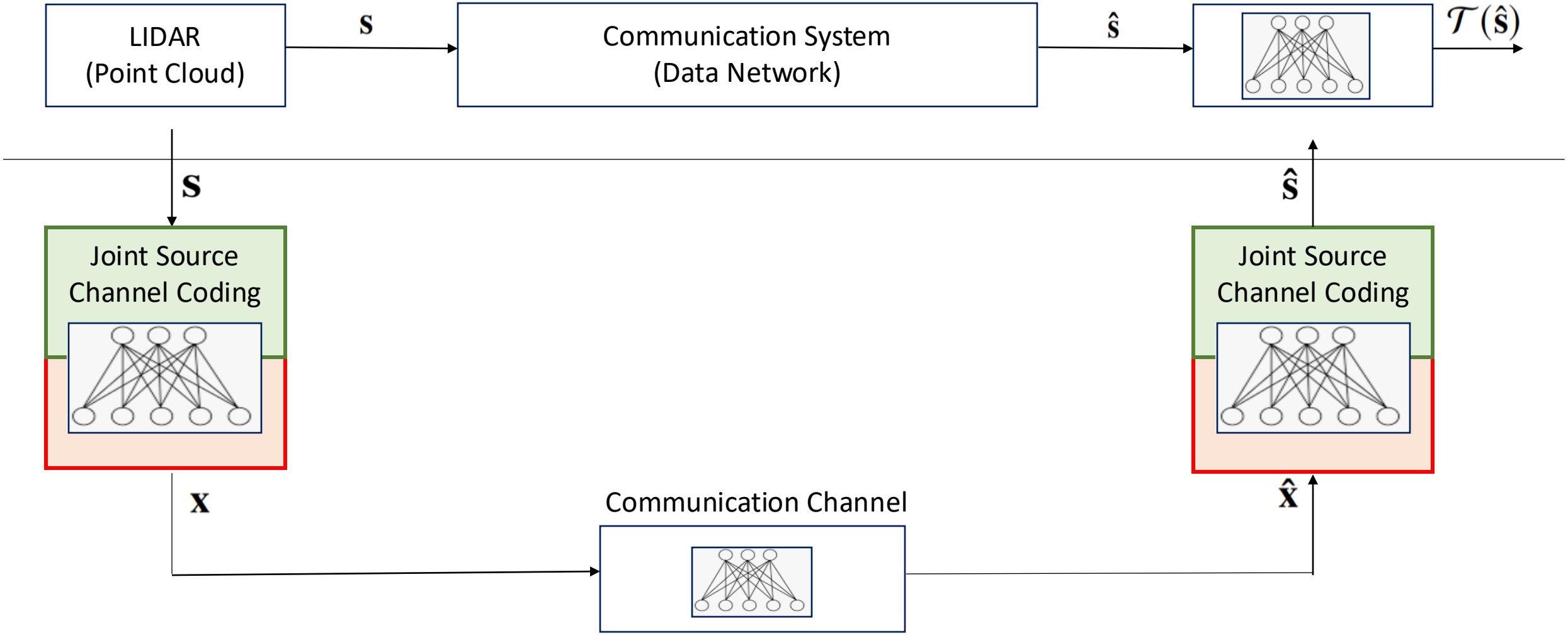

LIDAR
(Point Cloud)
s
Communication System
(Data Network)
ŝ
$\mathcal{T}(\hat{\mathbf{s}})$
s
Source Coding
z
Channel Coding
Waveform design
x
Channel Coding:
Classical: Efficient error correction codes
LDPC: Used in latest Wi-Fi&5G standards
Polar: Used in 5G standard
Deep Learning Based: Efficient design using DL architectures (autoencoders, transformer, RNN & LSTMs, etc.)
Communication Channel
Waveform design
Decoding
ẑ
Source Decoding
ŝ


LIDAR
(Point Cloud)
s
Communication System
(Data Network)
ŝ
T(ŝ)
s
Source Coding
z
Channel Coding
x
Waveform design
Effects of waveform processing and channel transmission can be implemented as **differentiable neural network layers**
Communication Channel
ŝ
Source Decoding
ẑ
Decoding
x̂
Waveform design



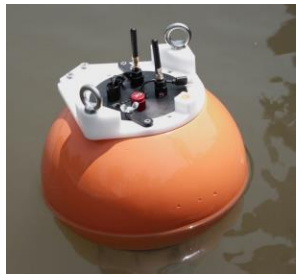
The whole communication and learning setup is one big neural network!

Can we learn it through appropriate architecture design and end-to-end training?



Deep Joint Source-Channel Coding (DJSCC) demonstrates improved performance

However, DJSCC still aims to replicate the source message at the receiver



Smart Buoy
(Time Series)

$\mathbf{s}$

Communication System
(Data Network)

$\hat{\mathbf{s}}$

Prediction Task

$\mathcal{T}(\hat{\mathbf{s}})$

$\mathbf{s}$

Source Coding

$\mathbf{z}$

Channel Coding

$\mathbf{x}$

Communication Channel

$\hat{\mathbf{x}}$

Decoding

$\hat{\mathbf{z}}$

Source Decoding

$\hat{\mathbf{s}}$

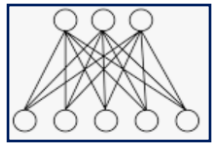
Object 3D
Reconstruction Task

Robotic hand
(video/3DPC)

$\mathbf{s}$

Communication System
(Data Network)

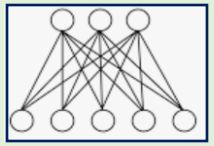
$\hat{\mathbf{s}}$



$\mathcal{T}(\hat{\mathbf{s}})$

$\mathbf{s}$

Source Coding



$\mathbf{z}$

Channel Coding

$\mathbf{x}$

Communication Channel

$\hat{\mathbf{x}}$

Decoding

$\hat{\mathbf{z}}$

Source Decoding

$\hat{\mathbf{s}}$

Stereo Image & 3DPC acquisition

Compression / Semantic Encoder

Streaming Platform & Transport

3D model & pose reconstruction

Decompression / Semantic Decoder

Streaming Platform & Transport

Uplink

Robotic workcell & Control Server

3GPP/O-RAN Testbed

5GCN

O-RIC

O-CU

O-DU

UE

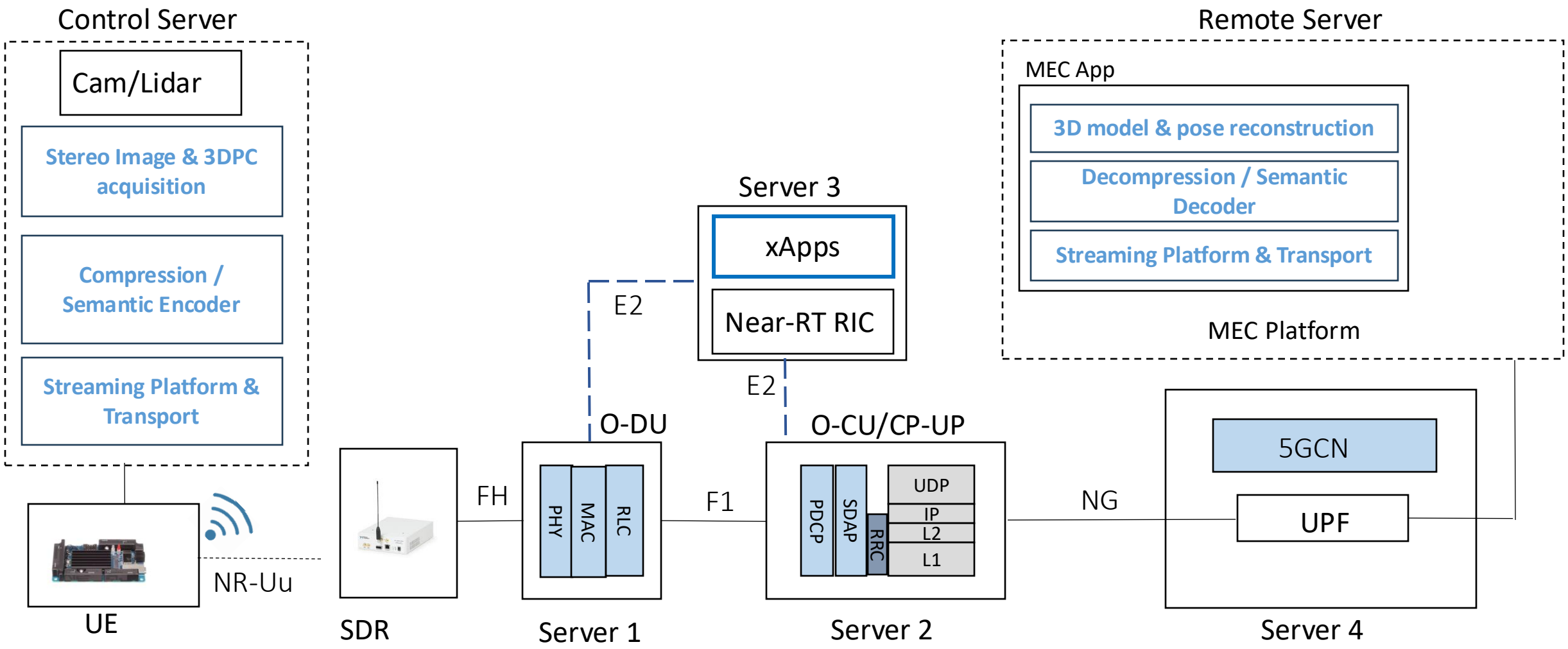
SDR

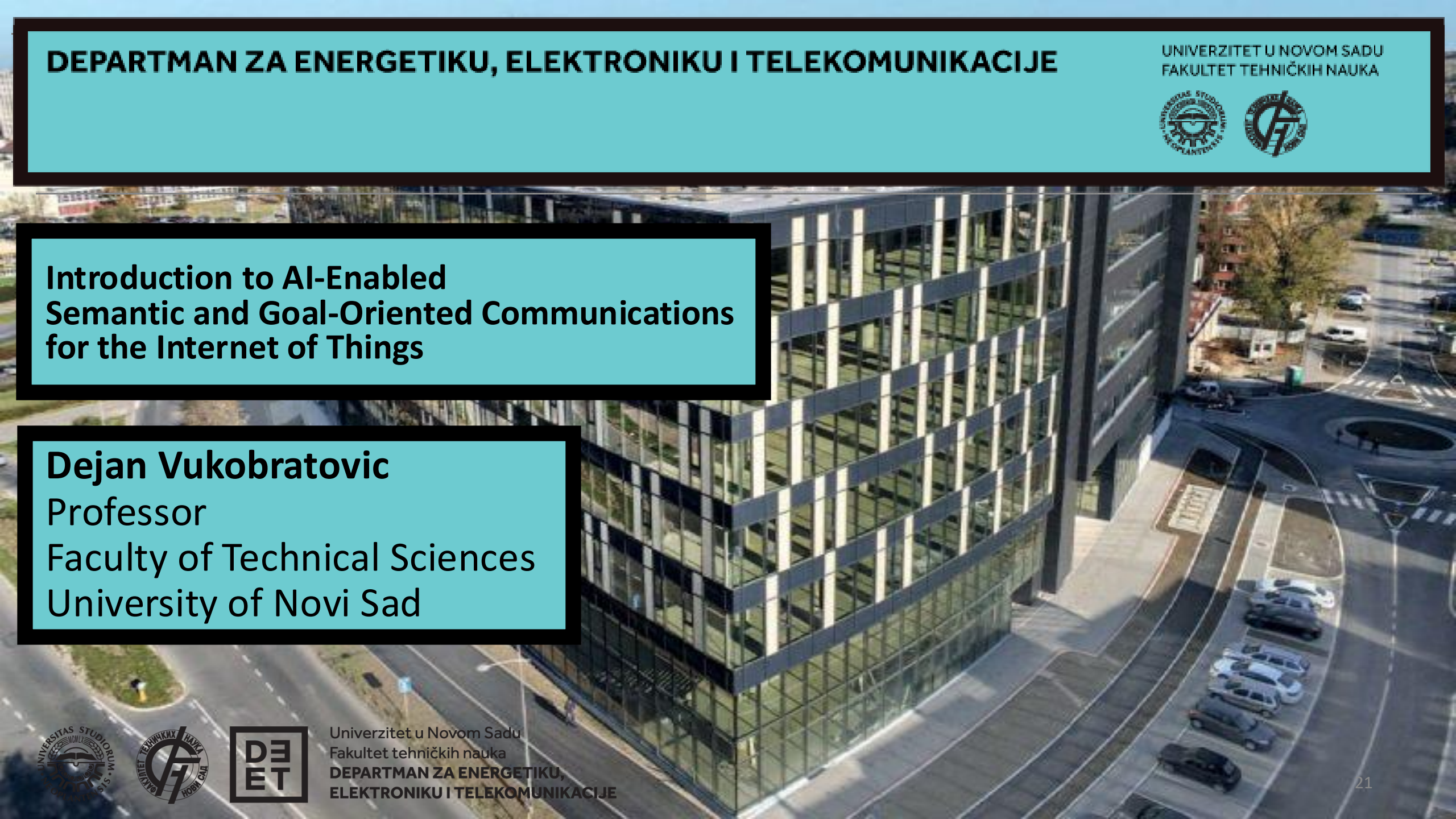
Remote Server

Downlink

Motion Commands & Task Execution

Motion Planning & Control


Control Server
Cam/Lidar
Stereo Image & 3DPC acquisition
Compression / Semantic Encoder
Streaming Platform & Transport
Remote Server
MEC App
3D model & pose reconstruction
Decompression / Semantic Decoder
Streaming Platform & Transport
MEC Platform
Server 3
xApps
Near-RT RIC
E2
E2
O-DU
O-CU/CP-UP
PHY
MAC
RLC
PDCP
SDAP
RRC
UDP
IP
L2
L1
FH
F1
NG
5GCN
UPF
NR-Uu
UE
SDR
Server 1
Server 2
Server 4



Introduction to AI-Enabled Semantic and Goal-Oriented Communications for the Internet of Things

Dejan Vukobratovic
Professor
Faculty of Technical Sciences
University of Novi Sad