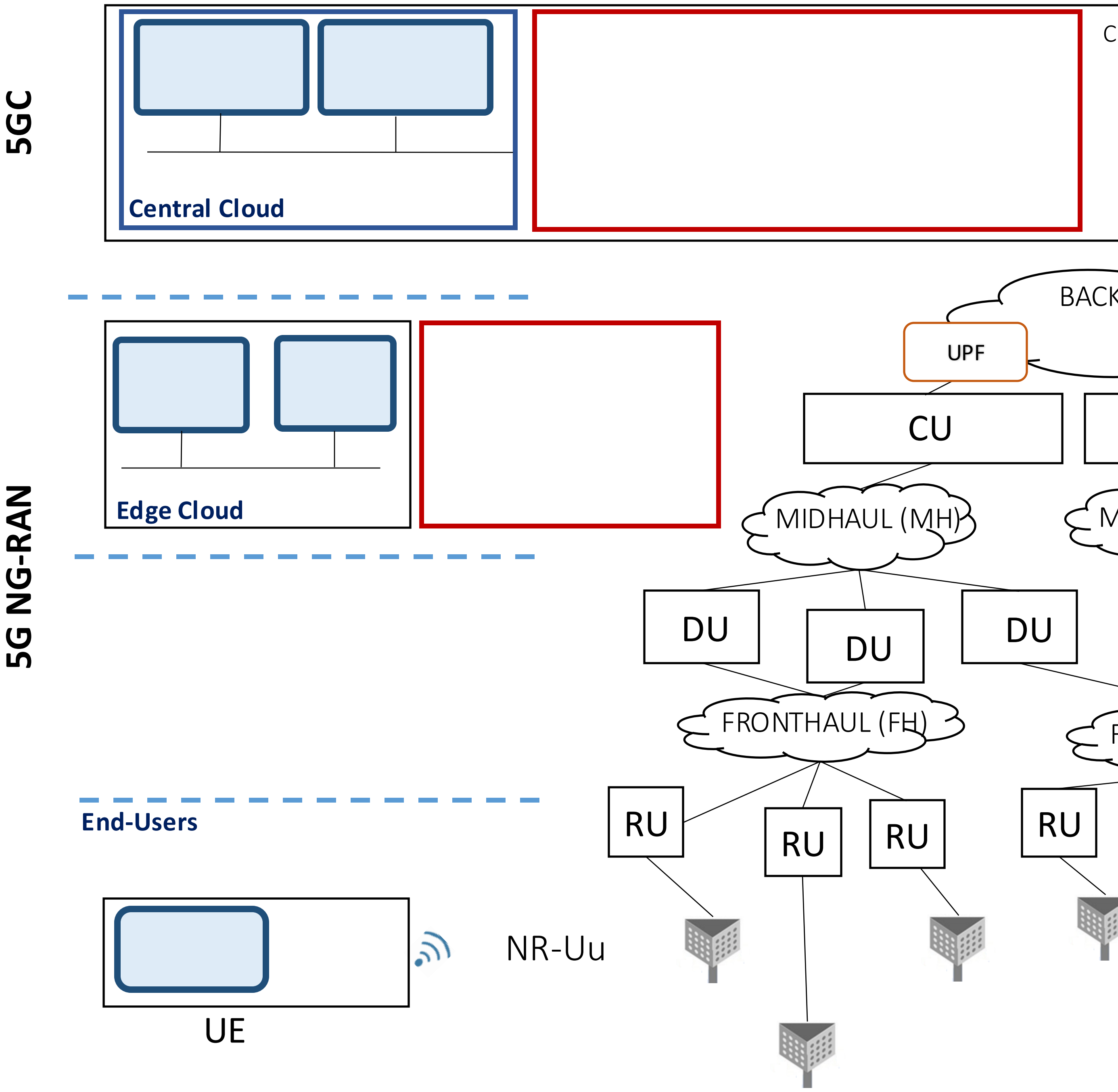
DEPARTMAN ZA ENERGETIKU, ELEKTRONIKU I TELE

Integrating Intelligence in 5G: How ML/AI Is Shaping 5G?

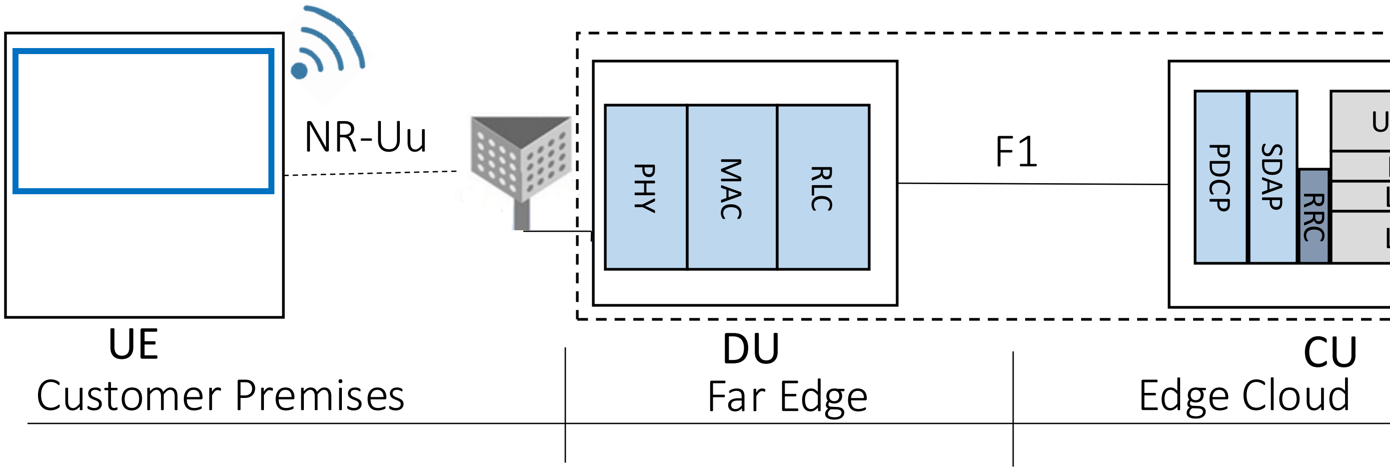
Dejan Vukobratovic
Professor
Faculty of Technical Sciences
University of Novi Sad

Univerzitet u Novom Sadu
Fakultet tehničkih nauka
DEPARTMAN ZA ENERGETIKU,
ELEKTRONIKU I TELEKOMUNIKACIJE



5GC
Central Cloud
5G NG-RAN
Edge Cloud
UPF
CU
MIDHAUL (MH)
DU
DU
DU
FRONTHAUL (FH)
End-Users
RU
RU
RU
RU
UE
NR-Uu

5G NG-RAN


NR-Uu
PHY
MAC
RLC
F1
PDCP
SDAP
RRC
UE
Customer Premises
DU
Far Edge
CU
Edge Cloud

Outline of the talk

- AI/ML in the Core Network
- AI/ML in the RAN
- AI/ML for the PHY
- AI/ML at the Application Layer

Outline of the talk

- **AI/ML in the Core Network**
- AI/ML in the RAN
- AI/ML for the PHY
- AI/ML at the Application Layer

5G Core (5GC) Network

Cloud-Native Service-Based Archite

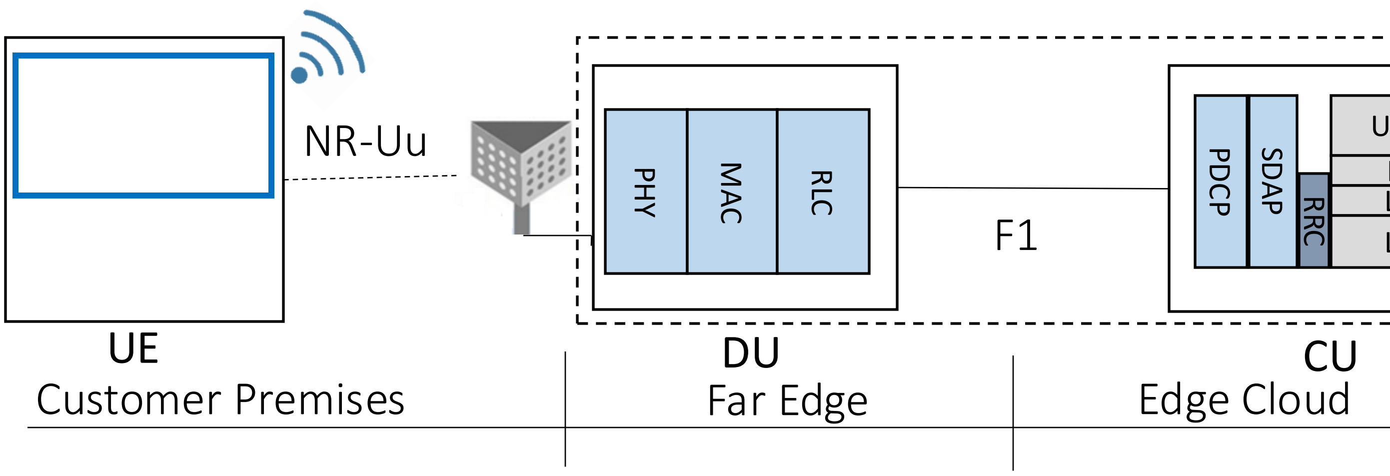
- Collection of virtual Network Functions
- Sofware-based implementation in virtua
- NFs offers one or more services via thei

Support for Network Slicing

- Support for services with different requ
- Creating multiple virtual networks on a
- 5G slice provides complete network fun

3GPP **TS 23.501**, “System architecture for the 5G System (

5G NG-RAN



Rommer, S., Mulligan, C., Hedman, P., Olsson, M., Frid, L., and Sultana,

AI/ML in the 5G Core Netw

ML/AI Integration in 5GC

- Initial Steps started in Rel. 15 in 5GC
  - **NWDAF** – Network Data Analytics Functior
- **Goal:** Enable automated data collection

Evolution of NWDAF

- Initial function (Rel. 15) to provide netw
- Provide data collection and analytics exp
- UE application data collection (Rel. 17)

3GPP **TS 29.520**, “5G System; Network Data Analytics Services; S

Network Data Analytics Fu

Main Functions

- Data Collection Interface for Network Node
- Predefined Data Analytics Functions
- Data and Analytics Exposure Interface for A

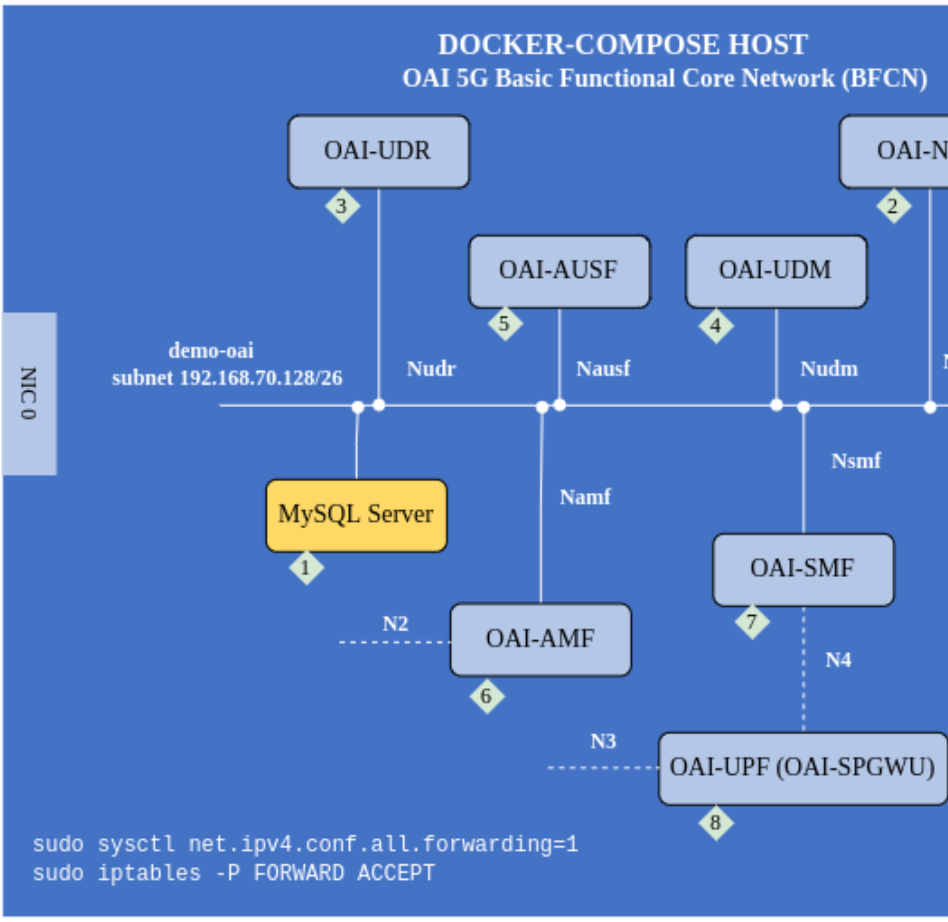
NWDAF Services

- **Analytics Subscription/Analytics Information,** Predictive QoE, Slice SLA Assurance
- **ML Model Information/Provision** provides inf retrieval
- **Distributed Implementation** – Central NWDAF NWDAF (low-latency use cases)

Chouman, A., Manias, D.M. and Shami, A., „Towards supporting i implementation and initial analysis," *IWCMC 2022,* pp. 324-329,

5G Core (5GC) Open Sourc

OpenAir Interface (OAI) 5GC



https://gitlab.eurecom.fr/oai/cn5g/oai-cn5g-fed/-/blob/master/dc

NWDAF support: https://gitlab.eurecom.fr/oai/cn5g/oai-cn5g-nwc

ML/AI in 5GC: A Sample o

Scaling and Orchestration of 5GC Network

- Sheoran, A., Fahmy, S., Cao, L., & Sharma, P., *AI-Drive* *Internet Computing, 25(2), 18–25, 2021.*
- Atalay, T.O., Stojadinovic, D., Stavrou, A. and Wang, H Testbed: A Resource Consumption Study. *IEEE Wirele* *Conference (WCNC)*, pp. 2649-2654, 2022.

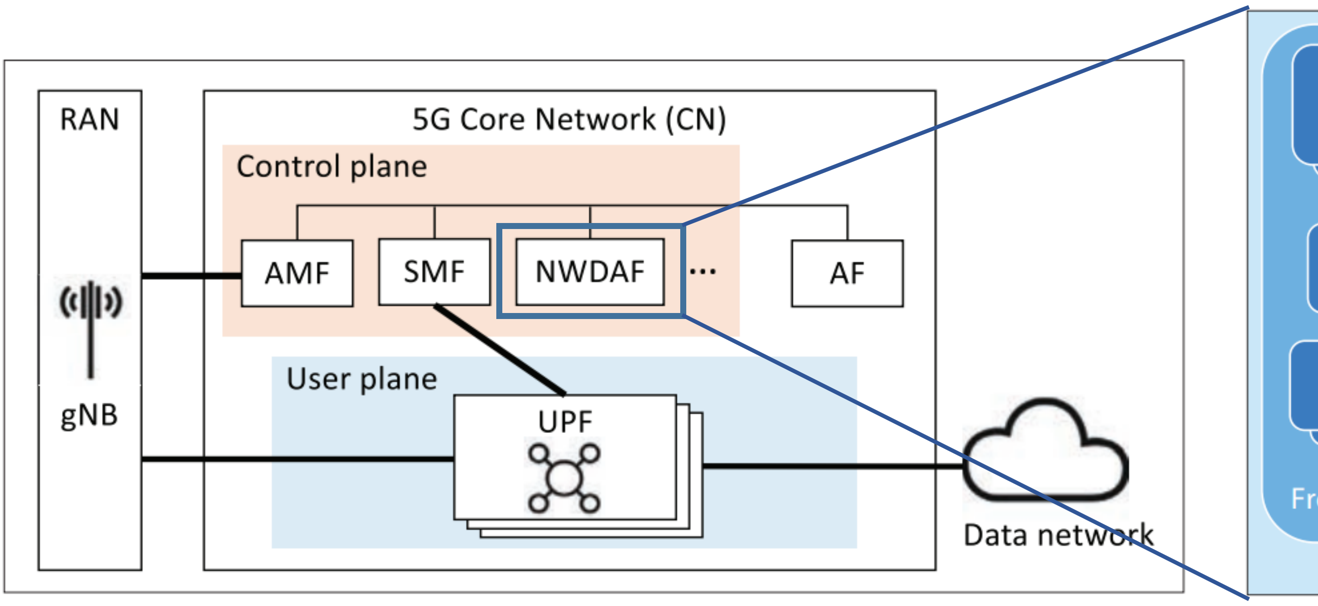
NWDAF (distributed) implementation study

- Lee, S., Lee, J., Kim, T., Jung, D., Cha, I., Cha, D., Ko, H. Implementation of Network Data Analytics Function i
- Hossain, M.A., Hossain, A.R., Liu, W., Ansari, N., Kiani, collaborative learning approach in 5G+ core networks

NWDAF ML/AI-based Functions

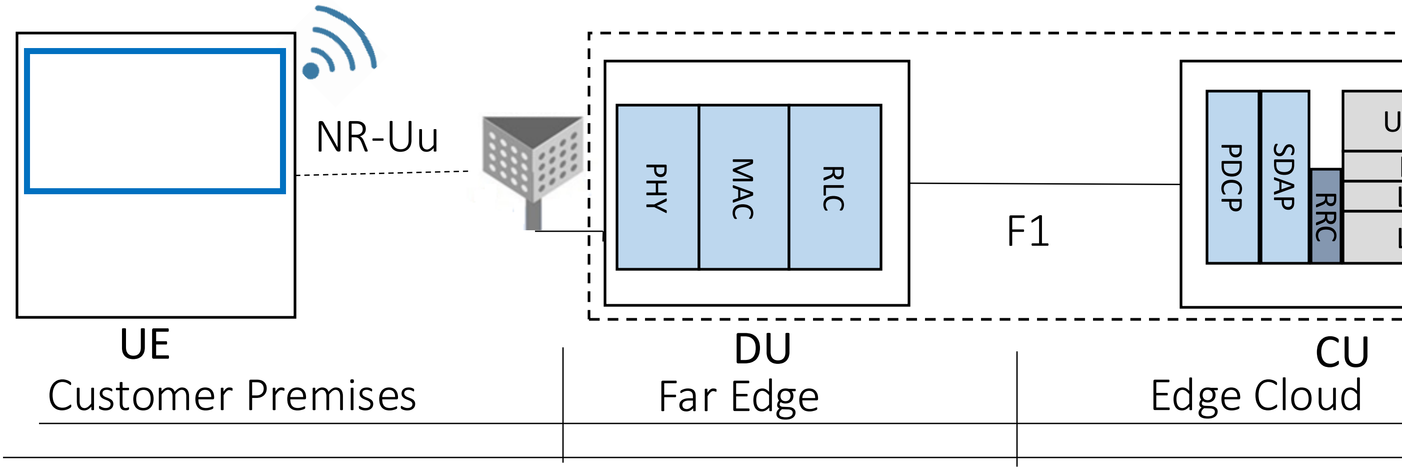
- Murudkar, C.V., Chen, K.C. and Gitlin, R.D., Network A Network Operator's Perspective. *IEEE Communicatior*
- Jeong, J., Roeland, D., Derehag, J., Johansson, Å.A., U Eriksson, G., Mobility prediction for 5G core networks *Magazine*, *5*(1), pp. 56-61.

Example: Mobility Predict

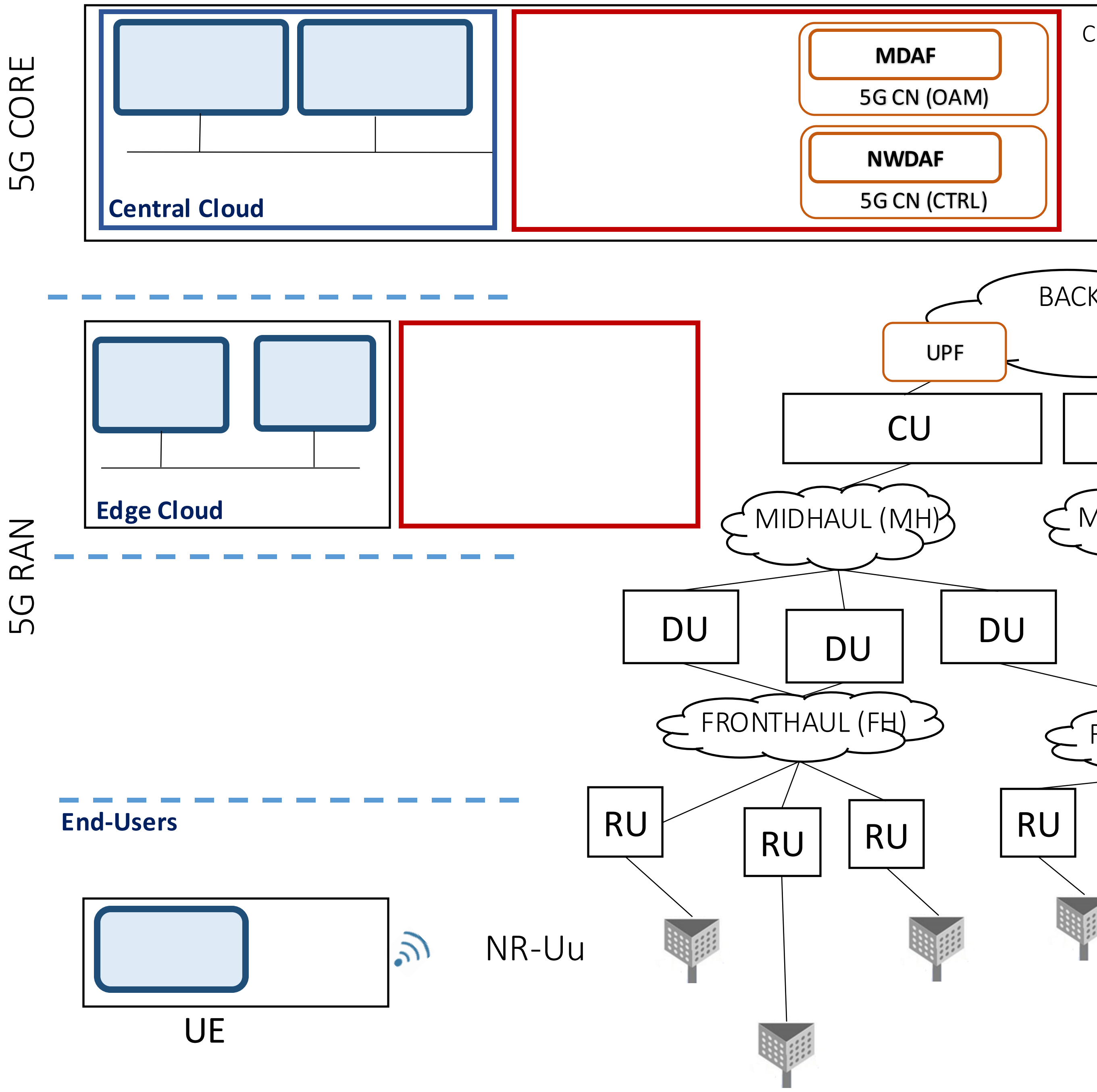


- **Example service:** mobility prediction service
- **Data required**: mobility events streamed by API
- **Life-cycle management**: Handles aperiodicity learning delay, LCM detects concept drift anc

Jeong, J., Roeland, D., Derehag, J., Johansson, Å.A., Umaashankar, prediction for 5G core networks. *IEEE Communications Standards*



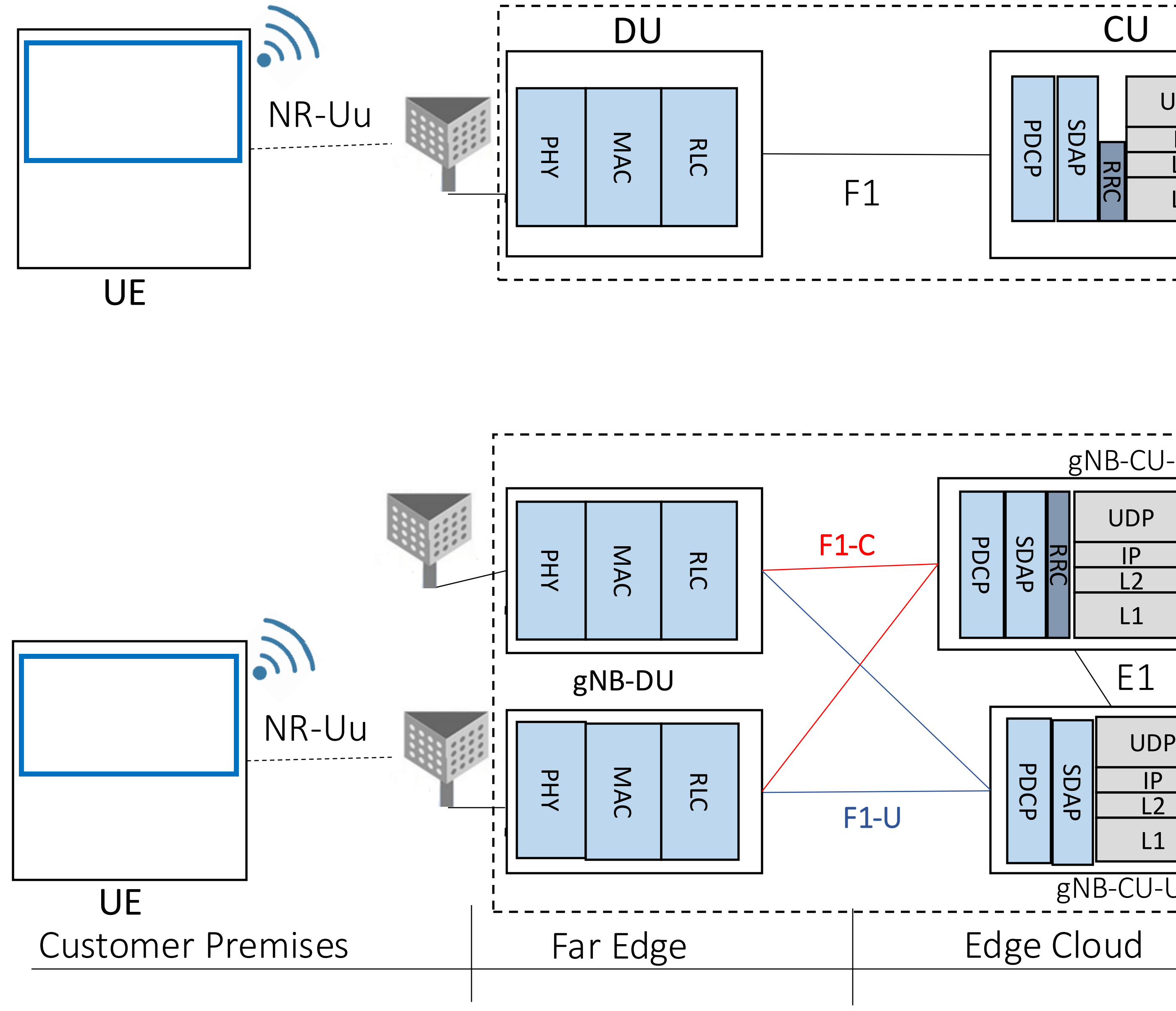
Ferrús, R., Sallent, O. and Perez-Romero, J., Data analytics architect
management in 5G radio access networks. *IEEE Communications M*



5G CORE
MDAF
5G CN (OAM)
NWDAF
5G CN (CTRL)
Central Cloud
5G RAN
UPF
CU
Edge Cloud
MIDHAUL (MH)
DU
DU
DU
FRONTHAUL (FH)
End-Users
RU
RU
RU
RU
NR-Uu
UE

Outline of the talk

- AI/ML in the Core Network
- **AI/ML in the RAN**
- AI/ML for the PHY
- AI/ML at the Application Layer



3GPP **TS 38.401**, “NG-RAN Architecture Description,” V18.

3GPP ML/AI study for 5G

Initial Technical Study Item

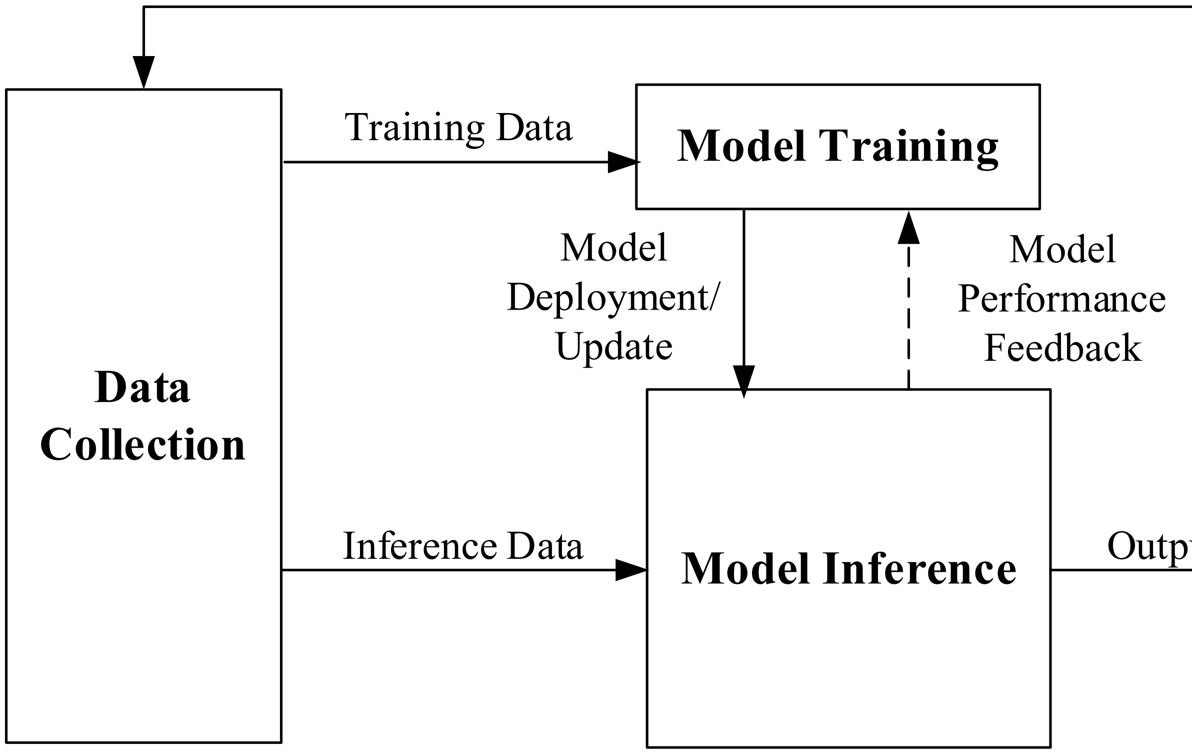
- 3GPP **TR 37.817**, “Study on enhanceme
  EN-DC,” V17.0.0, April 2022.

Main Outcomes

- Functional Framework for RAN Intellige
- AI-enabled RAN for three use cases:
  - **Network Energy Saving** via traffic offloadi
  - **Load Balancing** via prediction-based acros
  - **Mobility Optimization** via UE cell associati
- Led to Approval of Rel. 18 Study on AI/

https://www.3gpp.org/technologies/ai-ml-nr

3GPP ML/AI study for 5G I



Functional Framework for RAN

3GPP **TR 37.817**, "Study on enhancement for data collecti

Rel. 18 Work on ML/AI for

Follow Up Technical Study Item

3GPP **TR 38.843**, "Study on artificial intell
for NR air interface," V18.0.0., January 20

Main Targets

- General Framework for Enhancing Air In
- Main topics:
  - defining stages of AI/ML algorithms (mode
  - UE-gNB collaboration levels
  - required data sets for model training, valid
  - model life-cycle management (LCM)

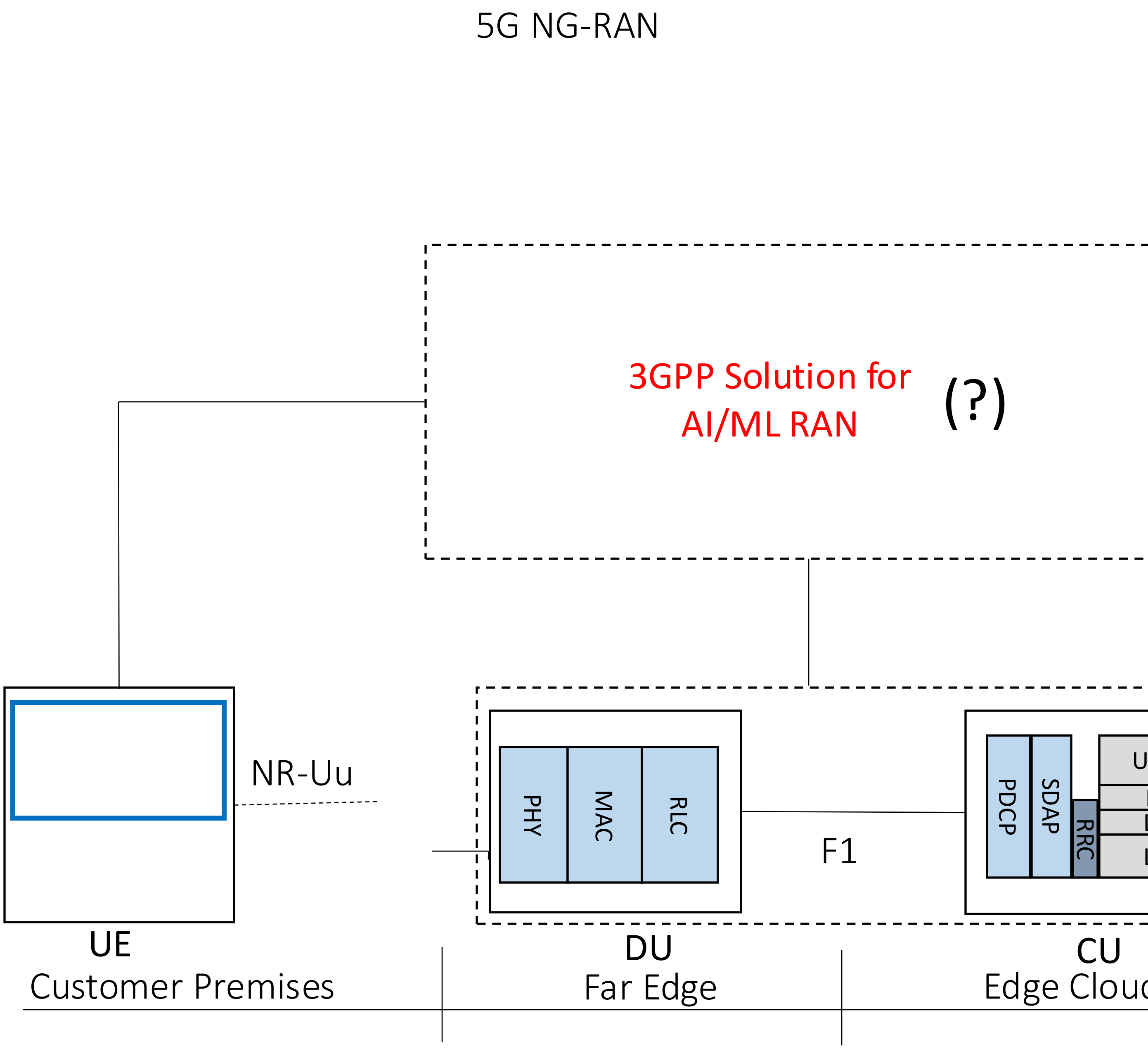
3GPP **TR 38.843**, "Study on artificial intelligence (AI)/machine learn

Rel. 18 Work on ML/AI for

Three use cases

- **CSI Feedback:** Use ML/AI to reduce CSI
domain CSI compression), improve feed
prediction (e.g., time domain CSI predic
- **Beam Management:** Use ML/AI to redu
and latency, and improve beam selectio
spatial-domain and time-domain downl
- **Positioning:** Improve positioning accura
heavy NLOS conditions. Use either direc
fingerprinting) or AI/ML assisted approa

3GPP **TR 38.843**, "Study on artificial intelligence (AI)/machine learn



Follow the development of 3GPP **TR 38.843**, “Study on artificial intelligence (AI)/ma

O-RAN ML/AI study for 5G

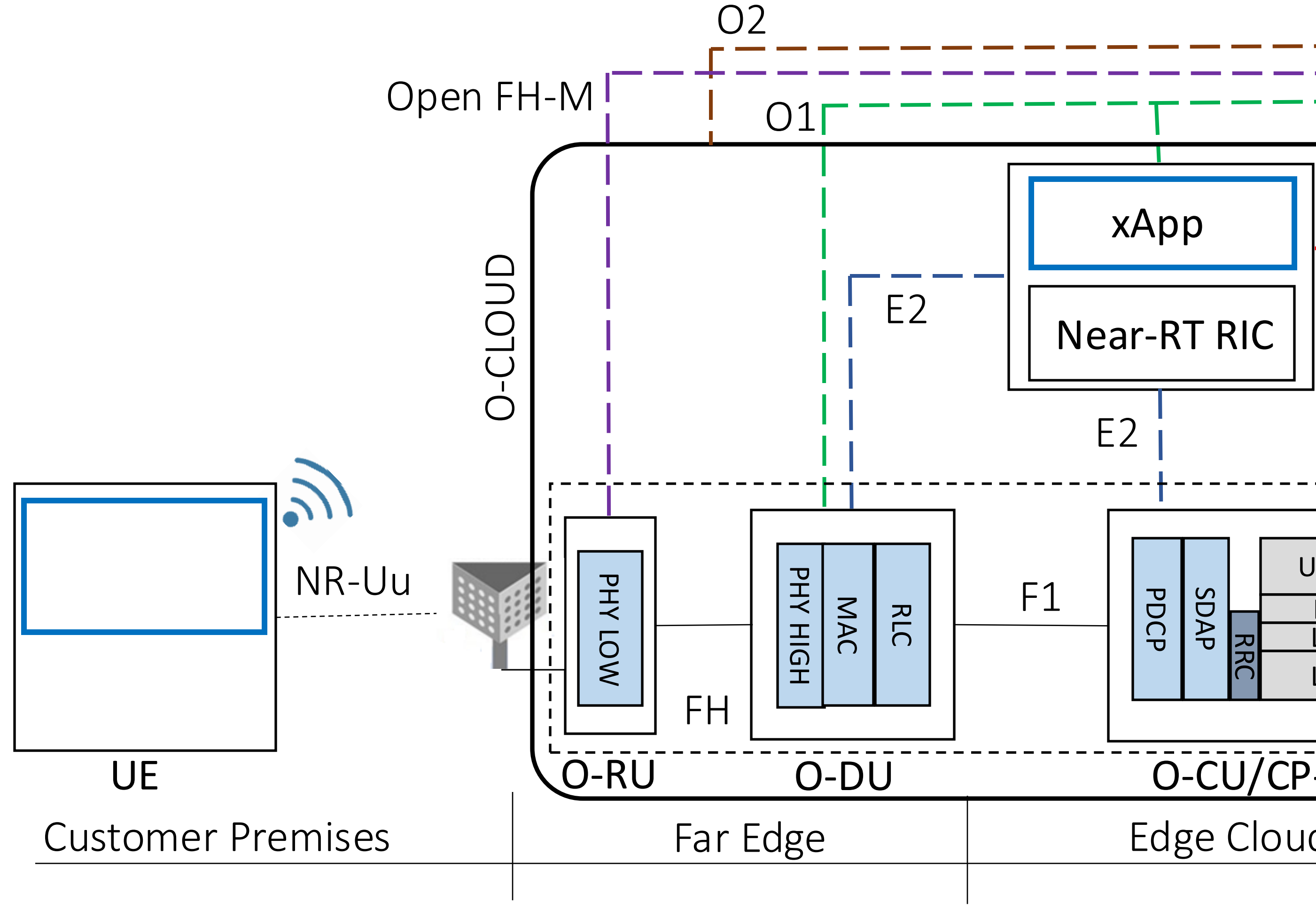
O-RAN Alliance

- https://specifications.o-ran.org/specific
- O-RAN.WG1.OAD-R003-v012.00: „O-RA
Technical Specification, 06/2024

O-RAN Architecture

- Disaggregation of NG-RAN into open, vi
driven architecture
- Augments 3GPP NG-RAN architecture a
with open interfaces (A1, E2, O1, O2)
- Introduces two Radio Interface Controll
(non-RT RIC) and near real-time RIC (ne

5G NG-RAN



- O-RAN.WG1.OAD-R003-v12.00: „O-RAN Architecture De

O-RAN Non-RT RIC

Service Management and Orchestra

- Responsible for RAN domain manageme
- **Non-RT RIC:** main SMO element that int
  NFs): near-RT RIC, O-CU-CP/UP, O-DU, C
- Fault, configuration, accounting, perforr

Non-RT RIC Functions

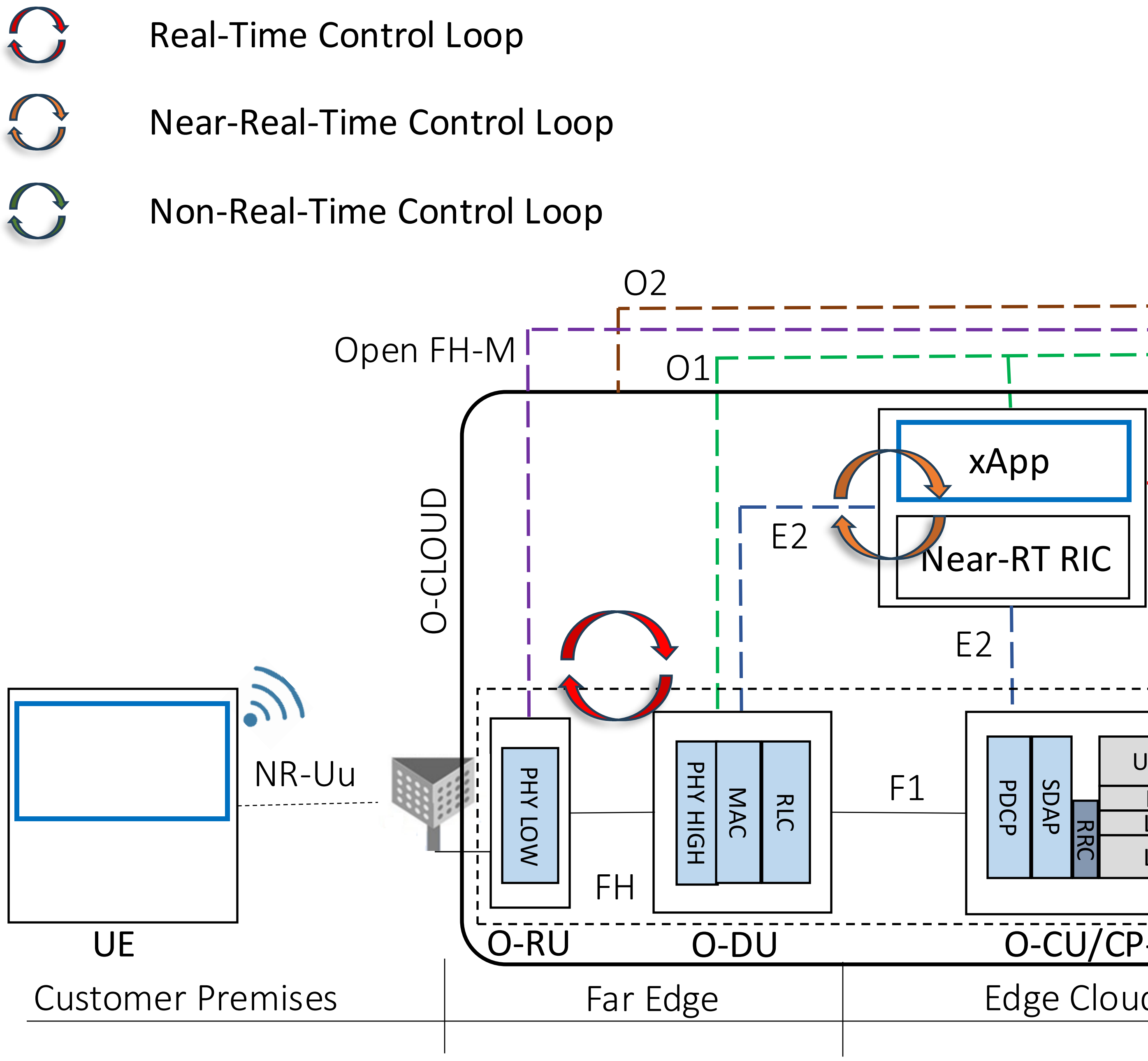
- Controls Near-RT RIC via A1 through: 1)
  model management, 3) enrichment info
- Intelligent Non-RT (> 1s) RAN optimizati
- **rApps** (Non-RT RIC applications): AI/ML
  resource optimization – actions triggere

O-RAN Near-RT RIC

Near-RT RIC Functions

- Enables AI/ML data-driven near-real tim
  optimisation of RAN O-NFs (O-CUs and
- Actions triggered via E2 interface
- **xApps** (Near-RT RIC applications): collec
  nodes and provide back value-added se
- In O-RAN, O-DU is split into O-DU (VNF)
  to 7-2x lower layers split (one of the 3G

M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O
research challenges," IEEE Communications Surveys & Tutorials, 25(2), pp



- O-RAN.WG1.OAD-R003-v12.00: „O-RAN Architecture De

ML/AI in 5G NG-RAN: A Sa

O-RAN xApps and rApps implementation

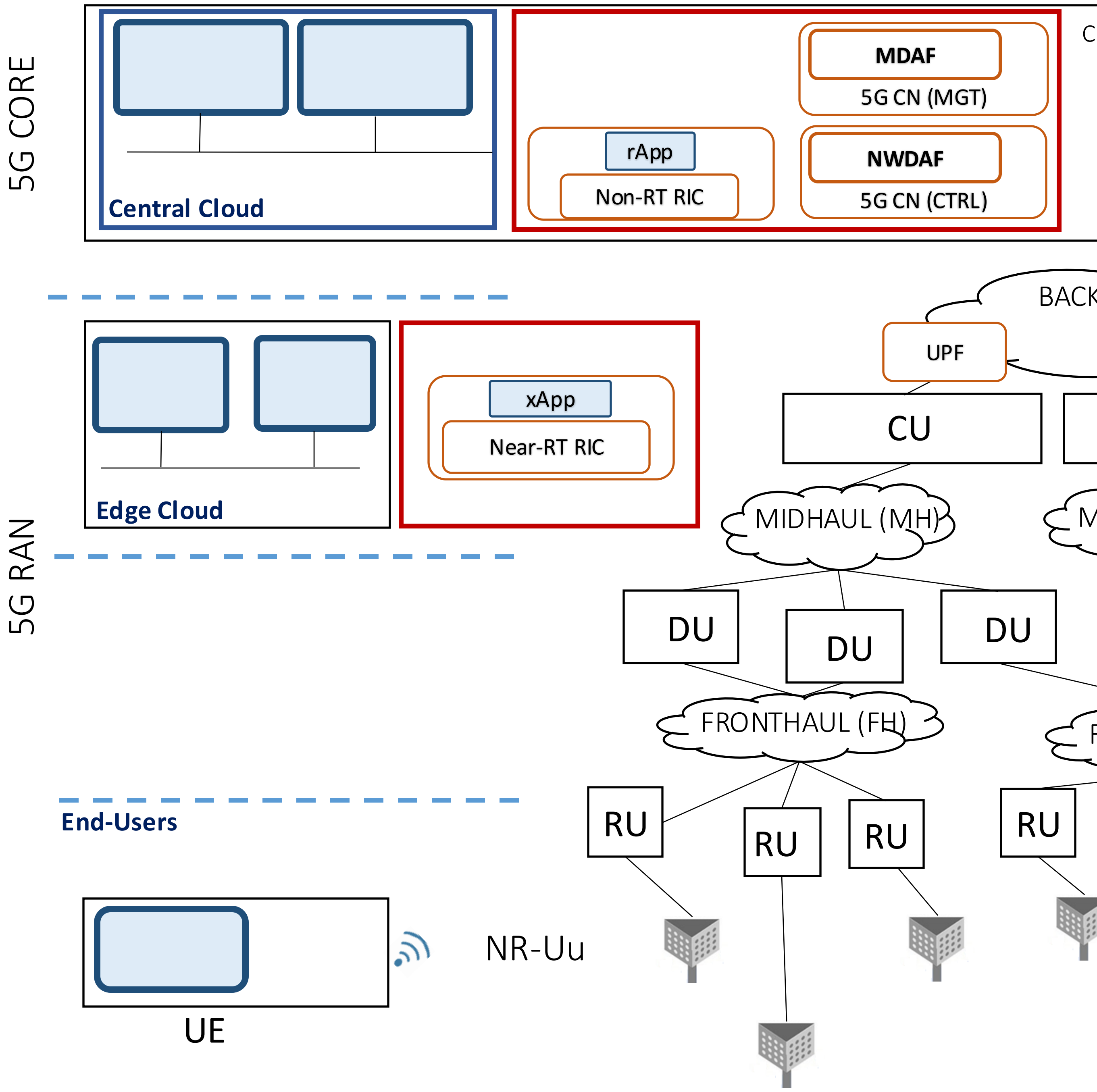
- Kouchaki, M. and Marojevic, V., Actor-Critic Network fo
  Design, Deployment, and Analysis, *IEEE Globecom Wor*
- Vilà, I., Sallent, O. and Pérez-Romero, J., On the Implem
  based Capacity Sharing Algorithm in O-RAN, *IEEE Globe*

O-RAN Open Source Testbed Implementation

- Open AI Cellular (https://github.com/openaicellular) - U
  Marojevic, V., Reed, J.H. and Shah, V.K., Prototyping ne
  with SDRs. *arXiv:2205.13178*.
- OpenRAN Gym (https://openrangym.com/) - M. Polese
  Melodia, "ColO-RAN: Developing Machine Learning-bas
  Control on Programmable Experimental Platforms," IEE
- Both OAI and srsRAN are developing their O-RAN comp

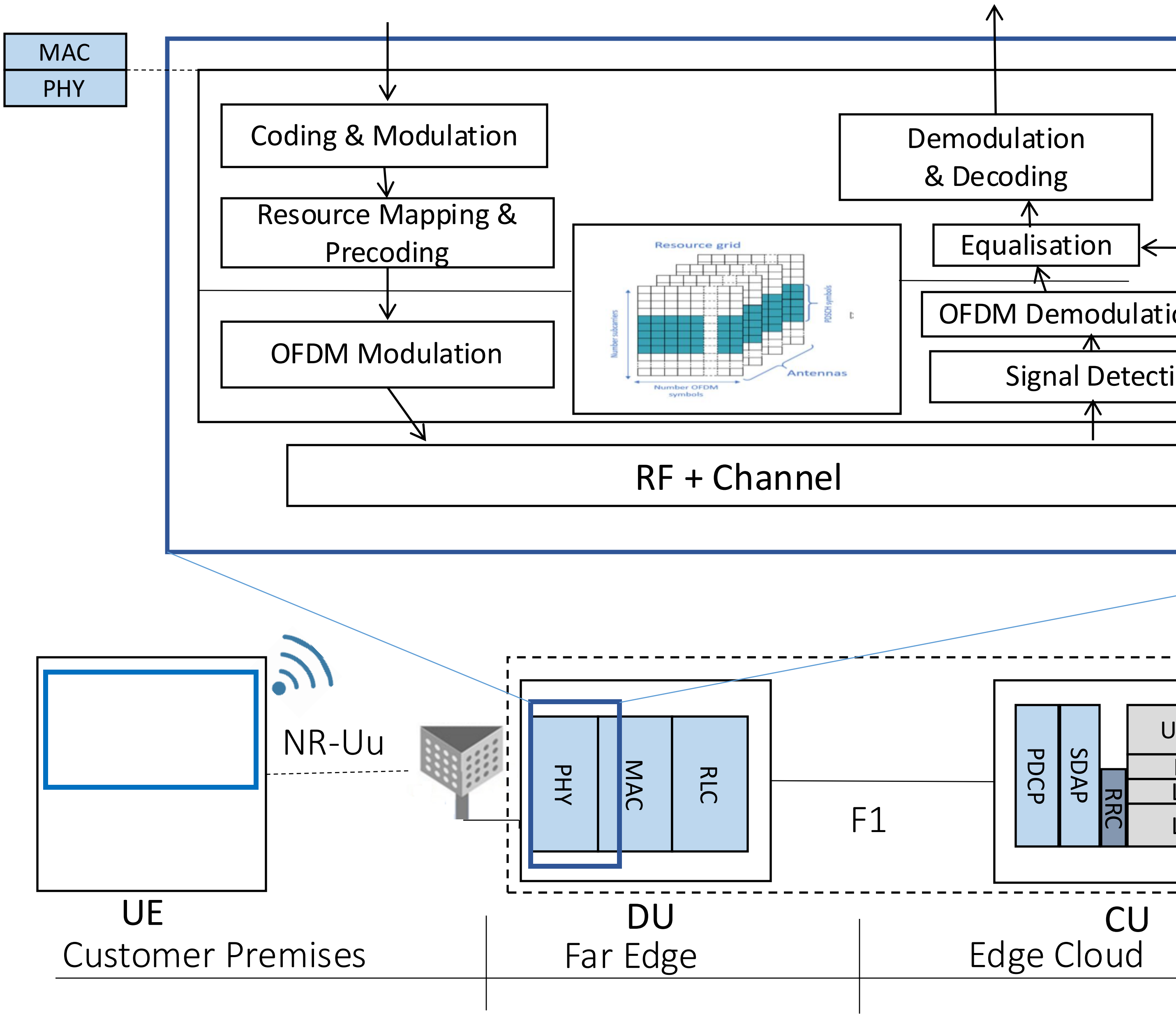
OAI: https://openairinterface.org/oai-5g-ran-project/

srsRAN:https://docs.srsran.com/projects/project/en/late
source/index.html

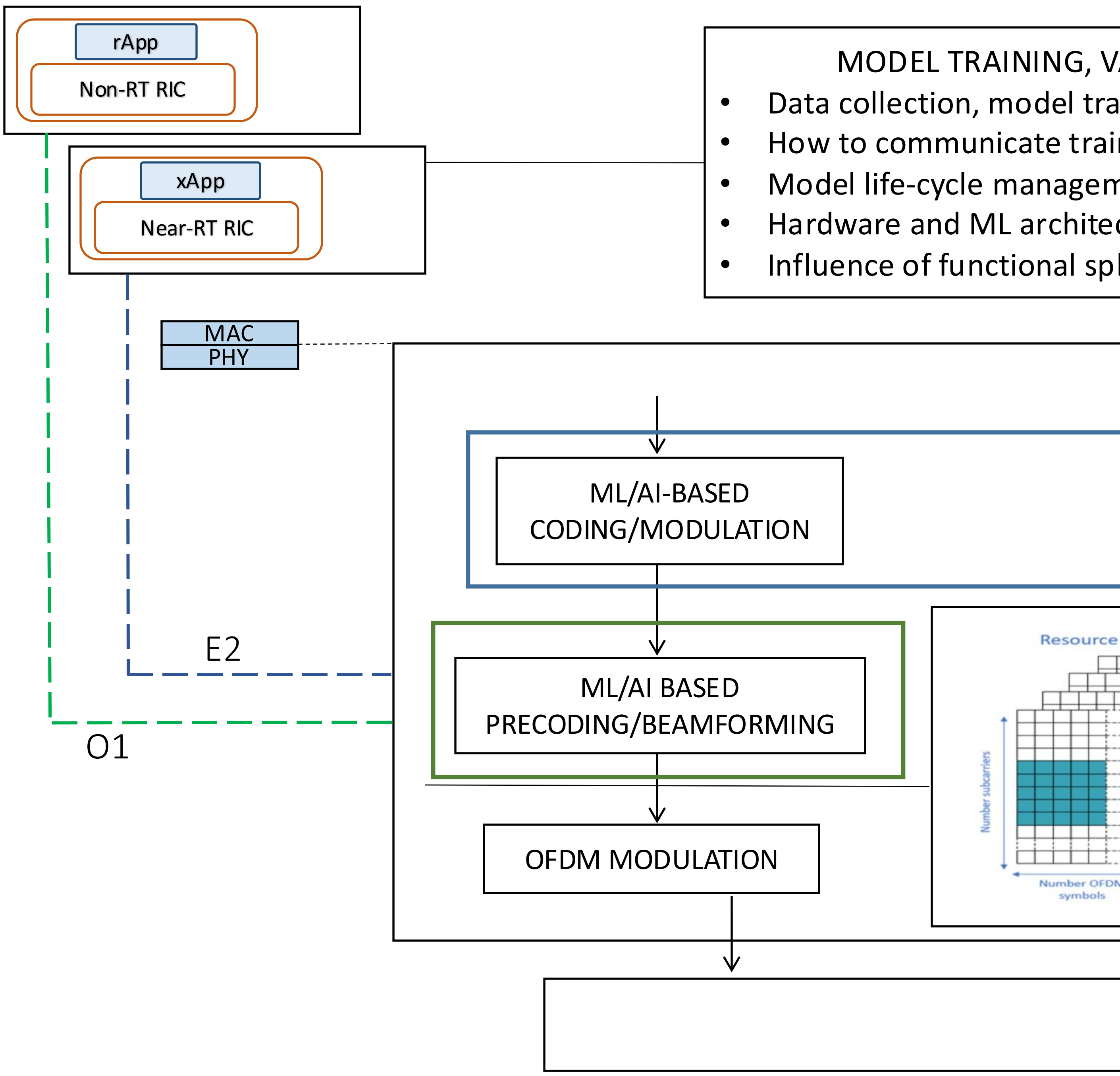

5G CORE
Central Cloud
rApp
Non-RT RIC
MDAF
5G CN (MGT)
NWDAF
5G CN (CTRL)
5G RAN
Edge Cloud
xApp
Near-RT RIC
UPF
CU
MIDHAUL (MH)
DU
DU
DU
FRONTHAUL (FH)
End-Users
RU
RU
RU
RU
UE
NR-Uu

Outline of the talk

- AI/ML in the Core Network
- AI/ML in the RAN
- **AI/ML for the PHY**
- AI/ML at the Application Layer



MAC
PHY
Coding & Modulation
Resource Mapping &
Precoding
OFDM Modulation
Resource grid
Antennas
Demodulation
& Decoding
Equalisation
OFDM Demodulatio
Signal Detecti
RF + Channel
NR-Uu
PHY
MAC
RLC
F1
PDCP
SDAP
RRC
UE
Customer Premises
DU
Far Edge
CU
Edge Cloud



rApp
Non-RT RIC
xApp
Near-RT RIC
MAC
PHY
E2
O1
MODEL TRAINING, V
Data collection, model tra
How to communicate trai
Model life-cycle managem
Hardware and ML archite
Influence of functional spl
ML/AI-BASED
CODING/MODULATION
ML/AI BASED
PRECODING/BEAMFORMING
OFDM MODULATION
Resource
Number subcarriers
Number OFDM
symbols

ML/AI-based Signal Detec

Deep Learning-based Timing Offset and Carrier Fr

- First batch of baseband signal processing steps (b

O’Shea, T., Karra, K., and Clancy, T.C., "Learning approximate
information", *Proc. IEEE 27th Int. Workshop Mach. Learn. Sig*

- Extensions to I/Q imbalance, phase noise and po

Pihlajasalo, J., Korpi, D., Riihonen, T., Talvitie, J., Uusitalo, M.A
OFDM Waveforms Using Deep Learning Receiver, *IEEE SPAW*

- DL-based timing offset and CFO estimation for W

Ninkovic, V., Valka, A., Dumic, D. and Vukobratovic, D., Deep
frequency offset estimation in IEEE 802.11 ah, *IEEE Access*, *9*

ML/AI-based Channel Esti

Deep Learning-based Channel Estimation

- DNN trained to map received pilots and data sym
  symbols (without intermediate channel estimatio

Ye, H., Li, G.Y. and Juang, B.H., Power of deep learning for cha
systems, *IEEE Wireless Communications Letters*, *7*(1), pp. 114

- Channel estimation as CNN-based 2D image reco

Soltani, M., Pourahmadi, V., Mirzaei, A. and Sheikhzadeh, H.,
Communications Letters, *23*(4), pp. 652-655, 2019.

- Extensions to mmWave, massive MIMO, RIS, Thz
- Recommended paper on DL-based Channel Estim

Hu, Q., Gao, F., Zhang, H., Jin, S. and Li, G.Y., Deep learning fo
performance, and comparison. *IEEE Transactions on Wireless*

- Recommended paper on implementation aspects

Haq, S.A.U., Gizzini, A.K., Shrey, S., Darak, S.J., Saurabh,
Augmented Wireless Channel Estimation for Preamble-
Chip, *IEEE Transactions on Very Large Scale Integration*

ML/AI-based Coding/Mod

End-to-End Autoencoder-Based Deep Learning

- End-to-end learning of Communication Systems u

O'Shea, T. and Hoydis, J., An introduction to deep learning fo
Cognitive Communications and Networking, *3*(4), pp. 563-57

- Deep Autoencoder-based PHY design in OFDM sy

Dörner, S., Cammerer, S., Hoydis, J. and Ten Brink, S., Deep le
Journal of Selected Topics in Signal Processing, *12*(1), pp. 132

- Autoencoder-based UEP codes and Rateless code

Ninkovic, V., Vukobratovic, D., Häger, C., Wymeersch, H. and
Error Protection Codes. *IEEE Communications Letters*, *25*(11)

Ninkovic, V., Vukobratovic, D., Häger, C. and Wymeersch, H.,
Decoding Delay and Reliability, *IEEE ICC 2023, Rome, Italy, M*

- Open source libraries for end-to-end learning

https://developer.nvidia.com/sionna
https://developer.nvidia.com/aerial-sdk

ML/AI-based Coding/Deco

Neural-Enhanced Belief Propagation Decoding

Neural-network inspired Belief Propagation decoder

- Nachmani, E., Marciano, E., Lugosch, L., Gross, W.J., Burshtein, D. decoding of linear codes. *IEEE Journal of Selected Topics in Signa*

Principled Approach for Combining Data-Based and Mo

- Shlezinger, N., Whang, J., Eldar, Y.C. and Dimakis, A.G., 2023. Mod IEEE, 111(5), pp.465-499.

Decoding Using Graph-Neural Networks

Decoding LDPC codes using GNNs derived from code fa

- Cammerer, S., Hoydis, J., Aoudia, F.A. and Keller, A., Graph neural *Globecom Workshops (GC Wkshps)* (pp. 486-491), 2022.
- Ninkovic, V., Kundacina, O., Vukobratovic, D., Häger, C., A. Graell i Neural Networks. arXiv preprint arXiv:2408.05170, to appear, IEE

Decoding Using Transformers

Transformer-based approach to decoding LDPC codes

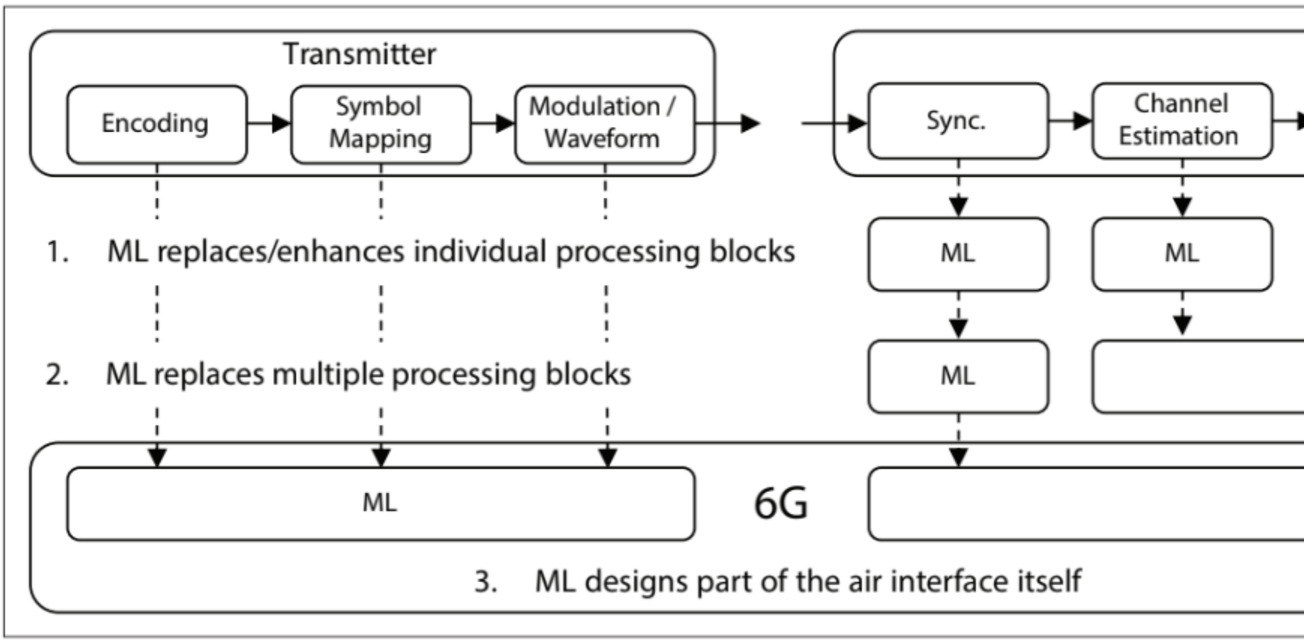
- Choukroun, Y. and Wolf, L., 2024. Learning Linear Block Error Cor

AI-Based PHY Design

Deep Learning for the Physical Layer

Hoydis, J., Aoudia, F.A., Valcarce, A. and Viswanathan, H., Tov

Communications Magazine, *59*(5), pp. 76-81, 2021.



Question: Which PHY blocks to replace with AI/M

ML/AI-based Beamformin

ML for Beam Alignment

Ma, W., Qi, C. and Li, G.Y., Machine learning for beam alignment in
Communications Letters, 9(6), pp. 875-878, 2020.

ML-based Initial Beam Alignment

Sohrabi, F., Chen, Z. and Yu, W., Deep active learning approach to a
alignment. *IEEE Journal on Selected Areas in Communications*, *39*(8

Joint Learning of Beams and Alignments

Heng, Y., Mo, J. and Andrews, J.G., Learning site-specific probing be
Transactions on Wireless Communications, *21*(8), pp. 5785-5800, 2

Overview of AI/ML Beamforming Methods

Al Kassir, H., Zaharis, Z.D., Lazaridis, P.I., Kantartzis, N.V., Yioultsis, T.
and Future Challenges of Deep Learning-Based Beamforming. *IEEE*

Beam Management

Khan, M.Q., Gaber, A., Schulz, P. and Fettweis, G., Machine Learning
Management: A Survey and Open Challenges, *IEEE Access*, *11*, pp.

Outline of the talk

- AI/ML in the Core Network
- AI/ML in the RAN
- AI/ML for the PHY
- **AI/ML at the Application Layer**

AI/ML model transfer in 5

Use Cases and Requirements for Mo

- **AI/ML operation splitting (split learnin**
  sensitive parts in UE, offload computati
- **AI/ML model data distribution**: adaptiv
  needed (efficient unicast/multicast)
- **Distributed/Federated learning over 5G**
  a central entity trains a global model by

TR 22.874 is extended in Rel. 18/19 to TS

- **Example:** maximum allowed downlink end-to-end laten
  required is 1.1 Gbps for image recognition related AI/M

- 3GPP TR 22.874, "Study on traffic characteristics and pe
  transfer," V18.2.0., December 2021.
- 3GPP TS 22.261, "Service requirements for the 5G syste

Outline of the talk

- AI/ML in the Core Network
- AI/ML in the RAN
- AI/ML for the PHY
- AI/ML at the Application Layer
- **Case Study: AI/ML for 5G Smart G**

Smart Grids and 5G Integra

- 5G network is evolving towards idea
  distributed communication and com

- Evolution includes native support fo
  (ML) and **Artificial Intelligence(AI)**-b

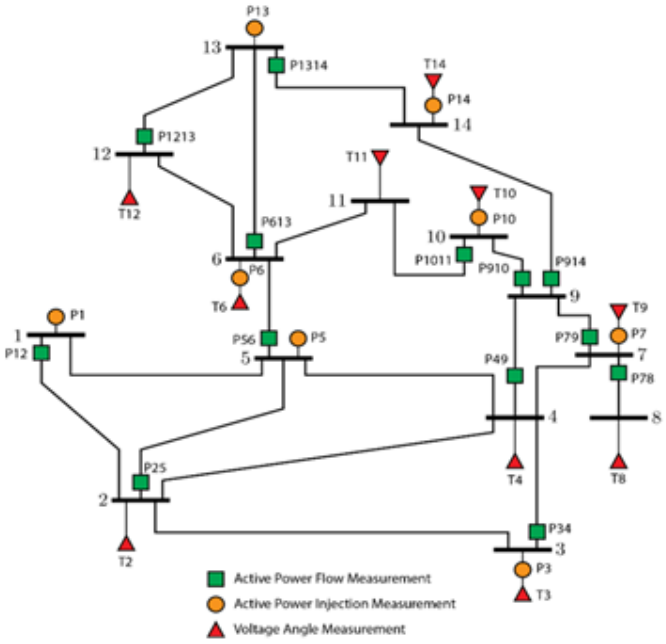
- Integration of Smart Grids and 5G ne
  Generation Partnership Project (3GF
  Release 18/19 of 5G standards

X. Xia, C. Mei, X. Zhou, S. Wang, H. Wang, and Y. Xing, "A review of 3gpp re
IEEE/CIC International Conference on Communications in China, 2021, pp.

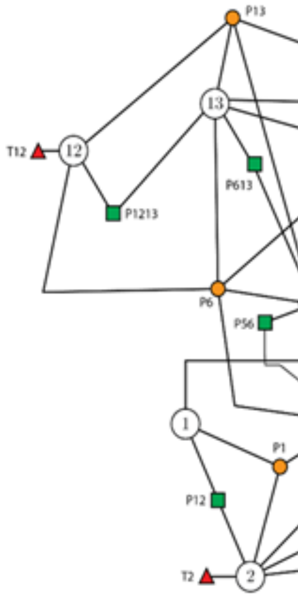
Distributed State Estimation

Model-based SE: We explored Gaussian Bel
solution of distributed large-scale state estin

Power system bus-branch model



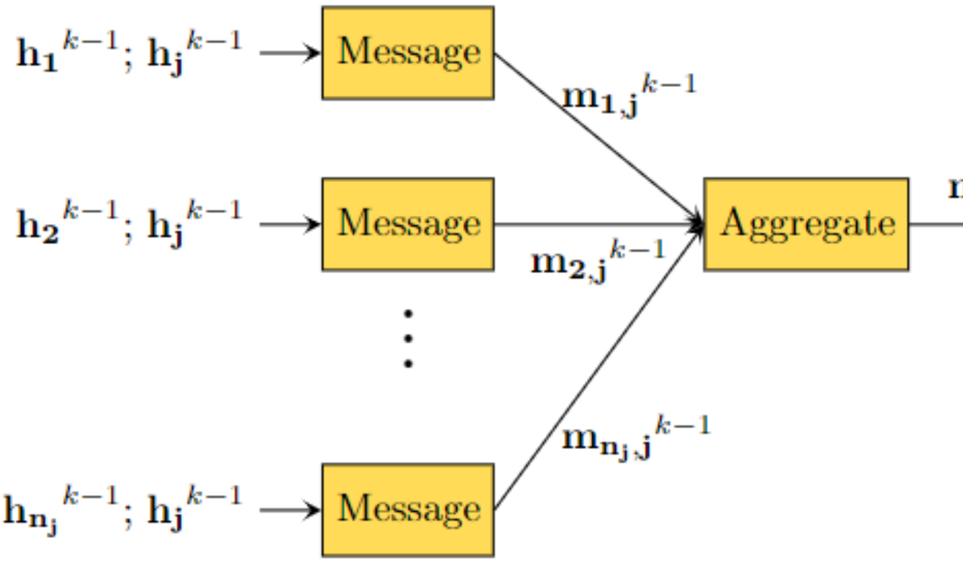
Probabilis



M. Cosovic, D. Vukobratovic, Distributed Gauss–Newton met
Propagation, IEEE Transaction on Power Systems, January 20
M. Cosovic, M. Delalic, D. Raca, D. Vukobratovic, Observabilit
factor graphs, IEEE Transactions on Power Systems, Septemb

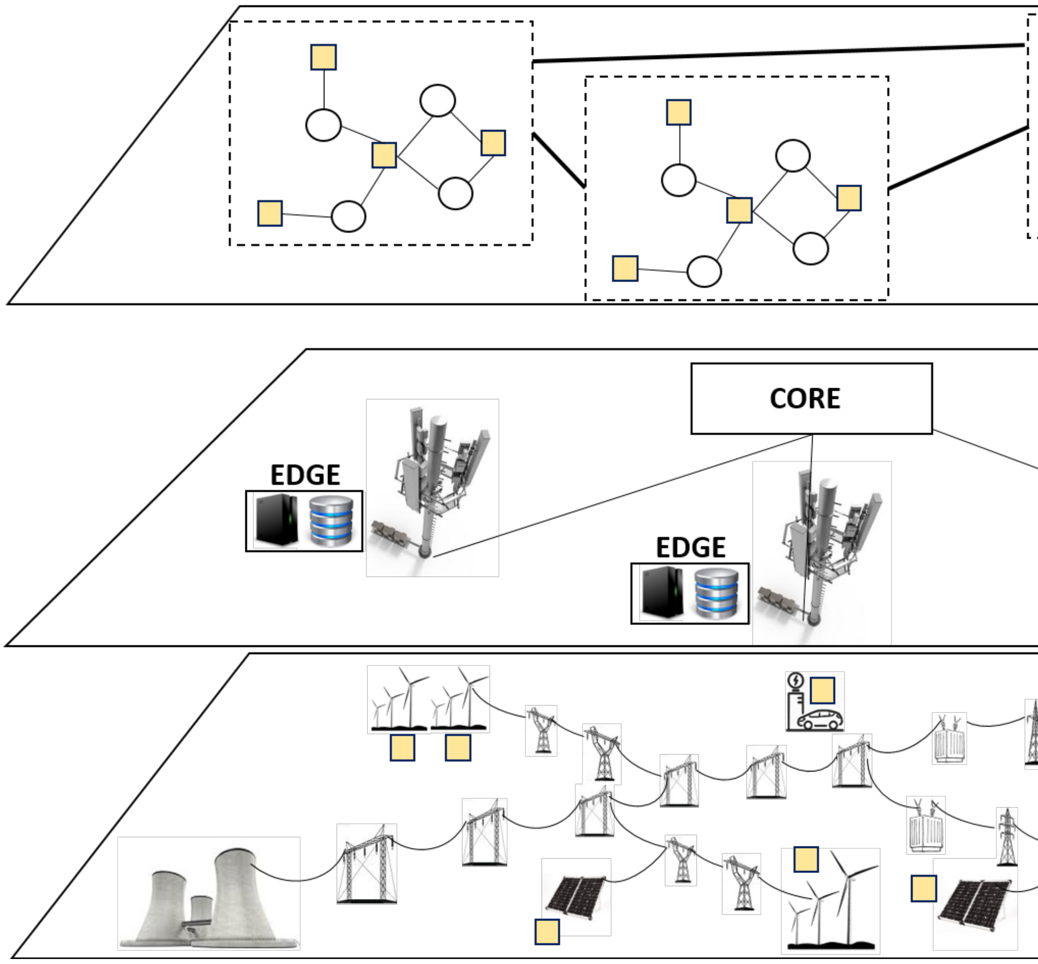
Distributed State Estimatior

Data-based SE: We explored message-passi
for solving distributed large-scale state estin

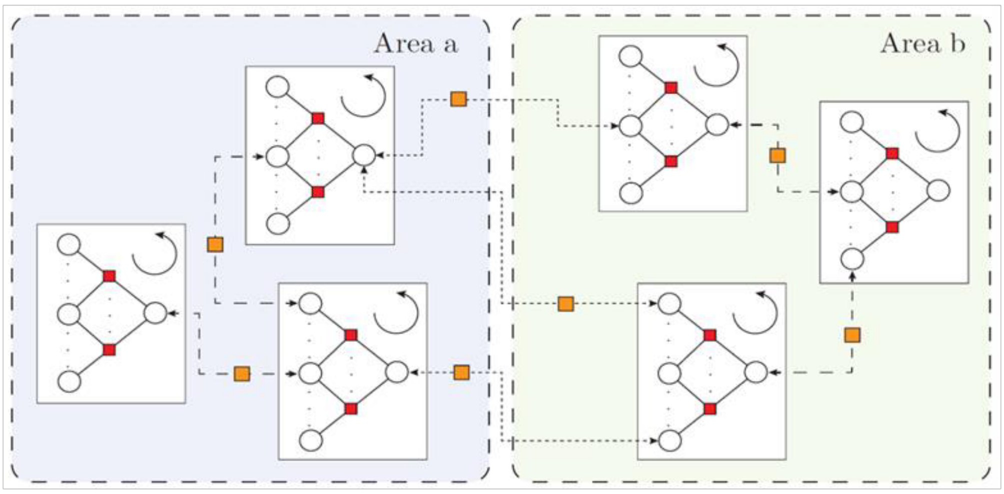


O. Kundacina, M. Cosovic, D. Vukobratovic: "State Estimation in
Networks," Int'l Conference on Probabilistic Methods in Power S
O. Kundacina, M. Cosovic, D. Miskovic, D. Vukobratovic: "Graph
Fast, and Scalable Linear State Estimation with PMUs," Sustainal

Large-Scale Distributed Al

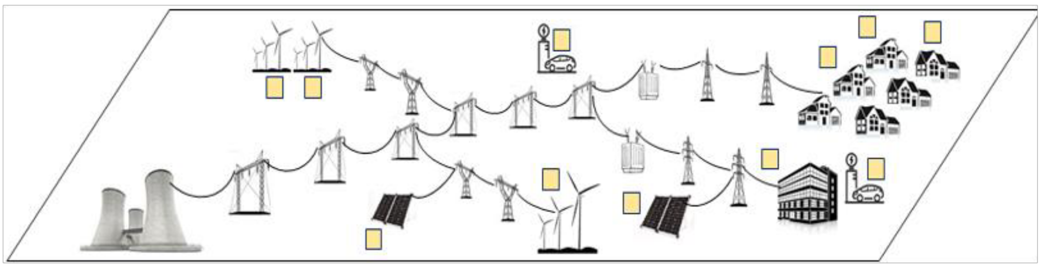


M. Cosovic, A. Tsitsimelis, D. Vukobratovic, J. Matamoros, C
Enabling Distributed State Estimation for Smart Grid, IEEE C



Multi Area BP SE (Stack of 40 RPis 4)

Power flow (Server)



M. Cosovic, D. Miskovic, M. Delalic, D. Raca, D. Vukobratovic, Distribut
Gaussian Belief Propagation, IEEE Internet of Things Journal, 2023.

Distributed SE and 5G Netw

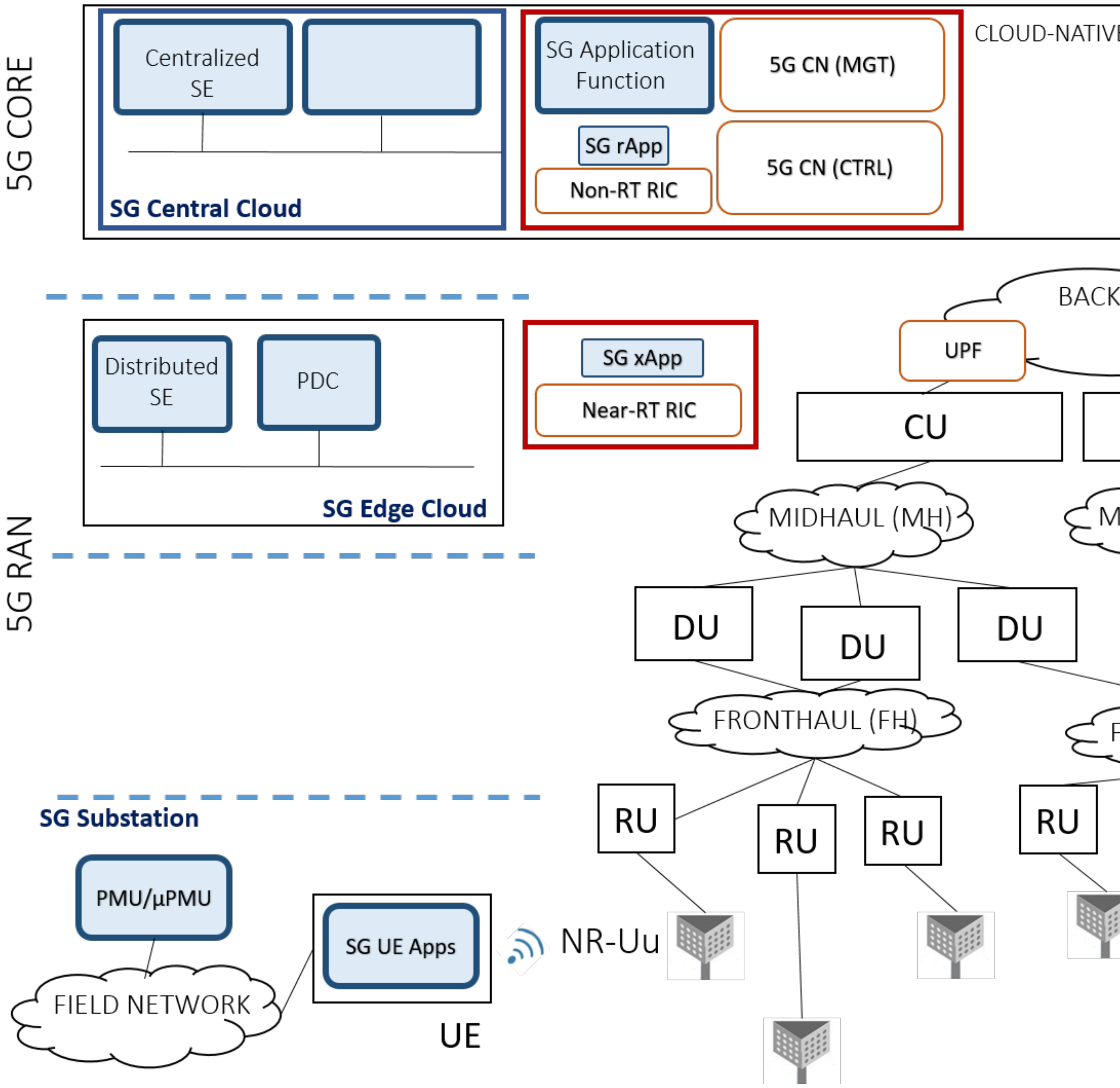
(Distributed) State Estimation – main functi

- **WAMS**: Wide-Area Measurement System
- **PMU**: Phasor Measurement Units

Goal: Explore 5G network as a platform for

- **Data collection:** WAMS over 5G mMTC/UR
- **Data Processing:** Distributed SE as a 5G edge/core function
- **Message-Passing Algorithms:** Model-base (GBP) and Data-based (GNN)

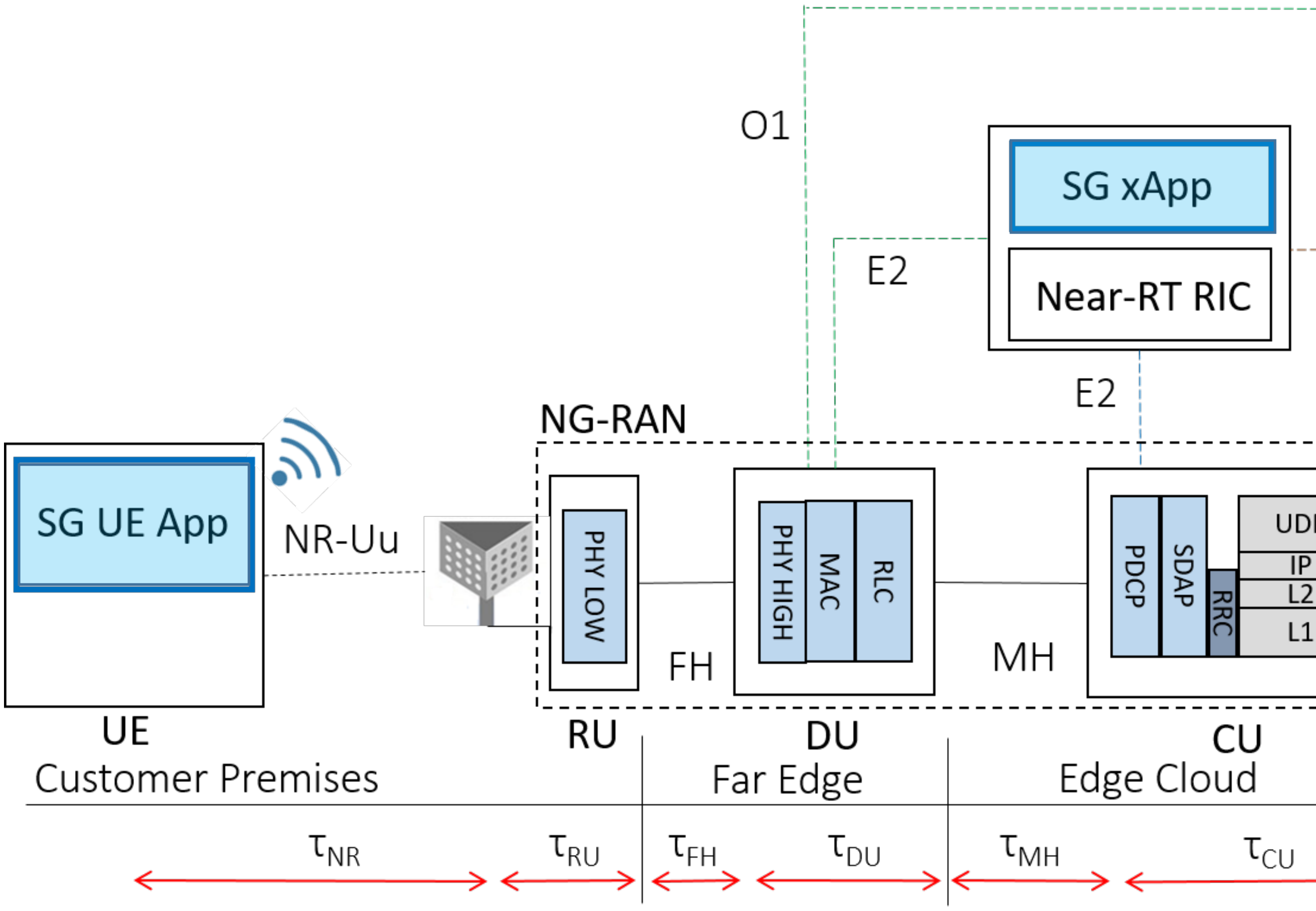
O. Kundacina, M. Forcan, M. Cosovic, D. Raca, M. Dzeferagic, D. Miskov
Distributed State Estimation via AI/ML-Empowered 5G Networks ," **IEE**



5G CORE
Centralized SE
SG Central Cloud
SG Application Function
5G CN (MGT)
SG rApp
Non-RT RIC
5G CN (CTRL)
CLOUD-NATIVE
5G RAN
Distributed SE
PDC
SG Edge Cloud
SG xApp
Near-RT RIC
UPF
CU
MIDHAUL (MH)
DU
DU
DU
FRONTHAUL (FH)
RU
RU
RU
RU
SG Substation
PMU/μPMU
FIELD NETWORK
SG UE Apps
UE
NR-Uu

Smart Grid Apps/Function

- **SG AF**: Third-party CN function (via NEF)
- **O-RAN Apps**: SG xApps and rApps



DEPARTMAN ZA ENERGETIKU, ELEKTRONIKU I TELE

Integrating Intelligence in 5G: How ML/AI Is Shaping 5G?

Dejan Vukobratovic
Professor
Faculty of Technical Sciences
University of Novi Sad