

---

# Building an Open AI-RAN Platform

OCUDU gNB Progress and OAI UE Alignment

Matt Pennybacker

DeepSig Inc. • [deepsig.ai](https://deepsig.ai)

OpenAirInterface Workshop • June 23, 2026



# Agenda

---

- 01 Introduction to OCUDU
- 02 AI-RAN dApps and Interfaces
- 03 How the OAI UE Fits In
- 04 OAI UE & OCUDU Interoperability
- 05 Summary and the Road Ahead

SECTION 01

# Introduction to OCUDU

---

01

# What is OCUDU?

---

- ▶ A **carrier-grade, open source O-RAN gNB** that fuses the **O-CU and O-DU** into one deployable stack
- ▶ An **independent codebase** that evolved from the **srsRAN** project
- ▶ Released under the **BSD 3-Clause** license (OpenMPI variant): permissive redistribution and modification, including **commercial use**
- ▶ **Embraces O-RAN community standards**: F1-C/F1-U, E1, E2, O1, with split-6, split-7.2x, and split-8 support
  - ▷ drops straight into real O-RAN testbeds
  - ▷ integrates easily with real O-RAN carrier deployments
- ▶ Funded by the **FutureG Office** via the National Spectrum Consortium, and organized under the **Linux Foundation**

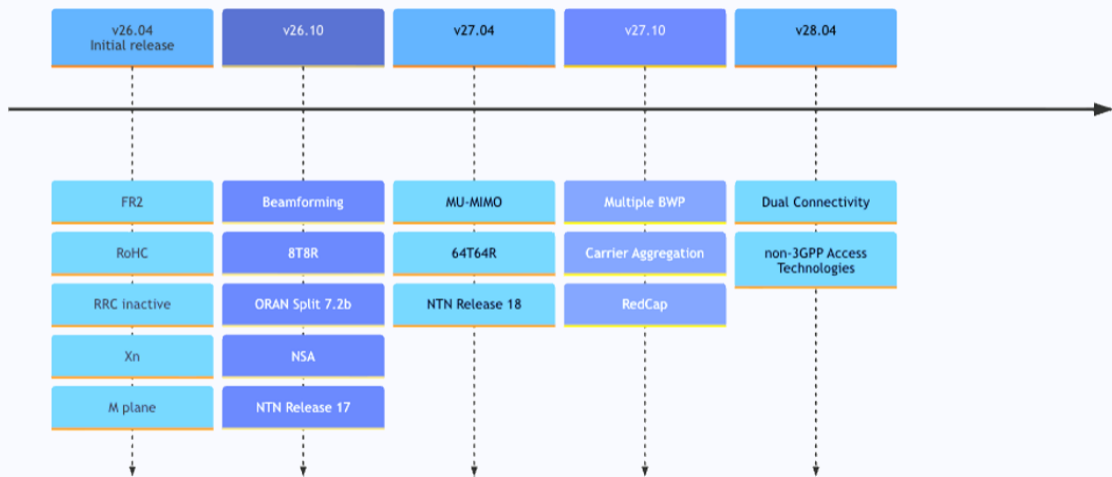
Learn more: [ocudu.org](https://ocudu.org) • [gitlab.com/ocudu](https://gitlab.com/ocudu)

# But this is an OAI workshop?

---

- ▶ This talk sits at the intersection of the **OAI UE** and the **OCUDU gNB**
- ▶ OCUDU is a gNB only; it has no usable 5G NR UE of its own
- ▶ OAI hosts its own Linux Foundation project with similar aims, **Duranta**, and the two differ in emphasis:
  - ▷ **Duranta**: “an open, **research-grade** RAN + UE reference stack”
  - ▷ **OCUDU**: “a permissively licensed, open source 5G and beyond CU/DU project designed for **commercial deployment**, broad industry adoption, and advanced research and development”

# The OCUDU roadmap



SECTION 02

# AI-RAN dApps and Interfaces

---

02

# dApps already exist; the question is latency

---

- ▶ dApps are an established O-RAN *concept*, proposed but not yet an adopted standard; the **E3** dApp interface is already implemented in OpenAirInterface
- ▶ For AI-RAN, the binding constraint is the **dApp ↔ RAN interface**: how a dApp exchanges data with the pipeline sets the achievable latency
- ▶ That budget decides **which AI-RAN tasks a dApp can run**, from per-slot inference down at the PHY to advisory control
- ▶ So the work is not “add dApps” but **classify the use cases** and match each to an interface that fits its budget

# Matching AI-RAN use case time budgets to interfaces

Class	Example task	Interface that fits	Time Budget
A: PHY-inline	neural RX, precoding	resident CPU ABI / GPU kernel	tens of $\mu\text{s}$
B: local control	power control, sched.	DU-owned descriptors, direct ABI	$\sim 100 \mu\text{s}$
C: advisory	telemetry, policy	E3 agent (ZMQ); ASN.1 for interop	ms

- ▶ **E3/ASN.1** is the right tool for Class C, but it cannot carry slot-timed tensors within budget
- ▶ **Class A and B** need resident, lower-latency interfaces, with the right choice **guided by benchmarks**

A forthcoming paper details the use-case classes, the full use-case list, and benchmarks.

# OCUDU as a portable AI-RAN reference platform

---

## Portable reference designs

OCUDU is the home for **dApp and AI-RAN reference designs** that stay portable across DU/CU stacks and hardware, unified by a common control envelope, model formats, and SMO lifecycle.

## DeepSig's role

DeepSig **engineers the open-source interfaces** and provides **AI-RAN reference applications**, such as a neural receiver and sensing, validated on real radios. A GPU-accelerated PHY is a starting point for enabling **GPU-resident processing** in the PHY.

SECTION 03

# How the OAI UE Fits In

---

03

# You can't build an air interface with one end

---

- ▶ A gNB is only half the link, so every PHY/MAC change must **close against a UE**: RACH, HARQ feedback, CSI reporting, the entire UL transmit chain
- ▶ Commercial UEs are **black boxes** we cannot instrument or change: no trustworthy L1 traces, no PUSCH/DMRS/SRS access, no non-3GPP features
- ▶ AI-RAN *and* 6G research need a UE open enough to read, instrument, and **modify on the transmit side**
- ▶ Closed-loop validation means **driving the UE into corner cases** on demand: low SNR, timing-advance edges, malformed signaling
- ▶ That UE is the **OAI nrUE**: open and in-tree with the gNB

# First: a software UE for debugging and AI-RAN dev

---

- ▶ Run the UE **entirely in software**, with no radios: full L1 in simulation, or an abstracted-PHY mode for MAC/RLC work
- ▶ Drive it with reproducible **38.901 TDL/CDL** channels and swept SNR for deterministic BLER curves
- ▶ **Full bilateral visibility of metrics**: no handset can give REs, LLRs, HARQ state, every decode on both ends
- ▶ The data engine for learned PHY: generate **labeled uplink traffic**, train, and evaluate a neural RX in a closed loop
- ▶ CI-friendly: catch F1/RRC/PHY regressions *before* they cost lab and OTA time

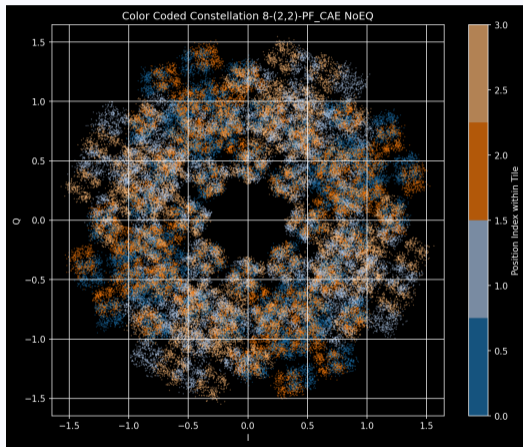
## Then: a flexible OTA UE for ISAC and beyond

---

- ▶ The same codebase now runs **over the air** on USRP-class SDRs, so the sim loop and the OTA loop stay identical
- ▶ An open UE gives full control over **what it transmits**: custom DMRS/SRS, pilot density, waveform, and CSI/measurement reporting
- ▶ Enables **ISAC**: range and Doppler from a 2D-FFT over the OFDM RE grid, with mono- and bistatic geometries on a **time-synchronized cooperative UE**
- ▶ A programmable node to close the gap between **38.901 simulation and the real channel**: impairments, timing, RF non-idealities

# Finally: an integration point for 6G and AI-native PHY

- ▶ **Two-sided learned waveforms:** neural TX (UE) and neural RX (OCUDU) as one **end-to-end autoencoder**, beyond QAM/CP-OFDM
- ▶ 6G is also studying other **two-sided models:** CSI compression/feedback, CSI prediction
- ▶ Co-training the **UE and gNB networks** requires updating the UE TX, which a closed handset forbids
- ▶ OCUDU + OAI UE = a full, **learnable air interface**, open at both ends of the link



A learned constellation for an exotic channel

SECTION 04

# OAI UE & OCUDU Interoperability

---

04

# Standing up OAI UE ↔ OCUDU

---

- ▶ A live, fully **containerized end-to-end software sim**: UE, channel, gNB, and core, each in its own container
- ▶ Closes the link with **no hardware**, and the same setup then moves over the air on a USRP
- ▶ OAI UE fits the need: a **3GPP standards-compliant UE** with strong 5G NR support, including TDD, and consistent with commercial UEs
- ▶ And it is **open source**: a fully accessible codebase, open to read, modify, and instrument on both ends

# Challenges before 2026.w05

---

- ▶ **Sparse documentation:** little existed on driving the OAI UE against a non-OAI gNB, just a 2024 workshop OAI UE + Nokia gNB demo and a few MRs or issues mentioning srsRAN or OCUDU
- ▶ **Virtual-radio incompatibility:** the OAI UE used ZMQ REQ/REP while OCUDU used RFsim, so the options were a ZMQ-to-RFsim shim or native ZMQ support in OAI
- ▶ **Blocking interop issues:** once on a USRP, several blocking and performance-limiting issues surfaced

# Blocking issues and fixes

---

Issue	Cause	Resolution
UE assert: max MIMO layers > 0	DCI 0_1/1_1 vs a 1-port UE	fallback DCI 0_0/1_0, or a 1-port uecap file
UE assert: no CSI on PUSCH	periodic CSI on PUSCH unsupported in the UE	disable CSI-RS
gNB: no DL link adaptation	no CQI/RI feedback, MCS stuck at 8	force a fixed MCS

Issues 2 and 3 were a **catch-22**: DL link adaptation needs CSI-RS, but CSI-RS tripped the UE assert. A fixed MCS held the line until the OAI UE gained **CSI-on-PUSCH support** (MR !3766).

# Significant improvements

---

- ▶ **Feb 2026** (2026.w05): OAI UE can send CSI reports on PUSCH (MR !3766)
- ▶ **May 2026** (2026.w17): native ZMQ radio in OAI (MR !3975)
- ▶ **June 2026** (pending):
  - ▷ TRS-based frequency-offset estimation and correction (MR !3574)
  - ▷ improvements to UCI on PUSCH at the UE (PR !139)

A run of upstream OAI merge requests has steadily improved interop, several driven by what this bring-up surfaced.

## Deep dive: CSI mapped to the wrong symbol

---

- ▶ The gNB reads CSI on PUSCH from the first non-DMRS symbol (symbol 0), but the OAI UE wrote it to the first non-DMRS symbol *after* the first DMRS symbol (symbol 3).
- ▶ The first failed slot stays in the gNB HARQ buffer, and the retransmission carries no CSI but soft-combines with it, yielding low confidence.
- ▶ With 30 kHz SCS, a DDDSU pattern, and a 20 ms CSI period (8 UL slots) on ZMQ with no fading or noise, the result is a constant **25% UL BLER**, and only when CSI is enabled.

### Status

Fixed in PR !139 (commit 4c79163).

SECTION 05

# Summary and the Road Ahead

---

05

# What we are building

---

- ▶ **OCUDU**: an open, carrier-grade O-CU/O-DU gNB, built to **host dApps** and open AI-RAN interfaces
- ▶ **The OAI UE**, aligned and extended, playing three roles:
  - ▶ a software UE for **debugging and AI-RAN dev**
  - ▶ a flexible **OTA UE** for ISAC and live experiments
  - ▶ an **integration point** for 6G and two-sided learned waveforms
- ▶ Together: a **complete, open, learnable air interface**, open at both ends of the link and reproducible by the community
- ▶ DeepSig has formed **two new OCUDU working groups**: hardware acceleration and AI-RAN

Get involved: [gitlab.com/ocudu/work\\_groups](https://gitlab.com/ocudu/work_groups)

# WG-HWACCEL (Hardware Acceleration)

---

## Mission

Incubate OCUDU acceleration across heterogeneous compute: prototype, benchmark, and document before upstream.

## Key objectives

- ▶ Prototype L1 and compute-intensive acceleration paths
- ▶ Incubate the initial NVIDIA GPU / CUDA L1 reference design
- ▶ Repeatable benchmarks and baseline results
- ▶ Reference designs, containers, and quickstarts
- ▶ An open path for additional hardware backends

**Initial release available now.**

# WG-AIRAN (AI-RAN)

---

## Mission

Incubate AI-RAN apps, dApps, interfaces, data paths, benchmarks, and examples on OCUDU, positioning it as the open AI-RAN platform.

## Key objectives

- ▶ Priority AI-RAN app and dApp use-case study
- ▶ Practical OCUDU interfaces aligned with O-RAN concepts
- ▶ Data, telemetry, and evaluation paths for AI algorithms
- ▶ Reproducible benchmarks and KPIs (“AI-RAN-Perf”)
- ▶ Grow the community: researchers, vendors, operators

**Initial release planned for late June / early July.**

---

# Thank you

Matt Pennybacker • pennybacker@deepsig.ai  
Building an Open AI-RAN Platform  
[www.deepsig.ai](http://www.deepsig.ai)

