



The Optical Network Digital Twin: New Frontiers Beyond QoT Estimation

Rocco D'Ingillo^{1,2}

1 PLANET - PHYSICAL LAYER AWARE NETWORKING · DET, POLITECNICO DI TORINO, ITALY

2 SYNPHONYA · STARTUP, POLITECNICO DI TORINO SPIN-OFF

01 The twin as it stands

What an ONDT actually is : the passive inventory layer, the active simulation layer, and the closed loop between them.

02 QoT-aware operations

What GNPY already lets operators do: multi-vendor and multi-domain interoperability, tighter margins, shared spectrum.

03 Sensing

The same fiber that carries your traffic can sense it, extending the twin from signal quality to physical integrity.

04 Language

LLMs grounded in the twin's live topology and telemetry: promising for management, still risky for closed-loop control.

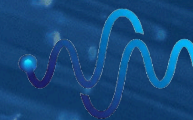
*...and then the two open questions that keep coming back in all four: **open interfaces** between the twin's layers, and **open, shared data** to keep every layer accurate.*

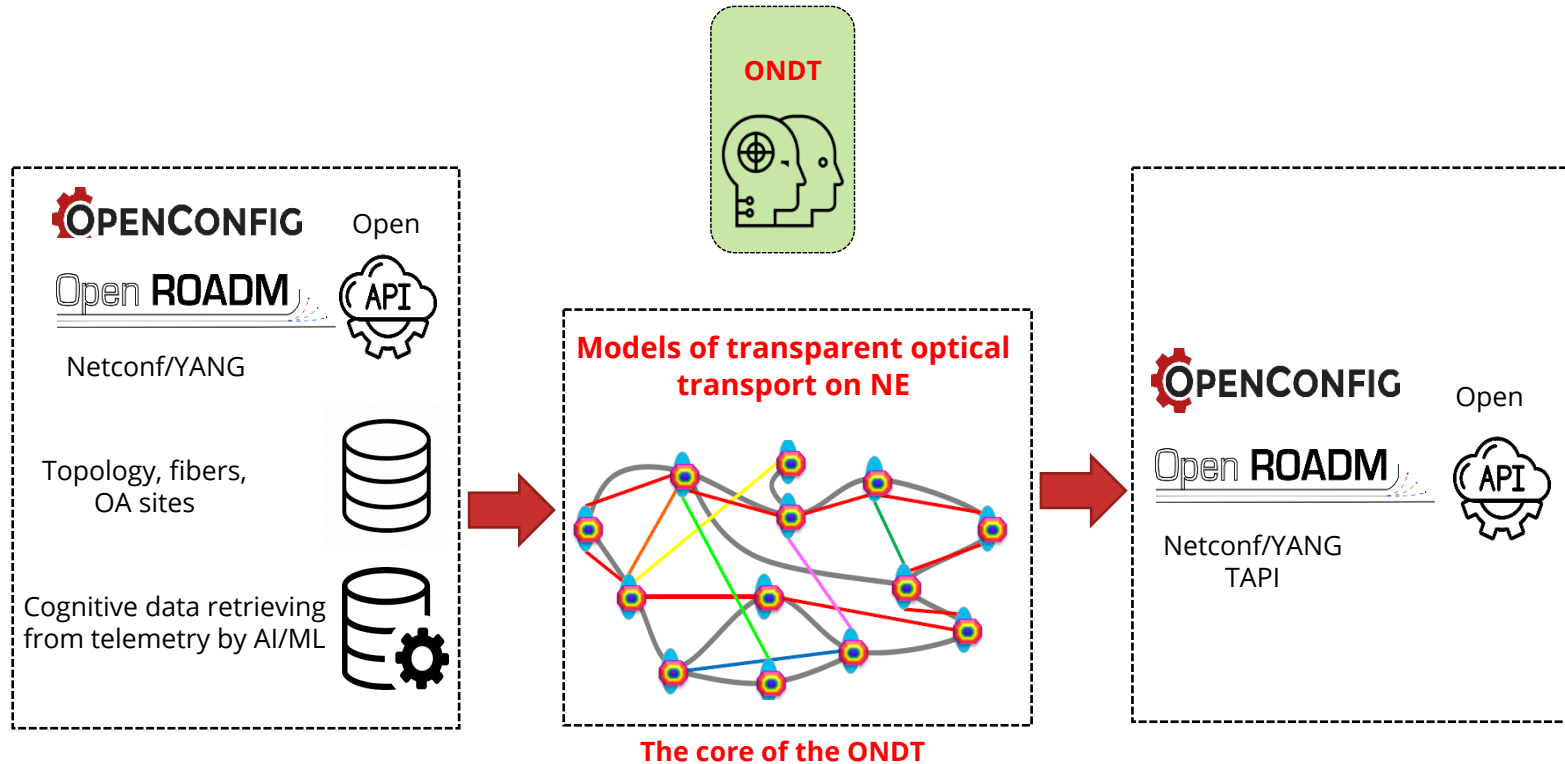


PART 1

THE TWIN AS IT STANDS

PASSIVE LAYER, ACTIVE LAYER,
AND THE LOOP THAT CLOSES BETWEEN THEM





Passive layer: the inventory

What is out there. Topology, fiber types and lengths, span losses, amplifier sites, ROADM configurations, transceiver capability sets. Slow-changing, and mostly wrong in every operator's database, which is the point.

Active layer: the simulation

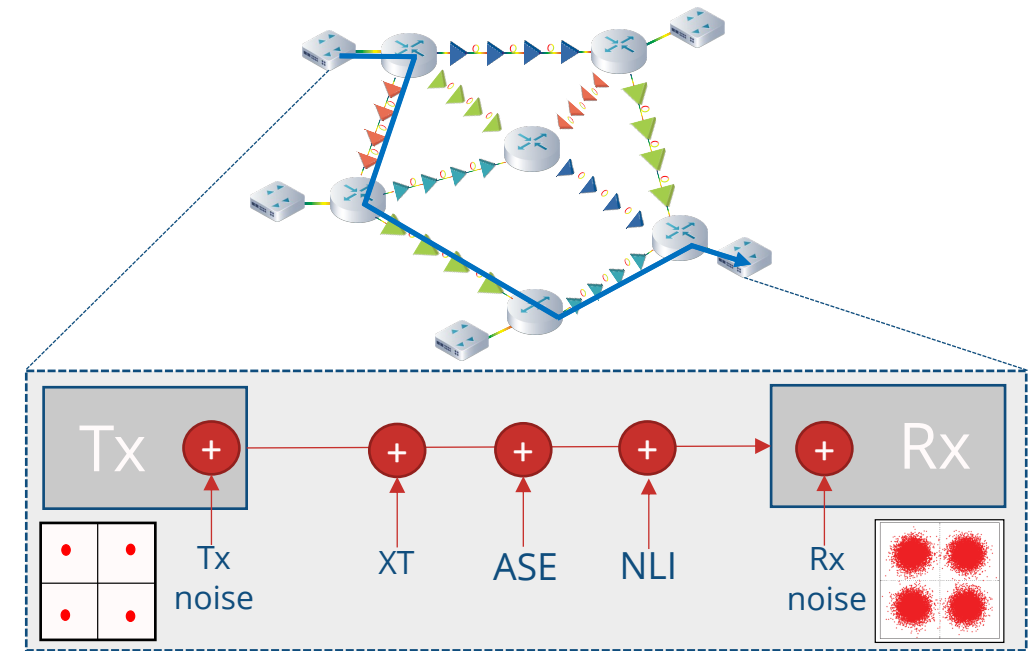
What the light does. Propagation, amplification, filtering, crosstalk, nonlinearity. Given the passive layer plus a spectral load, it returns a QoT for every channel on every route in seconds, not hours.

ONE NUMBER: THE GENERALIZED SNR

- A transparent lightpath is very well approximated as a **dual-polarization AWGN channel**.
- That is a strong statement, and it is the reason any of this is tractable: every impairment the line adds (amplifier noise, nonlinear interference, crosstalk, transceiver noise) can be accumulated into **one scalar**.

$$\text{GSNR} = \text{signal power} / (\text{ASE} + \text{NLI} + \text{XT} + \text{TRX noise})$$

- Each transceiver operating mode has a **GSNR threshold**. Compare the two numbers and you have your answer: the lightpath closes, or it does not.



Why an operator should care about a single number

- It makes the physical layer **comparable**. Two vendors' line systems, two generations of transceiver, two candidate routes. One metric, one comparison.
- It makes the physical layer a **weighted graph**. Once every link has a GSNR contribution, routing and spectrum assignment become an algorithm, not a negotiation.
- It makes margin **explicit**. The gap between predicted GSNR and the threshold is your margin, in dB, per lightpath. Visible, and therefore arguable.

- **GNPy**: the Gaussian Noise model in Python — developed in the **OOPT-PSE** working group of the Telecom Infra Project. Open source, vendor-neutral, Apache-licensed.
- Started life as a design and planning tool. It is now used as a **service inside SDN control**.
- Takes a JSON description of topology, equipment and services; returns GSNR per channel per route. No vendor's proprietary planning tool in the loop.
- This is what makes the twin a **community asset rather than a product**: the model everyone argues about is the same model.

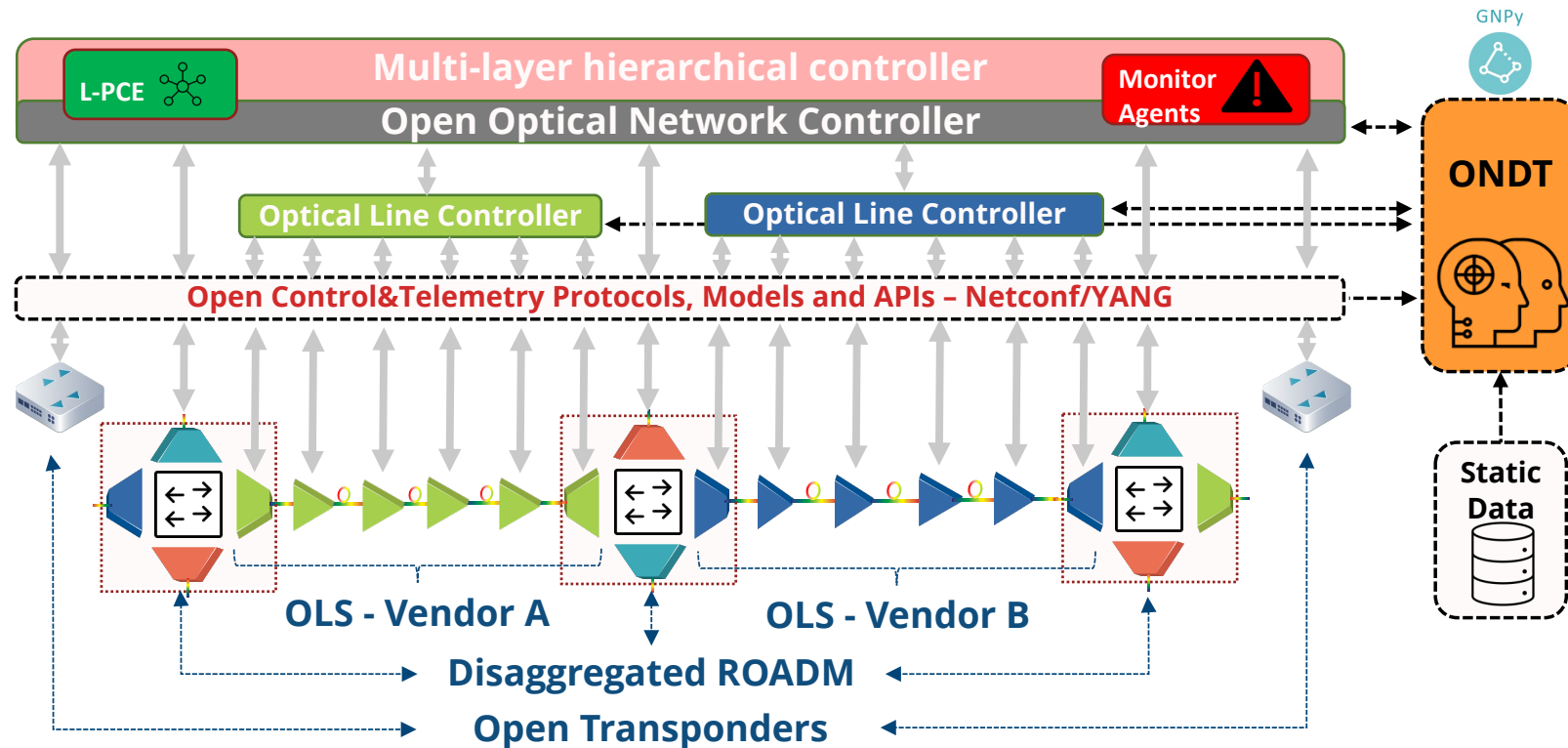
What GNPy is asked for

- Will this lightpath close, at this rate, on this route?
- Which of these three routes has the most headroom?
- If I add forty channels here, what happens to the ones already running?
- This vendor's amplifiers or that vendor's: what does the difference cost me in reach?

Validated, repeatedly, in the field — not only in simulation

- Multi-vendor greenfield testbeds and production brownfield networks · lab and commercial equipment · fed from SDN controllers
- Grammel et al., OFC 2018 · Filer et al., JLT 2018 · Auge, OFC 2019 · Kundrat et al., OFC 2020 · Ferrari et al., OFC 2020 · Ferrari et al., JOCN 2020 · Ferrari et al., ECOC 2020 · D'Amico et al., JOCN 2021 · D'Amico et al., OFC 2022

WHERE THE TWIN SITS IN THE CONTROL PLANE



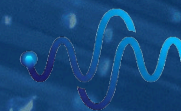
- **L-PCE**: lightpath computation engine: asks the twin whether a candidate route and mode will close, before committing spectrum.
- **Optical line controller**: one per line system, per vendor. Sets amplifier gain and tilt; reconciles its own line against the twin.
- **Monitoring agents**: read telemetry, compare against the twin's prediction, and raise the difference as a signal.



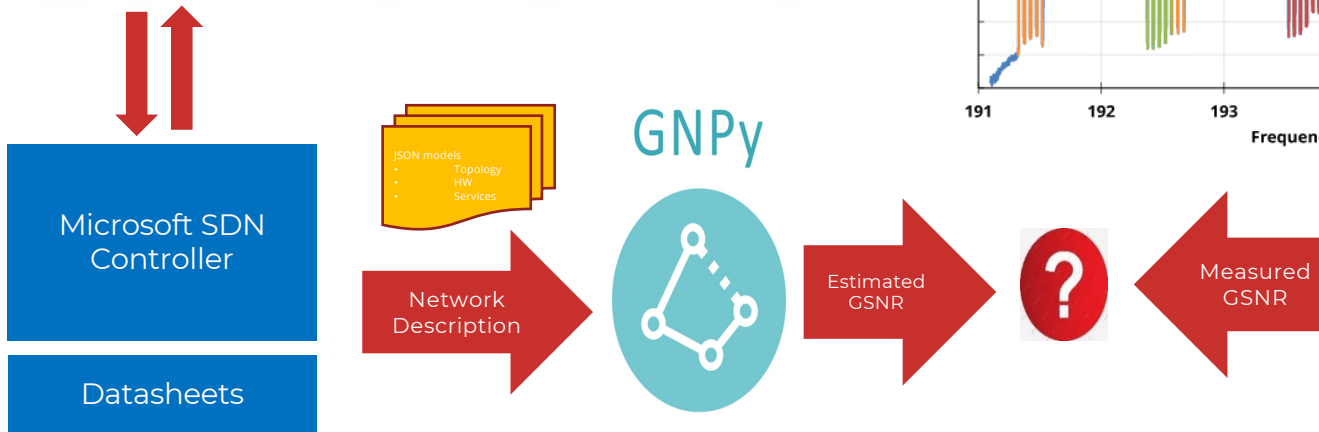
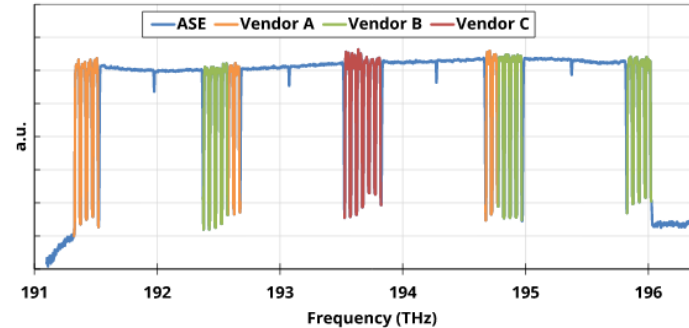
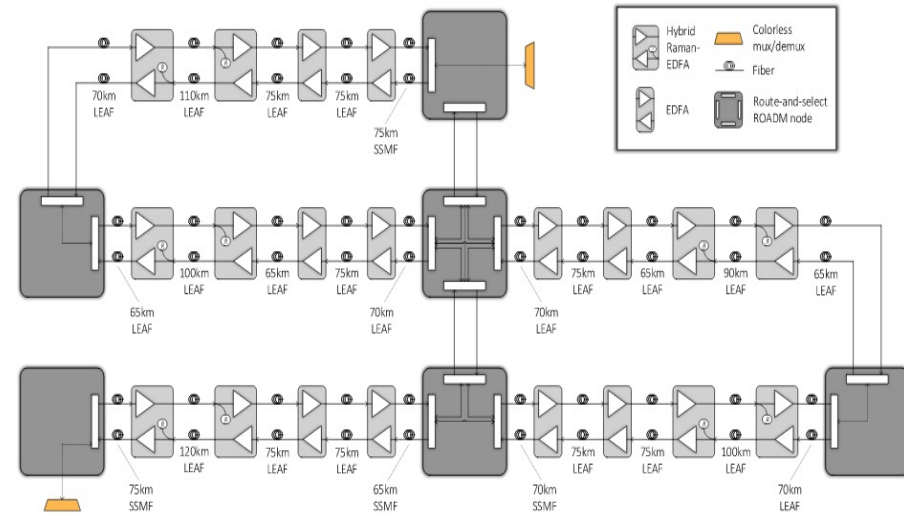
PART 2

QOT-AWARE OPERATIONS

MULTI-VENDOR INTEROPERABILITY,
HONEST MARGINS, AND COORDINATED SPECTRUM



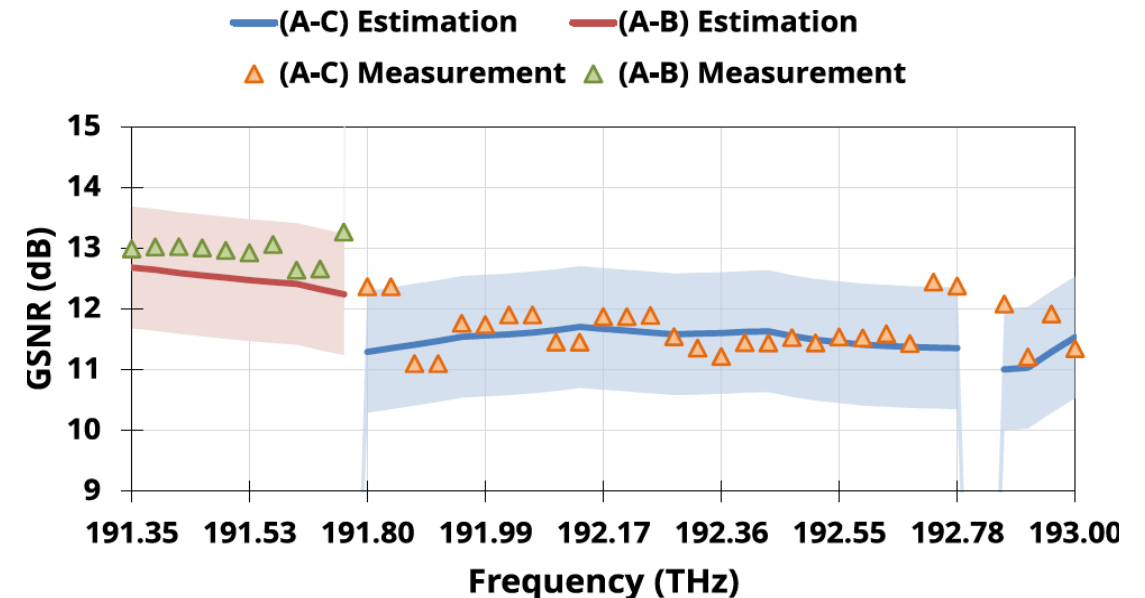
GREENFIELD: A MULTI-VENDOR TESTBED, 500+ CASES



- 6 ROADMs; mixed-fiber line systems from **400 km to 2000 km**.
- **Multi-vendor, multi-rate, multi-format** transceivers across the full C band.
- **500+ tested cases.**
- Crucially: GNPY was fed **from the operator's own SDN controller** (topology, hardware inventory and services as JSON and from datasheets). Not from measurements of the line it was being asked to predict.
- Estimated GSNR was then compared against measured GSNR, case by case.

A. Ferrari et al., "GNPy: an open source application for physical layer aware open optical networks", JOCN, 2020

- A segment of the **Microsoft core network**, carrying real customer traffic, not a testbed.
- Mixed amplification: lumped EDFAs **and** hybrid Raman-EDFA spans.
- Path A–B: **1900 km**, 10 channels. Path A–C: **2150 km**, 28 channels.



± 1.2 dB

full error range, all cases

78 %

of errors conservative

2150 km

longest path validated

Live

customer traffic throughout

This is the result that changes the conversation. A vendor-neutral open-source model, fed from an operator's own inventory, predicted a **deployed, mixed-amplification, two-thousand-kilometer production network to within 1.2 dB** (it was conservative in nearly four cases out of five).

A. Ferrari et al., "Softwarized optical transport QoT in production optical network: a Brownfield validation", ECOC 2020

- Traditional design is **worst-case stacking**. Assume the worst fiber, the worst aging, the worst spectral load, the worst amplifier, all at once. Then add a system margin on top because you are not sure.
- Nothing in that stack is wrong individually. But the joint probability of every worst case occurring on the same lightpath on the same day is **negligible**, and you paid for it anyway.
- Margin is not free. It is paid in **reach, in modulation format, and therefore in fiber**:
 - a few dB of GSNR is the difference between 16QAM and QPSK on the same route;
 - which is the difference between one transceiver pair and two;
 - which, at scale, is the difference between lighting a new route and not.

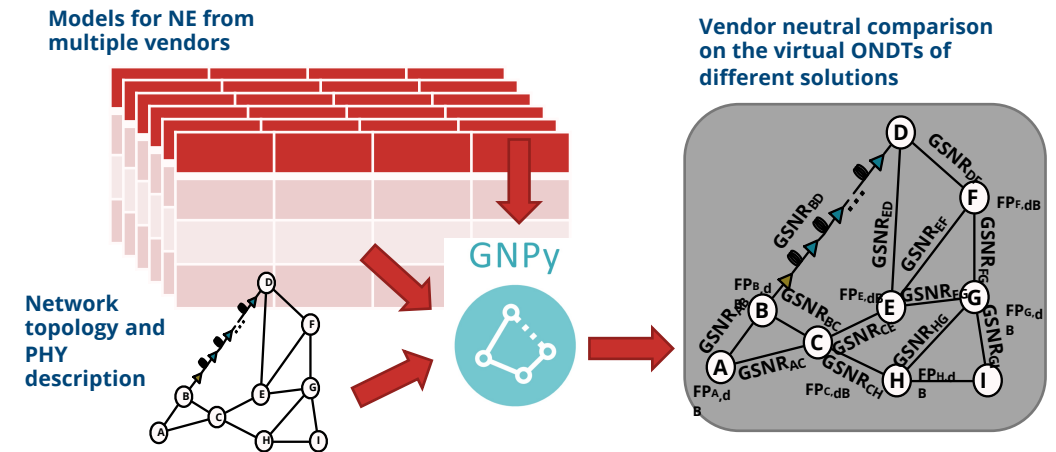
From worst-case to model-driven

- **Design margin**: for what the model genuinely cannot know. Shrinks as the passive layer improves.
- **Aging margin**: real, and it should be tracked over time rather than assumed at day one.
- **Unallocated margin**: the part you were holding because nobody had characterized the model's error. **This is the part the twin gives back.**

The shift is from **static, worst-case planning** to **model-driven decisions**: compute the margin this lightpath actually needs, given this route, this load and this characterized model error, and stop paying for the rest.

WHAT THAT BUYS YOU (2): MULTI-VENDOR, MULTI-GENERATION

- A vendor-neutral twin lets you evaluate network elements from **different suppliers, and different generations of the same supplier, on the same virtual line system.**
- You can ask questions that were previously unanswerable without a procurement cycle:
 - what does vendor A's amplifier cost me in reach versus vendor B's, on **my** topology?
 - can I put this year's 800G transceiver on a line engineered five years ago for 100G?
 - if I upgrade only the amplifiers on this route, what do I get?
- This is the difference between **interoperability on paper** (the interfaces match) and **interoperability in performance** (you know what the combination will actually do).



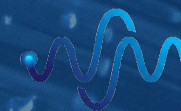
For an NREN this matters more than for a large carrier: you rarely refresh a whole line system at once, so your network is **permanently multi-generation and multi-domain**. The twin is how you keep a heterogeneous plant behaving like one network.



PART 3

SENSING

FROM A MODEL OF SIGNAL QUALITY
TO A MODEL OF PHYSICAL INTEGRITY



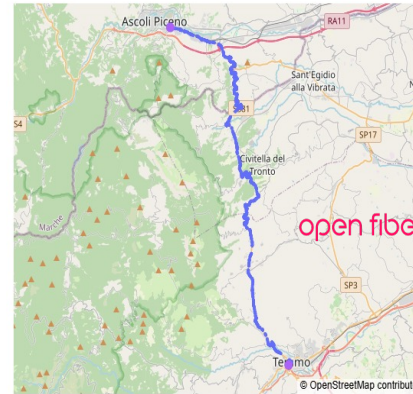
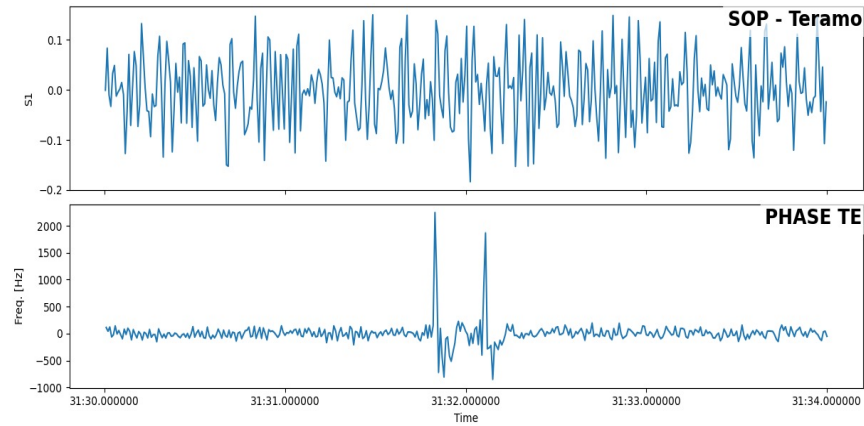
	Distributed sensing (e.g. DAS)	Integrated sensing (phase, SOP)
Localization	10 m	Not possible
Coexist with data traffic	Mostly not	Yes
Long links (unidirectional)	No, or until first amplifier	Yes
Maximum reach	<200 km	> 1000 km

- **DAS:** distributed acoustic sensing. It is the established technique, and it is excellent: coherent detection of Rayleigh backscatter gives roughly ten-meter spatial resolution.
- But it generally needs a **dedicated dark fiber**, it does not survive an amplifier (backscattered light is blocked) so its reach is one span, and it generates an enormous data volume.
- **Integrated sensing (ISAC)** measures what the transmitted signal itself accumulates: optical **phase** and **state of polarization**.
- It **coexists with live traffic by design**, and it reaches **beyond 1000 km**. The price is that it gives you no direct localization, which is a signal-processing problem, not a physics one.

DAS is the right instrument for a route you have decided to study; ISAC for **the network you already have**. And read that “not possible” as “**not from one one-way trace**”: two synchronized ends change the answer, as we will see.

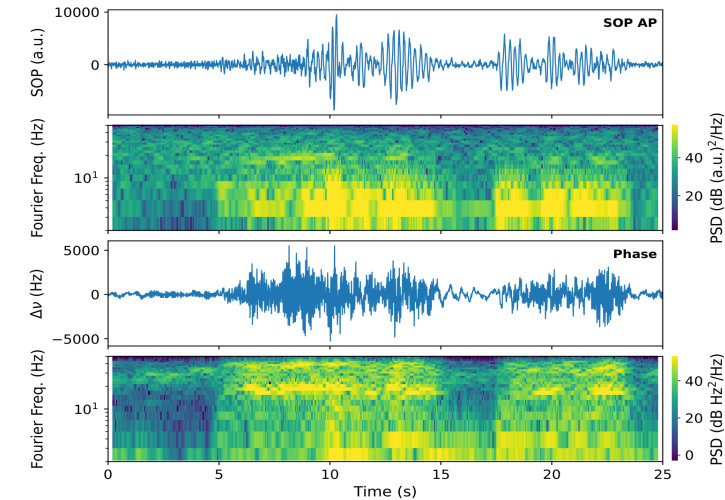
WHAT THE FIBER ACTUALLY HEARS

Vehicle passage — buried segment



- Short, impulsive perturbation.
- Clearly resolved **in phase**.
- **No reliable PBS-SOP signature** in this example: the polarization readout simply does not see it.

Tree impact on the aerial span

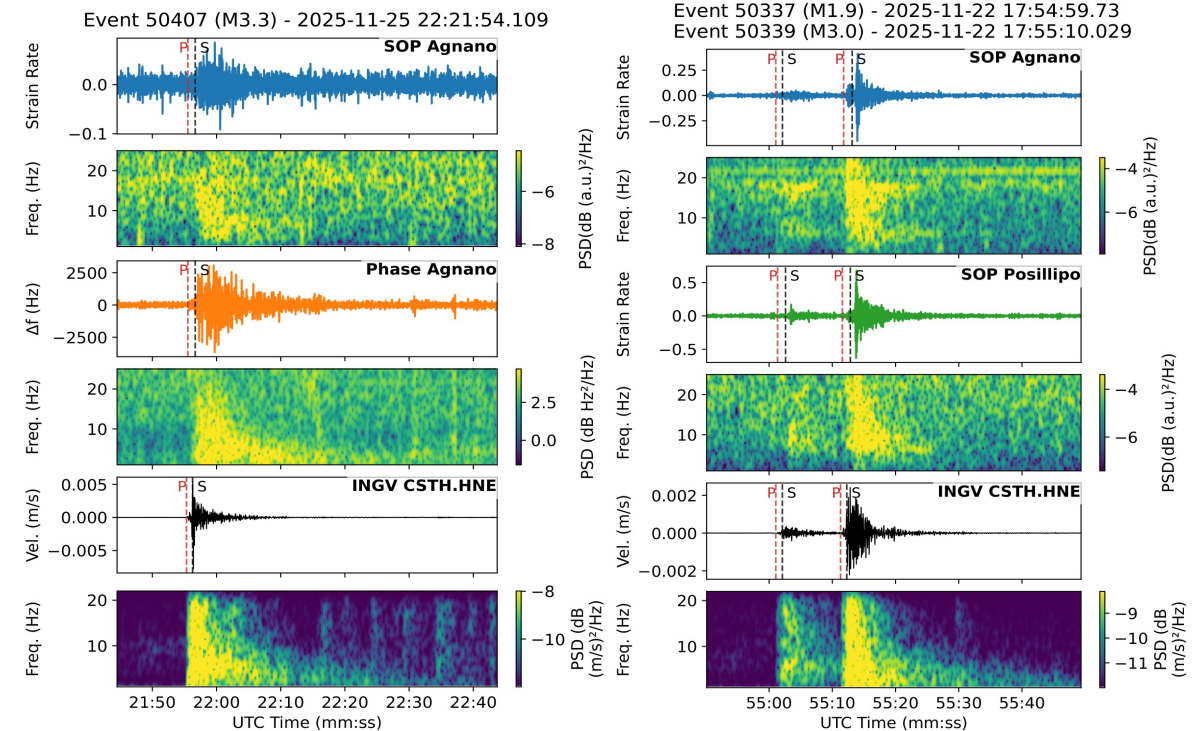


- Large, slow mechanical perturbation.
- Detected by **both** phase and PBS-SOP.
- Phase shows richer oscillatory dynamics; PBS-SOP mainly follows the slower polarization drift.

This pair is the honest summary of the technique. Phase sees more. Polarization sees the **strong and the slow**, which, for cable safety, is a large part of what you want, at a fraction of the cost per site.

TESTBED — CAMPI FLEGREI: SEISMIC EVENTS ON FTTH

- A testbed in **Campi Flegrei, near Naples** (densely populated, high seismic risk, and already covered by conventional seismometers), so results can be checked.
- Two buried **FTTH** routes in loop-back from one PoP: **4.25 km** (Agnano) and **7.25 km** (Posillipo). Built with Open Fiber, INGV, INRiM and Politecnico di Torino. Usage of a dedicated SOP-card.
- The SOP receiver is the low-cost one: a **1.25 Gb/s IMDD SFP+** at **-2 dBm**, sampled at 100 Hz, **coexisting with data services**. No service disruption was recorded.



M_e 1.9

smallest event detected in SOP

10 / 16

cataloged $M > 2$ events in the $M_{4.7}$ sequence

< 1 min

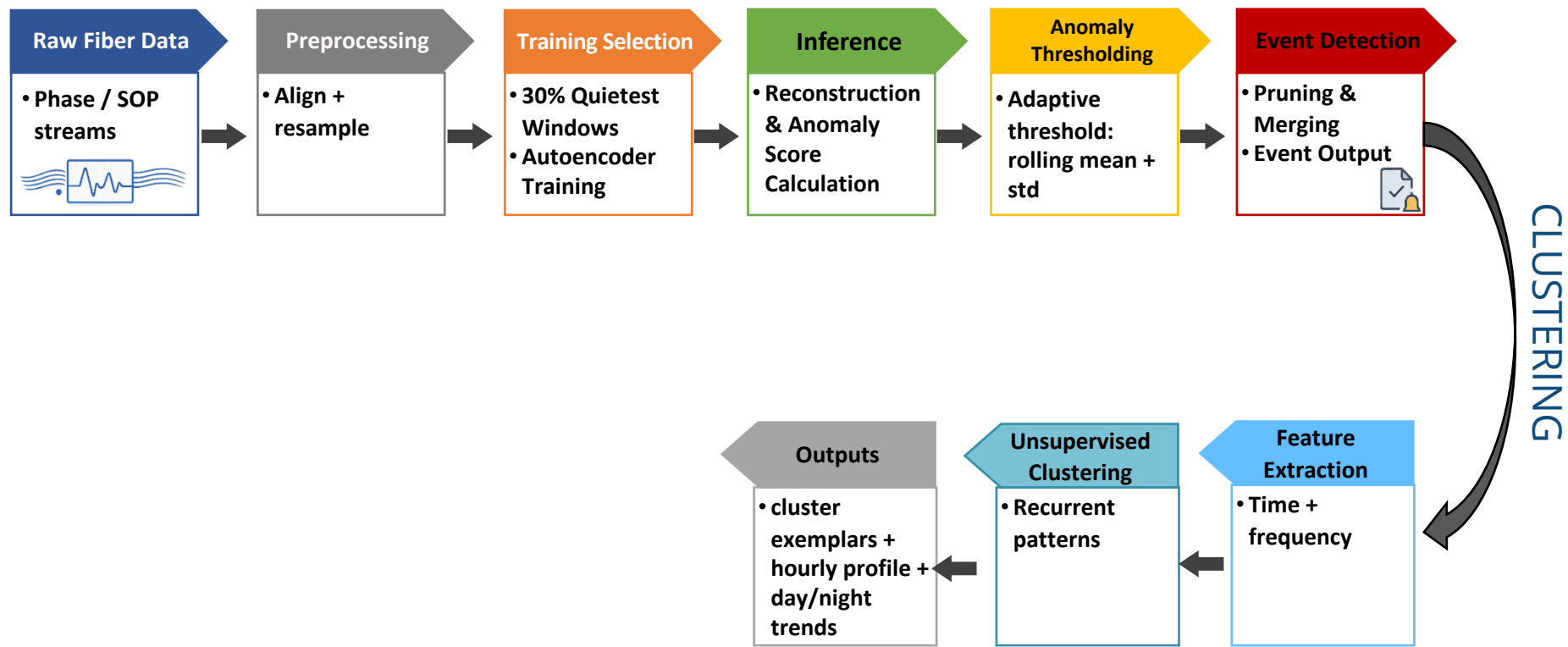
separation between events still resolved

0

service disruptions during the campaign

Cross-checked against coherent interferometry, a 22 km DAS reference on a separate cable, and the INGV CSTH seismic station. **Detection was manual and localization was not demonstrated:** that is the honest state of it.

AUTOENCODER



Align

Phase and SOP traces onto a common time grid.

Detect

An autoencoder trained **only on quiet background**. Reconstruction error becomes an anomaly score, with adaptive thresholding.

Cluster

Cluster the **detected events only**, not the quiet background. Time- and frequency-domain features per event.

Report

Cluster exemplars, hourly profiles, day/night and weekday/weekend trends.

- The pipeline you just saw is unsupervised because the field gives you no labels. But there is a second case: **a link where you control the excitation**, and there you can label.
- **GARR-T metropolitan ring, Turin.** 28.5 km live metro link, three SSF spans through two ROADMs; a CW probe at 1565.49 nm, 3 dBm; a Novoptel polarimeter sampling S_1 , S_2 , S_3 at **1500 Hz**; an Arduino-driven robotic arm inducing repeatable vibrations, against the real urban background of traffic, railway and construction.
- Features are deliberately cheap: Stokes parameters plus rolling mean and standard deviation, lags, and first-to-third derivatives. **XGBoost** outperforms the deep models tried alongside it.
- And the output is not “something happened”. It is **how bad** (five graded classes, from idle baseline to a critical, high-risk event).

Events	Vibration Intensity ; Risk level
Bl	Baseline ; Negligible (idle)
N-v1	Low ; (Ambient noise)
N-v2	Low- Moderate (Ambient noise)
N-v5	Moderate ; (Routine activity)
C-v6	Critical (High risk/Anomalous event)

Graded event classes on the live metro ring

98.02 %

accuracy, 5 classes, on the live link

0.996

AUC on the critical class

< 10 ms

inference, mid-range CPU — edge-deployable

1500 Hz

SOP sampling rate

The two approaches are complements, not rivals: **unsupervised to discover what your cable does**, supervised to **grade what you already know how to provoke**.



Closed-loop: Telemetry plane → AI/ML layer → Control plane → Proactive restoration before signal loss

Chain as demonstrated on the 31.07 km bidirectional loop — classification is 96.2 % there, 98.02 % on the 28.5 km detection link of the previous slide.

- Earlier the table said integrated sensing gives **no localization**. That is true of a single, one-way trace, and it is the standard objection to forward integrated sensing.
- It stops being true with **two ends and a common clock**. Measure SOP at both ends of a **31.07 km bidirectional loop**, synchronize by NTP, and cross-correlate: the time difference of arrival puts the event on the fiber.
- **Median localization error: under 23 meters**, on a buried metro ring.

Which completes the chain

- **Detect** — is this an anomaly?
- **Classify** — how serious is it?
- **Locate** — which stretch of cable?
- **Act** — alarm the controller, reroute **before** the signal is lost.

That last step is the whole argument of this talk, arriving from a different direction: **telemetry plane → AI/ML layer → control plane → proactive restoration**. It is the same closed loop as Part 2, with a different observable, which is exactly why sensing belongs inside the twin and not beside it.

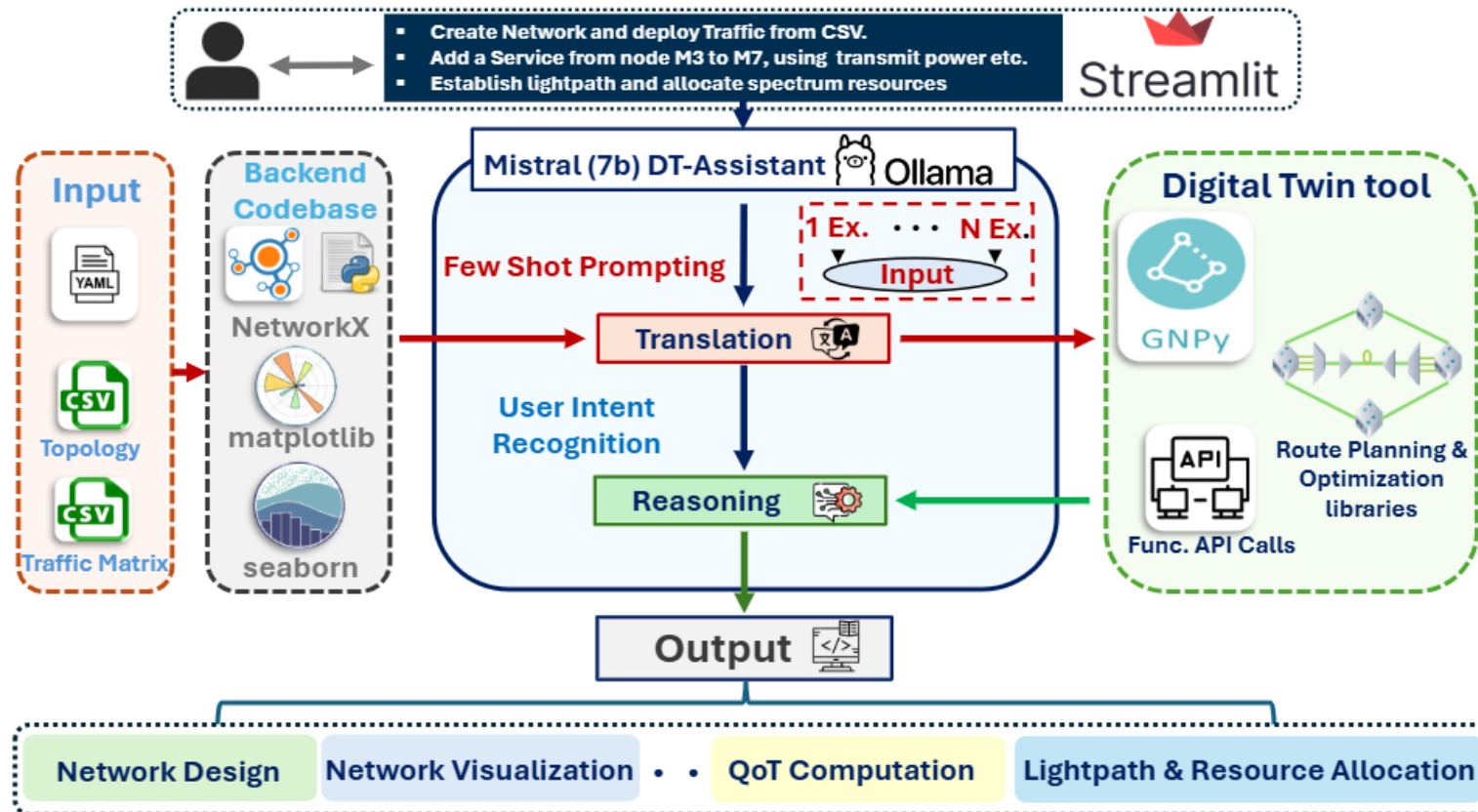


PART 4

LANGUAGE

LLMS GROUNDED IN THE TWIN,
AND WHERE THE LINE SHOULD BE DRAWN

THE ARCHITECTURE: LLM + RAG + DIGITAL TWIN



Reason

Four LLM stages: intent recognition, intent translation, function selection, execution planning.

Retrieve

RAG over heterogeneous sources: topology and telemetry CSV, vendor datasheet PDFs, JSON configs. Embedded, chunked, vector-searched.

Compute

GNPy: OSNR, GSNR, reach and feasibility. **The numbers never come from the model.**

Answer

Topology map, QoT results, grounded explanation, and follow-up questions in context.

WHAT OPEN MODELS ACTUALLY ACHIEVE

Model selection

4 open models compared. **Qwen2.5-7B** chosen for best balance of accuracy, memory footprint and reliable structured output.

83.3 % accuracy

Effect of RAG

Controlled experiment on retrieval-dependent tasks. Grounding is what suppresses invention.

59.7 % → 66.8 %
hallucination 10 % → 3.3 %

Full framework

Benchmark queries across task categories. Strong on topology and routing; **mixed multi-stage queries remain hard**.

77.6 % composite
3.8 % hallucination · 8.37 s

With GNPY in the loop

Intent-driven QoT: natural language → structured operations → physical-layer computation.

89.7 % across tasks

Design & analytics tasks

Topology generation, graph analytics, equipment-aware referencing (Qwen-2.5-14B-Instruct).

96.97 %, lowest hallucination

Read the second row as the headline: **retrieval, not model size, is what removes hallucination**. A 7B model grounded in your documentation beats a larger one guessing.

Reasonable today: management

- **Troubleshooting support.** “What changed on this OMS in the last six hours?”
- **Provisioning assistance.** Draft the service request; a human approves it.
- **Documentation retrieval.** The right vendor parameter out of a 300-page PDF.
- **Design exploration.** Generate candidate topologies and compare them.
- **Onboarding.** A junior engineer can interrogate the network in plain language.

Not yet: closed-loop control

- **Autonomous reconfiguration.** ~4 % hallucination is excellent for a chatbot and unacceptable for a system that moves traffic.
- **Multi-step planning under pressure.** Precisely the category that measures worst.
- **Anything without a human checkpoint.** The failure mode is a confident, well-formed, wrong action.
- **Accountability.** When an automated action causes an outage, “the model decided” is not an answer you can give a regulator.

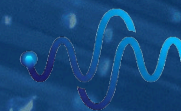
The useful framing: the twin is what makes the language layer **checkable**. Every claim it makes about your network can be recomputed by GNPpy. So put the human where the twin cannot verify.



PART 5

WHAT RECURS, AND WHAT IS MISSING

OPEN INTERFACES, OPEN DATA, AND HYBRID
PHYSICS/ML MODELS



- **TNC24, Rennes.** Renato Ambrosone (Politecnico di Torino) and Matteo Colantonio (**GARR**) asked a question that is not the accuracy question: GNPpy is extensively validated, but **is it hard to use?**
- So they built it. Two ROADMs, a transponder, an ASE source, a recirculating loop with an EDFA; 100 km and 125 km spans. Model it in GNPpy, run it, compare against the measured pre-FEC BER and power spectral distribution.
- **Verdict: it works.** The GSNR came out right.
- What it cost: deep topology knowledge, equipment models, device operating points, and a back-to-back BER-vs-OSNR curve for the transceiver. **That is the translation layer, priced.**

```
{  
  "type_variety": "myVarGainEDFA",  
  "type_def": "variable_gain",  
  "f_min": 🧐,  
  "f_max": 1f$e9,  
  "gain_flatmax": 🐒,  
  "gain_min": 🐒,  
  "p_max": 🐒,  
  "nf_min": 2$.*#,  
  "nf_max": 🧐  
}
```

The amplifier model an operator has to fill in. Every redacted field comes from an NDA or an in-lab characterization campaign. — Ambrosone & Colantonio, TNC24

Their three named limitations (**amplifier behavior under partial spectral load**, a **live telemetry interface**, and a **data model**) are the three gaps this talk closes on. And their call to action was this one:

“GNPpy has been developed mainly for design until now... NRENs should start contributing to influence the development.”

R. Ambrosone, M. Colantonio, “Empowering Optical Network Teams Through Digital Twins: a case study on implementing and utilizing GNPpy”, TNC24, Rennes, June 2024

- Every layer of the twin degrades without data, and each degrades differently:
 - the **passive layer** decays quietly: fiber records age, splices are added, nobody updates the database. This is where most of the residual QoT error lives, as the connector-loss sweep showed.
 - the **active layer** needs measurements to validate against. Component models (amplifiers above all) cannot be characterized from datasheets alone.
 - the **sensing layer** needs labeled events, and the labels do not exist. One operator's anomalies are only interpretable next to somebody else's.
 - the **language layer** needs grounded documentation, and vendor documentation is unstructured and heterogeneous by nature.

What is genuinely hard about this

- Span losses and route detail are commercially sensitive.
- Vendor performance data is contractually restricted.
- Failure data is reputationally sensitive.
- None of these objections is unreasonable.

What could be done anyway

- Anonymized and aggregated measurement sets: error distributions, not routes.
- Shared component characterizations, agreed with vendors.
- Open, labeled sensing datasets from testbeds, which is precisely why field testbeds matter.
- Research and education networks are unusually well placed to go first.

1) White-box Model (WB)

PROs	• Analytical models by physical laws • High generalizability • High accuracy in full spectral load
CONs	• Low accuracy with variable spectral load

2) Black-box Model (BB)

PROs	• High accuracy in varying spectral load conditions
CONs	• Model transparency, generalizability not ensured • Data-intensive (for ML) & not scalable in dynamic network environments

White box

Analytical, from physical laws. Interpretable and highly generalizable. **Accurate under full spectral load**, and progressively less so as the load varies.

Black box

Data-driven, e.g. a DNN. **Accurate under variable spectral load**. But opaque, data-hungry, and it does not transfer to the next device.

Grey box

The physical model plus a learned correction. Keeps the generalization, buys back the accuracy, and needs far less data than pure ML.

The twin is no longer only a QoT engine.

It is the place where the physical layer becomes computable, and once it is computable, it can be optimized, monitored, and eventually talked to.

The accuracy is good enough to act on.

± 1.2 dB on a live 2150 km production network. That is not a research claim; it is an engineering tolerance, and it should be spent.

Your fiber is already a sensor.

Forward sensing coexists with traffic, reaches beyond 1000 km, and extends the twin from signal quality to physical integrity.

Language belongs above the control loop, not inside it.

Grounded in the twin it is genuinely useful for management. Ungrounded, or in closed loop, it is a well-formed way to be wrong.

The bottleneck is shared, not technical.

Open interfaces between the twin's layers, and open data to keep every layer accurate. An NREN already said so at TNC24, and that call to action still stands.



THANK YOU!

rocco.dingillo@polito.it

rocco.dingillo@synphonia.com

PLANET — Physical Layer Aware Networking - DET, Politecnico di Torino - planet.polito.it

SYMPHONYA — synphonia.com

With thanks to V. Curri, E. Virgillito, F. Notarstefano, I. C. Dipto, G. Malik and the PLANET team,
and to Open Fiber, INRiM, INGV, SM-Optics, GARR, LINKS Foundation and Nokia.

EU Horizon projects SENSEI, NESTOR and ALLEGRO.