Machine Learning Models for Optical Communication Systems & Networks

João Pedro, Rui Manuel Morais

Machine Learning & Artificial Intelligence Take Center Stage

Making Headlines…

How Machine Learning Is Changing Cardiac Ultrasound

Artificial intelligence can speed workflow, take over time-consuming tasks, improve reproducibility and allow more time for patient care

Google's DeepMind is training Waymo's self-driving cars like StarCraft II bots

AI 'EMOTION RECOGNITION' CAN'T BE TRUSTED

The belief that facial expressions reliably correspond to emotions is unfounded, says a new review of the field

The California State Bar Is Considering Allowing Non-Lawyers (And Skynet) To Practice Law

If non-lawyers are allowed to practice, it is very likely to affect solo and small firms the most as they tend to serve the low-income/middle-class market.

Artificial Intelligence Can Now Create Perfumes And Doesn't Even Require Any Sense Of Smell

Will artificial intelligence take your job?

How AI will change your career over the next 20 years.

Is Machine Learning Needed* in Optical Communication Systems & Networks?

* i.e. does it add enough value for what it will cost?

Outline

- Applying Machine Learning
  - When & How is it Useful?
  - When is it Cost-Effective?
- ML in Optical Communication Systems & Networks
  - Enabling Trends
  - Challenges
  - (Some) Use Cases
- Conclusions

Applying Machine Learning

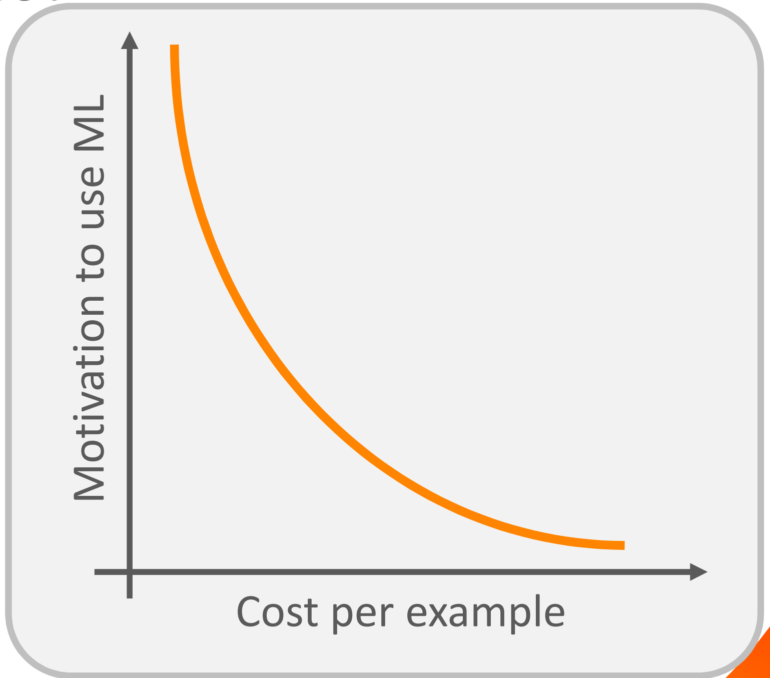
Infinera

Machine Learning: When & How is it Useful?

- ML algorithms can be seen as **universal function approximators**
  - In theory, **applicable to any problem**
  - **Probabilistic** instead of deterministic
- Type of problems that should rely on a probabilistic approach
  - Conventional approaches not applicable or undesirable due to
    - **Model-deficit**: e.g. no physics-based mathematical model exists for the problem due to insufficient domain knowledge
    - **Algorithm-deficit**: e.g. mathematical model is available but algorithms to run it are too complex
  - Not relevant the details of how the task is solved (i.e. compatible with “black-box” approach)
    - Task does not require application of logic, common sense or explicit reasoning based on background knowledge
    - Task does not require explanations about how the decision was made
  - Phenomenon or function being learned is stationary for a sufficiently long period of time
  - Sufficiently large labelled training data set exists or can be created

Machine Learning: When is it Cost-Effective?

- ML algorithms purely **rely on data**
  - Differences in the solutions quality mostly arise from the data set quality, i.e., how representative of the full problem and how balanced is the **data set**
  - *"It's not who has the best algorithm that wins. It's who has the* ***most data****."* Andrew Ng
- When is ML cost-effective against existing / alternative approaches?
  - If a representative data set can be generated with low cost the motivation to use ML is high
    - » Facebook and Google get labelled data almost for free (e.g. FB "10 Year Challenge")
  - If extended machinery and man power is required the motivation shrinks
    - » Increasing the cost per example: as a result of deploying more / upgrade existing monitoring devices, creating dedicated set ups to generate data, etc.
    - » Demands showing higher benefits from adopting ML



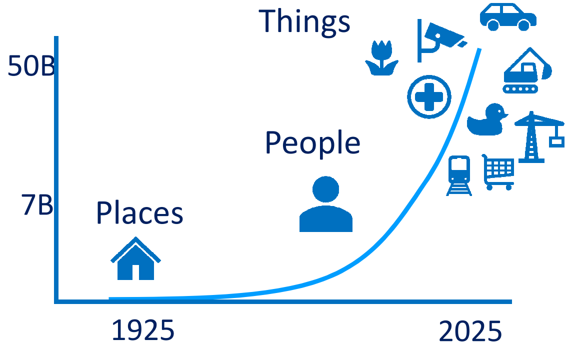
ML in Optical Communication Systems & Networks

Infinera®

Evolution to 5G / Rise of IoT…

Network density will increase with 5G

10x density, into every neighborhood
Much larger network footprint, elements and services
Need to keep OpEx low and workforce constant



ARPU challenge will continue, unlimited consumption will translate into wireless

Landline: Bandwidth x 10 over last years while ARPU ~ ½
Wireless: US changed from bandwidth by use to unlimited consumption
5G can be deployed as fixed-wireless substitution

Operators seek to build 5G networks much faster than 4G

Roll-out of 4G network took 3y for initial build and another 5y to get it everywhere
Roll-out implies huge CapEx and increases OpEx associated to maintenance, upgrades, etc.

Critical To…

Lower CapEx

Lower OpEx

…Core Networks under Pressure

More capacity…

Increase capacity per channel

Maximize capacity of installed fiber infrastructure, postponing expensive fiber roll-outs / additional fiber lease

… while keeping expenditures low…

Reduce cost per bit transported

Reduce footprint and power consumption

Simplify network operation

Facilitate network expansion

…and fostering innovation

Faster introduction of new features

Shorter time-to-market of new services

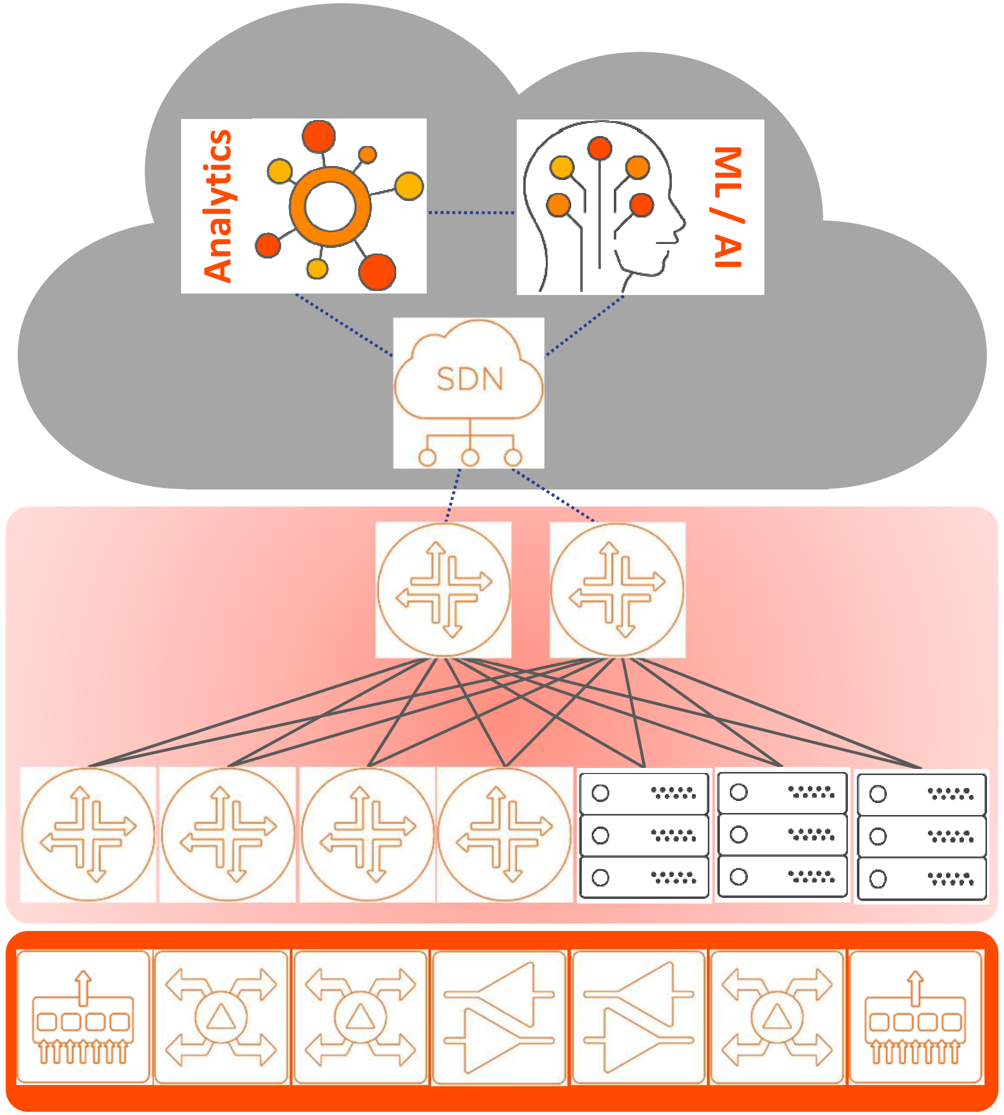
Critical To…

Lower CapEx

Lower OpEx

Setting The Vision

The Autonomous, Self-Learning & Self-Driving Network



Wouldn't it be nice if your network could …

… auto-configure

- » Routing / spectrum / modulation format based on real-time physical layer data
- » Full automation of optical layer

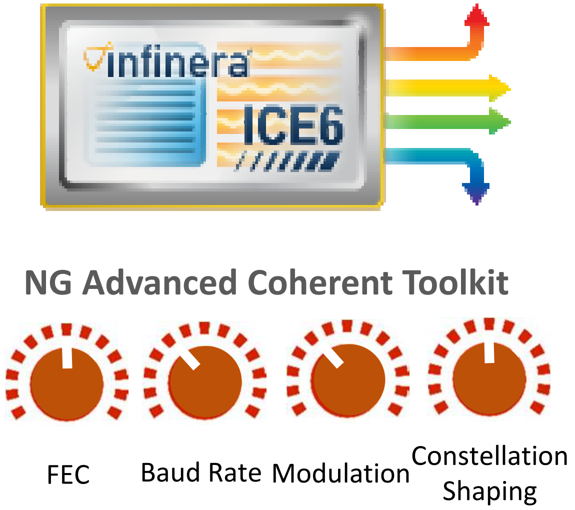
… self-heal

- » Root cause analysis & failure prediction
- » Automatic repair, preventive maintenance

… predict traffic and self-optimize

- » Automate network re-configuration
- » Suggest network augmentation
- » Avoid congested areas

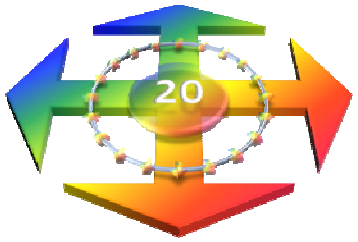
Enabling Trends: Programmable Optical Layer

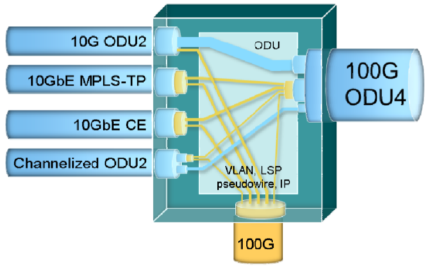


Flexible coherent line interfaces

- High spectral efficiency
- Fine grained trade-off capacity and reach
- Rely on advanced digital signal processing to mitigate physical impairments

Performance data can be collected at the receiver end of a lightpath



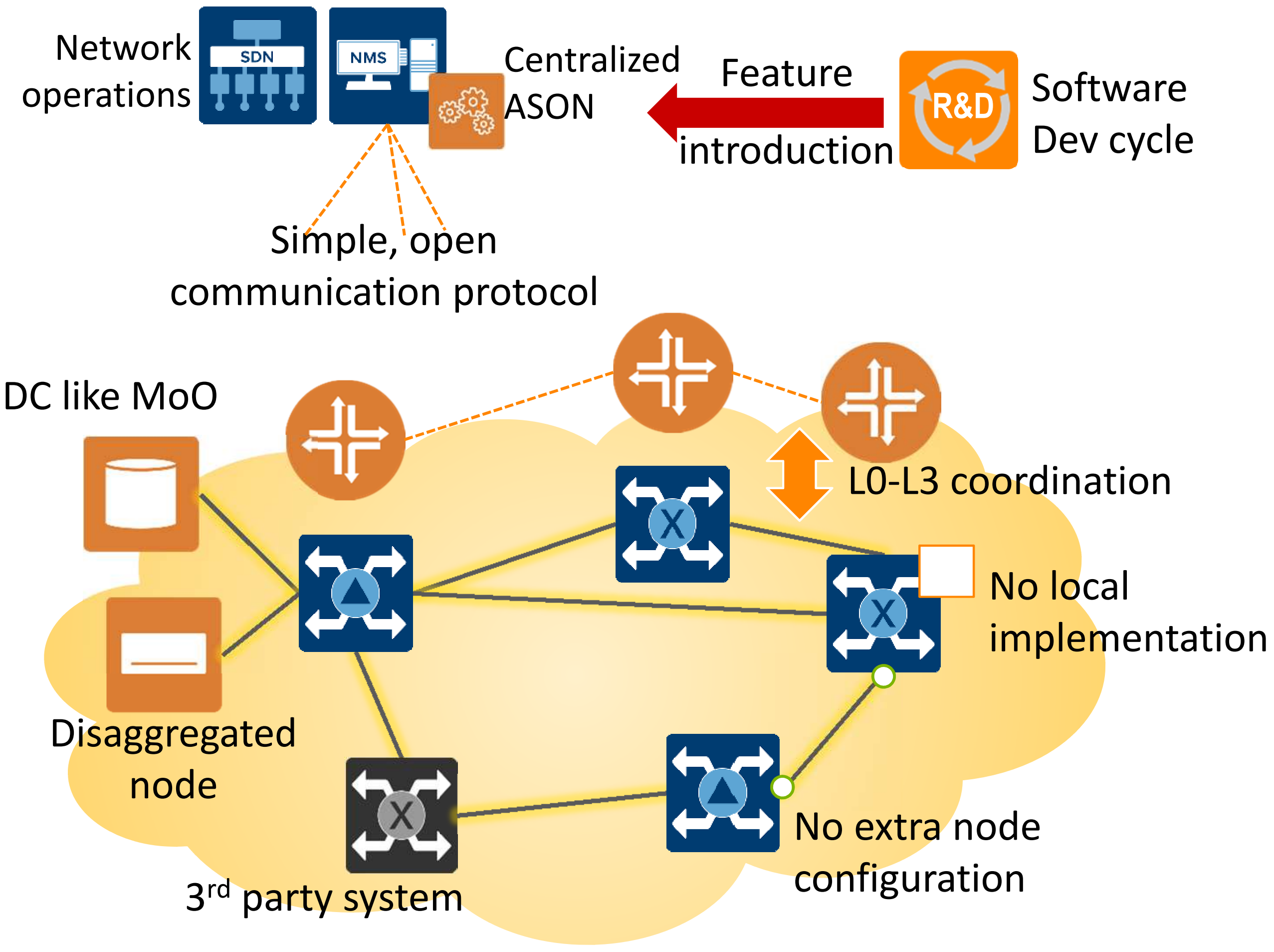


Flexible network elements

- Colorless, Directionless, Contentionless (CDC) ROADMs facilitate reuse of line interfaces, reduce need to send technician to the field, enable advanced restoration schemes
- Hybrid ODU/packet switches enable to mix and mux multiple types of client traffic over the same channel

Network can be remotely reconfigured and adapted to changing conditions

Enabling Trends: SDN Control



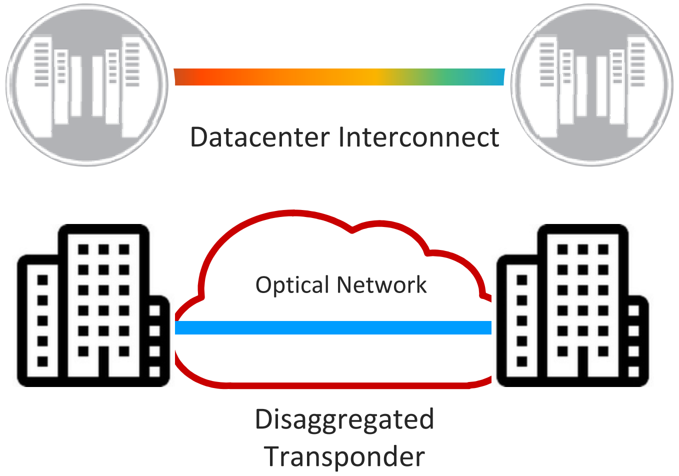
Centralized intelligence

- Global view of network resources & centralized decisions
- Simplified operations via simpler protocols/functions
- Shorter time-to-market via central SW upgrade only
- Improved network resource utilization through optimized multi-layer and multi-domain routes
- Improved resilience via multi-layer L0-L3 coordination in multi-service and IP-Optical networks

Centralized computation facilitates using virtually limitless CPU and storage*

*when compared to distributed control plane relying on network element capabilities

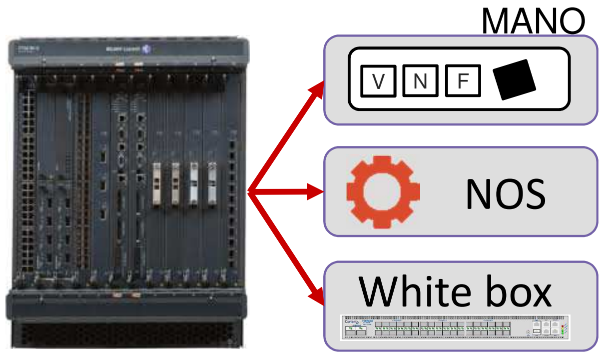
Enabling Trends: Disaggregation



Disaggregated transport platforms and line systems

- Multi-vendor interoperability
- Open standards and protocols
- Simultaneously meet the requirements of ICPs, CSPs, enterprises

Estimating end-to-end performance requires agreeing on a common model and/or "learning" it from the devices, systems or lightpaths

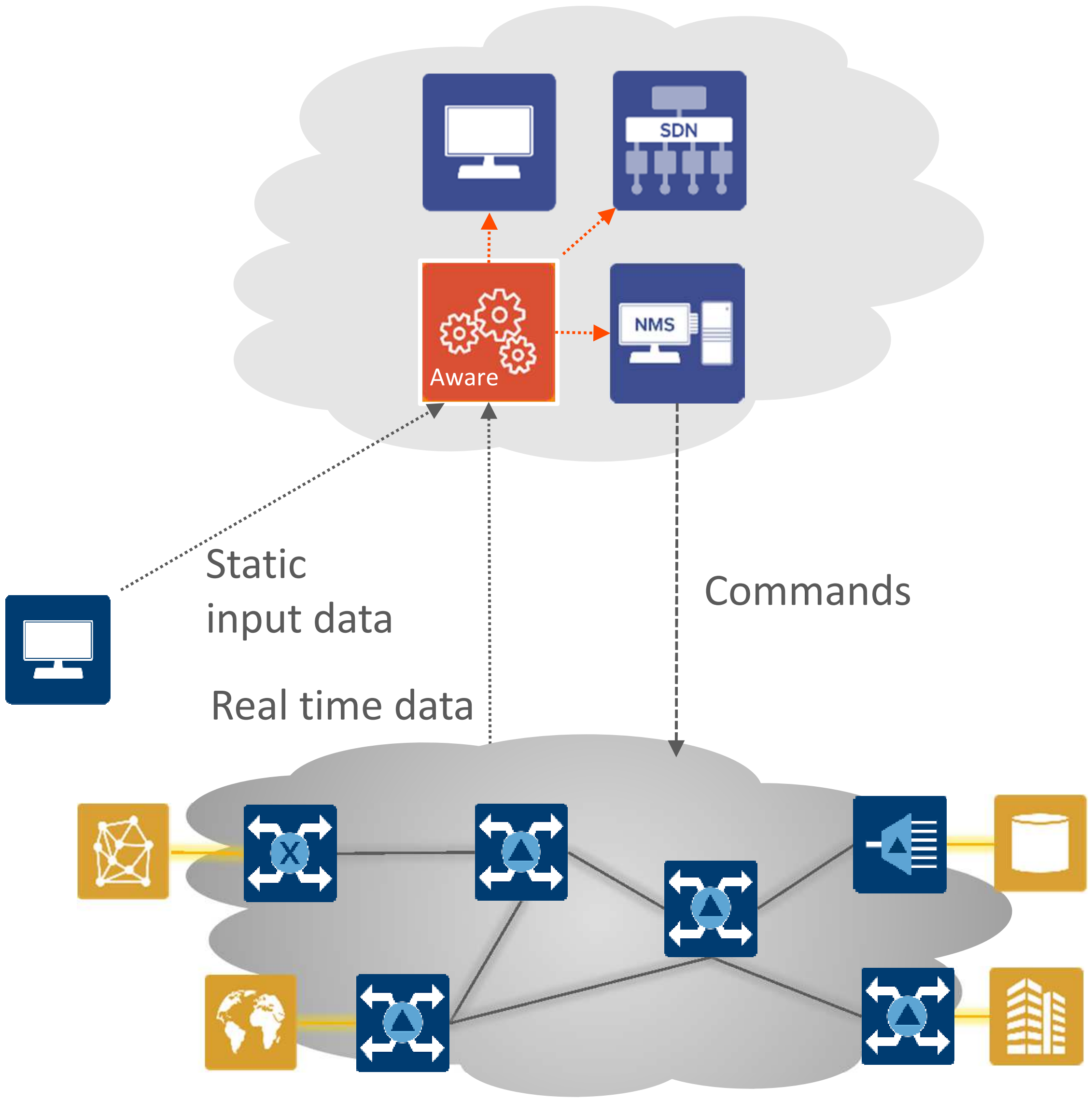


Disaggregated switches/routers

- Open carrier classed disaggregated whitebox switch/router & open software supported by any industry standard whitebox switch/router
- Break vendor lock-in, reduce costs, improve scalability

Common control & orchestration across multiple layers, prone to adopt innovative solutions

Enabling Trends: Low Margins Provisioning



Real time performance planning

- Rely on margin measurements during operation
- Avoid long term projections and worst case assumptions
- Accommodate short/mid term uncertainties, balancing bandwidth and reliability
- Avoid margin stacking / over-dimensioning
- Monitor performance evolution of lightpaths (i.e. service health check)

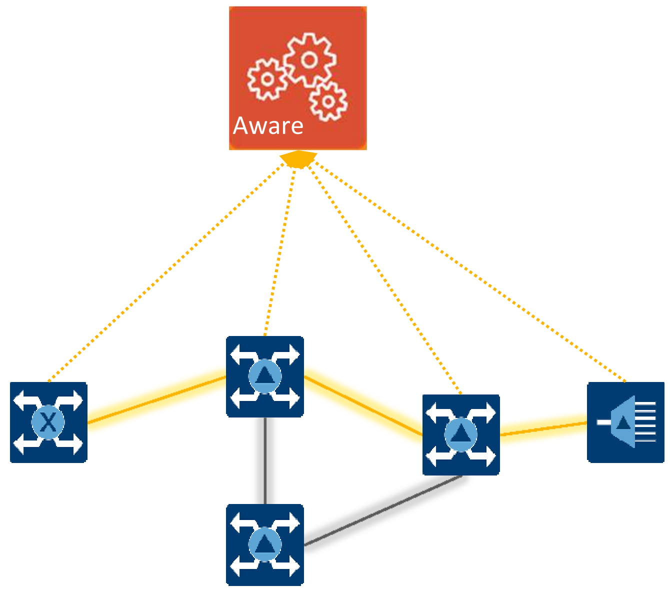
Data intensive approach facilitates estimating performance degradation and trend analysis

Enabling Trends: An Example

Aware – Real-time optical performance

Accurate margin prediction in multi-vendor / disaggregated networks

20%
Optical layer savings

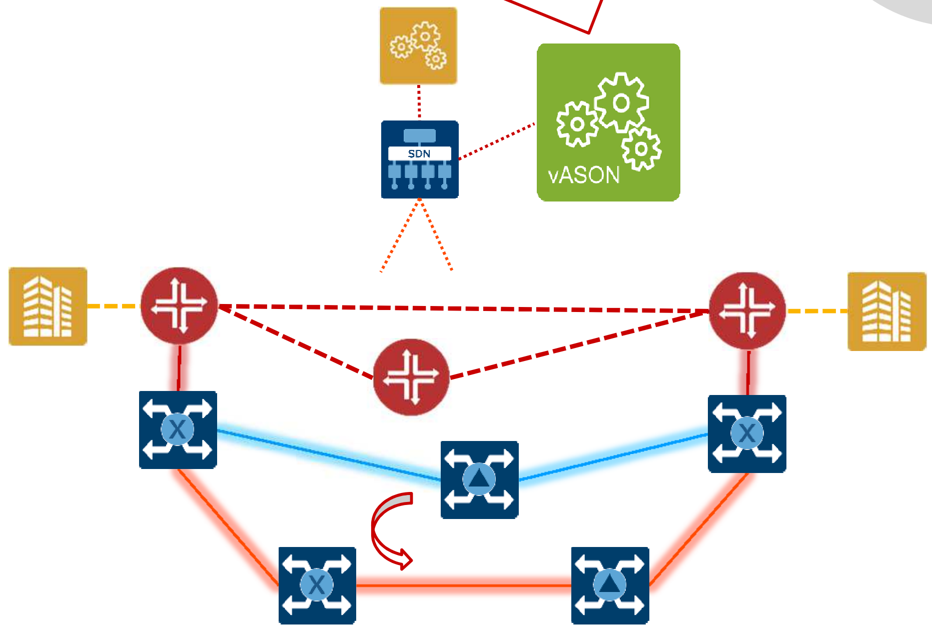


Creates input data for ML algorithms

vASON – Multi-layer/vendor resilience

High resilience for IP-Optical / disaggregated networks

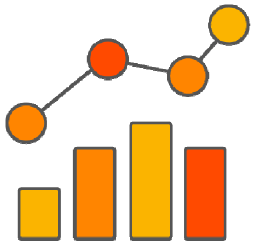
40%
Savings in multi-layer



Path selection can be enhanced with ML

Challenges: A Mature Ecosystem

- Optical networks have been successfully deployed and operated for decades
- Significant investment done in (non-ML) models and processes to operate them
  - ML models need to show improvements in key metrics
  - Usage of ML models has to be cost-effective



- Established practices and labor force accustomed to deterministic approaches
  - Shift in mindset takes time (and money)
  - Practical questions need to be answered: e.g. if there is problem, who's fault is it? The algorithm, the data set? Who collected the data, who supplied the equipment?

- Existing network infrastructure may not be ready to exploit ML
  - Data collection, storage and processing need to be enhanced
  - Additional investment, when needed, has to pay off

Challenges: Data Sources & Access

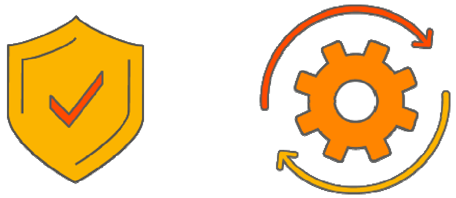
- Not every community has privileges to directly access every source of data

User Community	Data Sources
"Open Access" (e.g. broad research community)	▪ Analytical and simulation-based models and databases that are public ▪ Academic lab setups (usually limited in size and with limited access to device and sub-systems internal details)
System Vendors	▪ Same as "Open Access", plus ▪ Detailed characterization of (their) devices and sub-systems properties (but with limited visibility of phenomena arising in a complex, heterogeneous, live network)
Network Operators	▪ Same as "Open Access", plus ▪ Specific deployment instances (with visibility over multiple planes, layers, domains, vendors, enforced policies, with impact of time-varying effects)

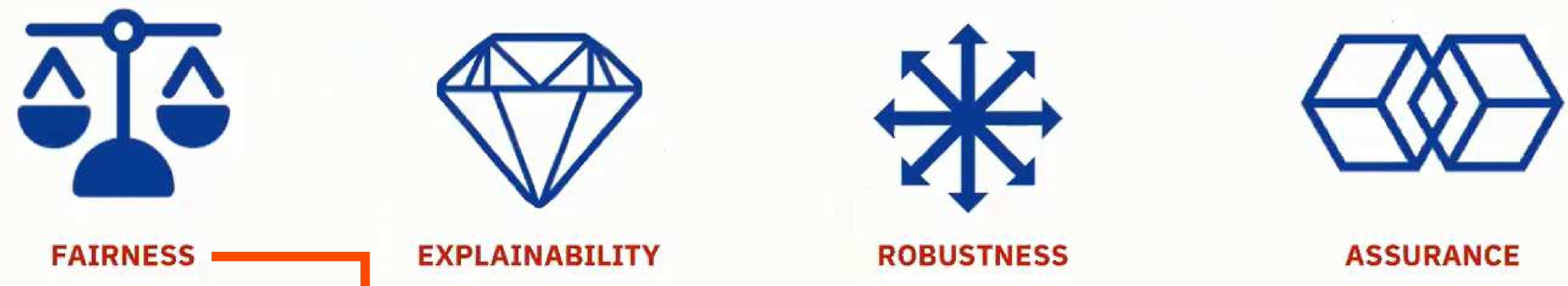
- Partnerships within and between different communities (in various forms) can be key to create and exploit meaningful and rich data sets
  - What are the incentives for sharing data?
  - How to handle confidential data?

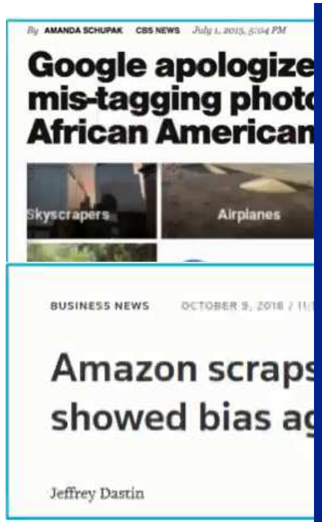
Challenges: ML Lifecycle Management

- Recall that ML applications are useful if
  - Phenomenon or function being learned is stationary for a sufficiently long period of time
  - Sufficiently large labelled training data set exists or can be created
- ML models cannot be seen as standalone applications immutable during their lifecycle
  - Monitoring tools need to be employed in order to control biases of the data as well as changes in the behavior of the problem
  - *What does it take to trust a decision made by a machine (other than that it is 99% accurate)?*
    - Is it fair?
    - Is it easy to understand?
    - Can someone have tampered with it?
    - Is it accountable?



Challenges: ML Lifecycle Management



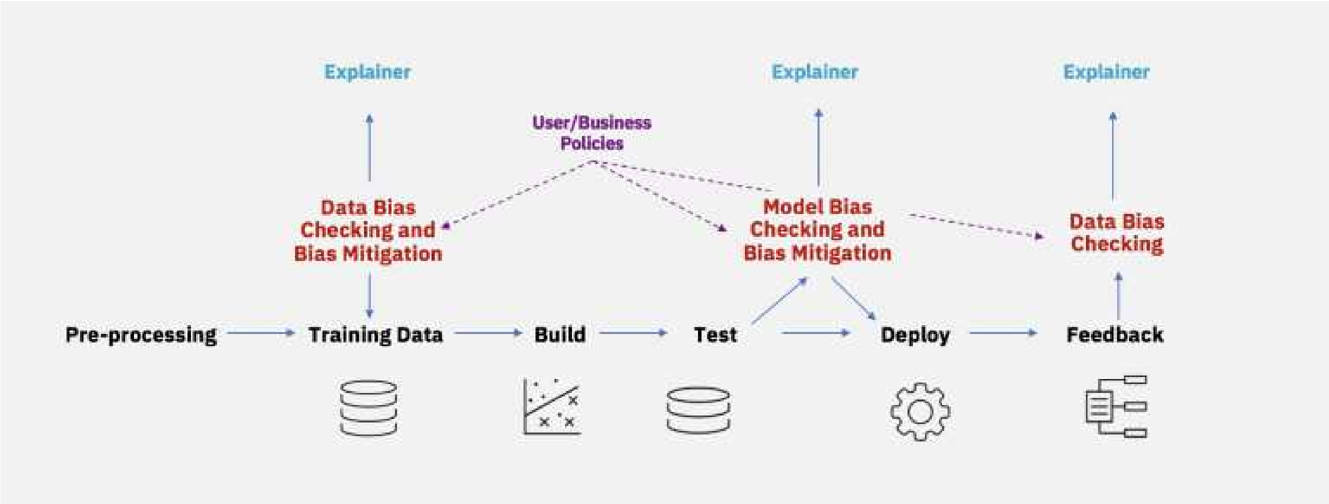


AI Fairness 360

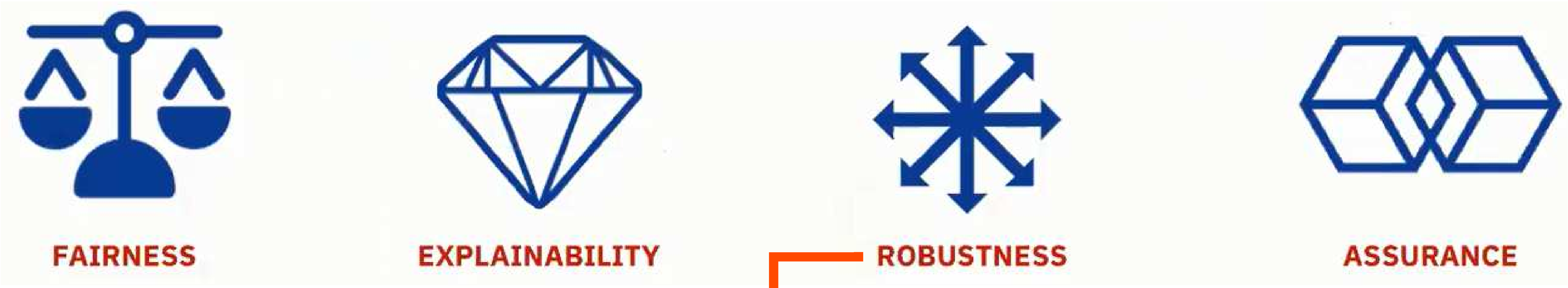
https://github.com/IBM/AIF360

AIF360 toolkit is an open-source library to help detect and remove bias in machine learning models.

The AI Fairness 360 Python package includes a comprehensive set of metrics for datasets and models to test for biases, explanations for these metrics, and algorithms to mitigate bias in datasets and models.



Challenges: ML Lifecycle Management



IBM ART – Adversarial Robustness Toolbox

Evasion attacks
- FGSM
- JSMA
- BIM
- PGD
- Carlini & Wagner
- DeepFool
- NewtonFool
- Universal perturbation

Evasion defenses
- Feature squeezing
- Spatial smoothing
- Label smoothing
- Adversarial training
- Virtual adversarial training
- Thermometer encoding
- Gaussian data augmentation

Poisoning detection
- Detection based on clustering activations
- Proof of attack strategy

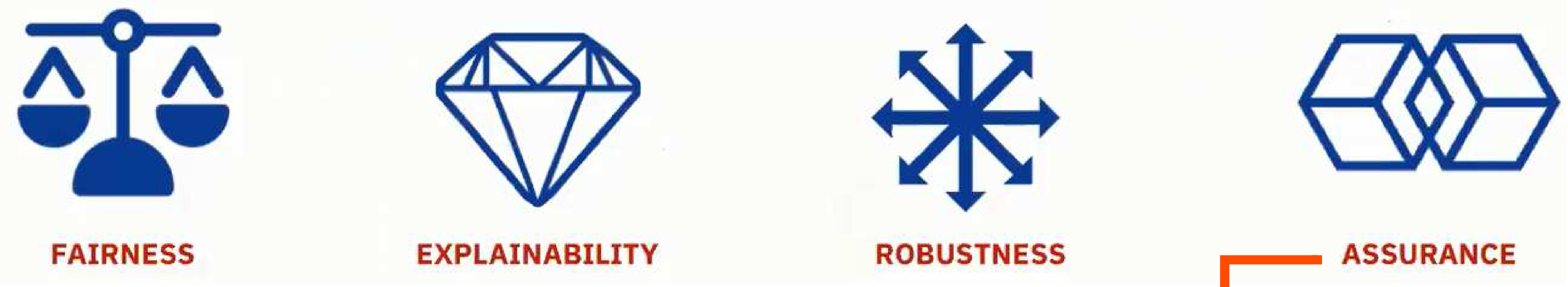
Evasion detection
- Detector based on inputs
- Detector based on activations

Robustness metrics
- CLEVER
- Empirical robustness
- Loss sensitivity

Unified model API
- Training
- Prediction
- Access to loss and prediction gradients

Infinera

Challenges: ML Lifecycle Management



Datasheets for Datasets

https://arxiv.org/abs/1803.09010

- The ML community has no standardized way to document how and why a dataset was created, what information it contains, what tasks it should and should not be used for, and whether it might raise any ethical or legal concerns
- Datasheets for datasets will facilitate communication between dataset creators and users, and encourage the ML community to prioritize transparency and accountability

Supplier's Declaration of Conformity (SDoC)

https://arxiv.org/abs/1808.07261

- A SDoC is a transparent, standardized, but often not legally required, document used to describe the lineage of a product along with the safety and performance testing it has undergone
- SDoC for AI services should contain purpose, performance, safety, security, and provenance information for examination by consumers

(Some) Use Cases

- ML models for optical communication systems & networks are being investigated for a wide array of use cases
  - From modelling individual device (e.g. EDFA) to capturing network-wide behavior (e.g. lightpath QoT)
  - From modelling physical-layer impairments (e.g. BER, OSNR) to predicting traffic flows / patterns
- Use cases can also differ in their potential impact and complexity of use
  - “Low-hanging fruits” (data already available or easy to acquire, low effort to deploy ML model) vs. “starting from scratch” (data collection and ML model usage demand significant effort)
  - “Incremental benefit” vs. “Paradigm shift in network operation”
- Two use cases are illustrated
  - QoT estimation focusing on non-linear impairments
  - Advanced failure prediction

(Some) Use Cases: Non-Linear Impairments Estimation

- Traditional methods to estimate optical performance
  - Simplified Heuristic / Analytical Models
    - ✓ Fast execution, enabling a trade-off accuracy vs complexity when they are designed
    - ✗ Limited accuracy (usually set to be conservative)
    - ✗ Valid only within a defined operation range and well-defined conditions
  - Complex Numerical Simulation
    - ✓ High accuracy
    - ✗ …but requires detailed and accurate characterization of the lightpath
    - ✗ Complex and very slow to execute
  - Live Monitoring
    - ✓ Most reliable performance estimation for established lightpaths
    - ✗ Not self-sufficient to estimate performance of unestablished lightpaths
    - ✗ Only applicable when the network is operating

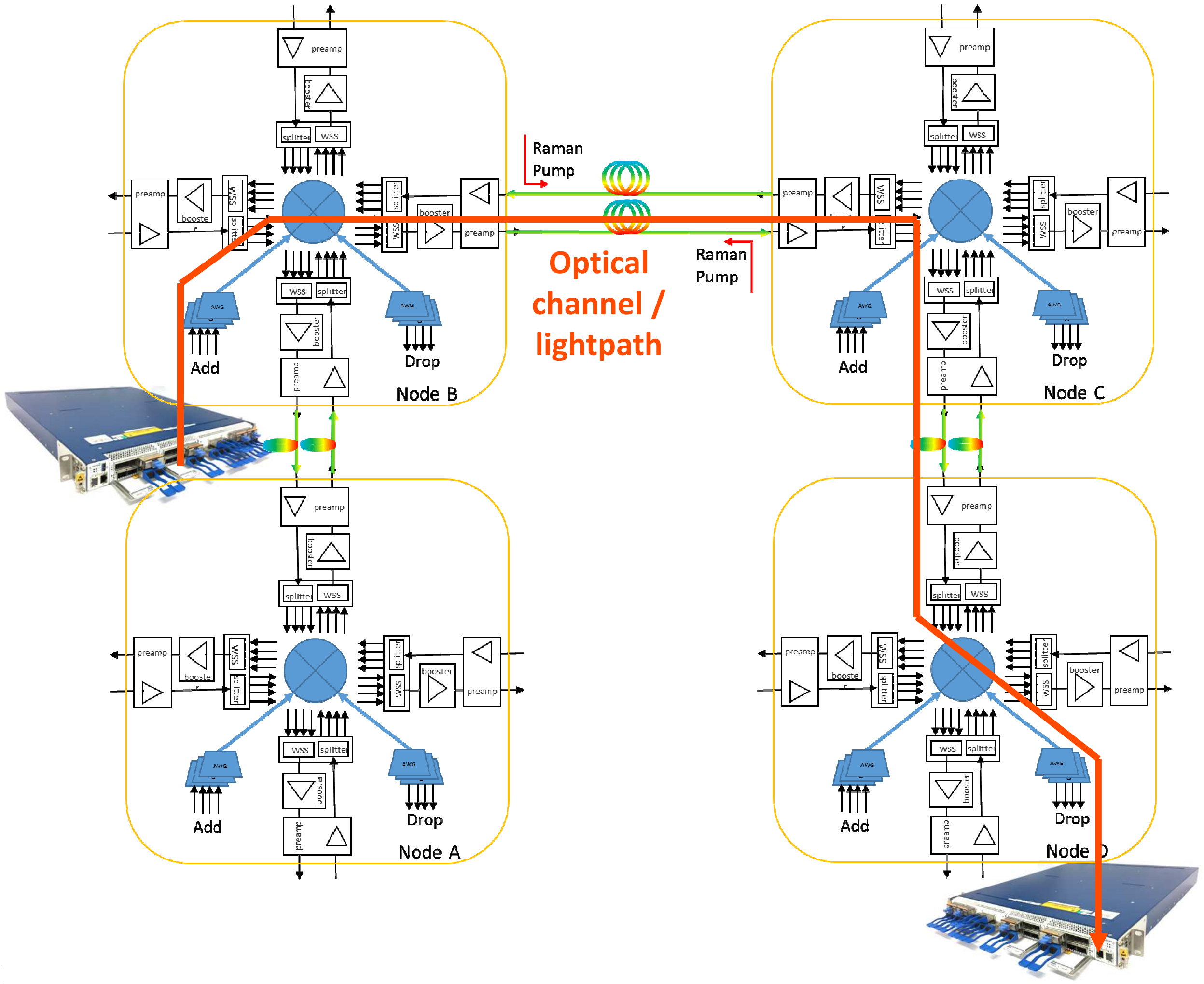
Dominant method for offline planning tools and online provisioning platforms

(Some) Use Cases: Non-Linear Impairments Estimation

- Ideally, optical performance estimation would meet the following targets
  - Fast execution
    - ✓ Enabling applications with strict execution time requirements (e.g. optical restoration)
    - ✓ Allowing network planners to explore more network configurations in tender planning
  - High accuracy
    - ✓ Enabling CapEx optimized tender planning (i.e. “better planning tools and tender planning”)
    - ✓ Improving resource utilization in running networks (i.e. “better online provisioning platforms”)
  - Easily embed measurements from a monitoring system
    - ✓ Further improving accuracy and improving resource utilization in running networks
  - Seamlessly support more than one mode of operation
    - ✓ Enabling to easily switch between, for example, a low precision industry standard optical model and a high precision proprietary optical model in the same engine

(Some) Use Cases: Non-Linear Impairments Estimation

- Focus on a sub-set of physical impairments: Non-linear (NL) impairments
  - Difficult to compute, leading to cumbersome models
  - Relevant in high symbol rate, flexible-grid networks
- Exemplify potential of ML models to improve trade-off accuracy vs computational effort
  - Baseline: incoherent Gaussian Noise (iGN) model
    - » Simple and fast, but usually conservative due to several assumptions, such as neglecting coherent interference, equal number of channels in the vicinity of the center channel, etc.
  - Detailed: Enhanced Gaussian Noise (EGN) model
    - » More accurate, but complex / slow to compute in order to account for self-channel, cross-channel and multi-channel effects, modulation format dependency, etc.



(Some) Use Cases: Non-Linear Impairments Estimation

Fiber Parameters

- Total number of spans
- Span length [km]
- Attenuation [dB/km]
- Effective area [μm^2]
- Dispersion coefficient [ps/(nm.km)]
- Non-linear refractive index [$10^{-20} m^2/W$]

Channel Parameters

- Total number of channels
- Power [W]
- Frequency [THz]
- Bandwidth [GHz]
- Modulation format
- etc.

NLI-ANN Model

Non-linear OSNR [dB]

Standalone ANN

- NL OSNR estimation computed by a single ANN: ANN returns absolute value
- Implementation granting the fastest execution time

Hybrid iGN-ANN

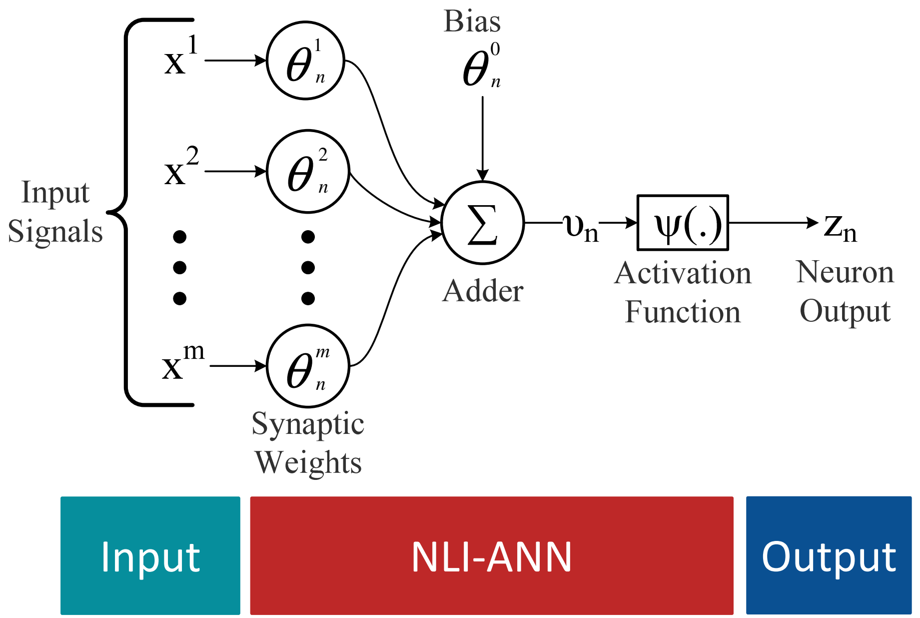
- NL OSNR estimation is a mix of iGN model and ANN: first estimation with closed-formula iGN model followed by correction factor with ANN
- ANN can be configured to only make fine adjustments
- Mitigates “human anxiety” over relying only on ML methods

(Some) Use Cases: Non-Linear Impairments Estimation

- Proof of Concept: **Hybrid iGN-ANN***

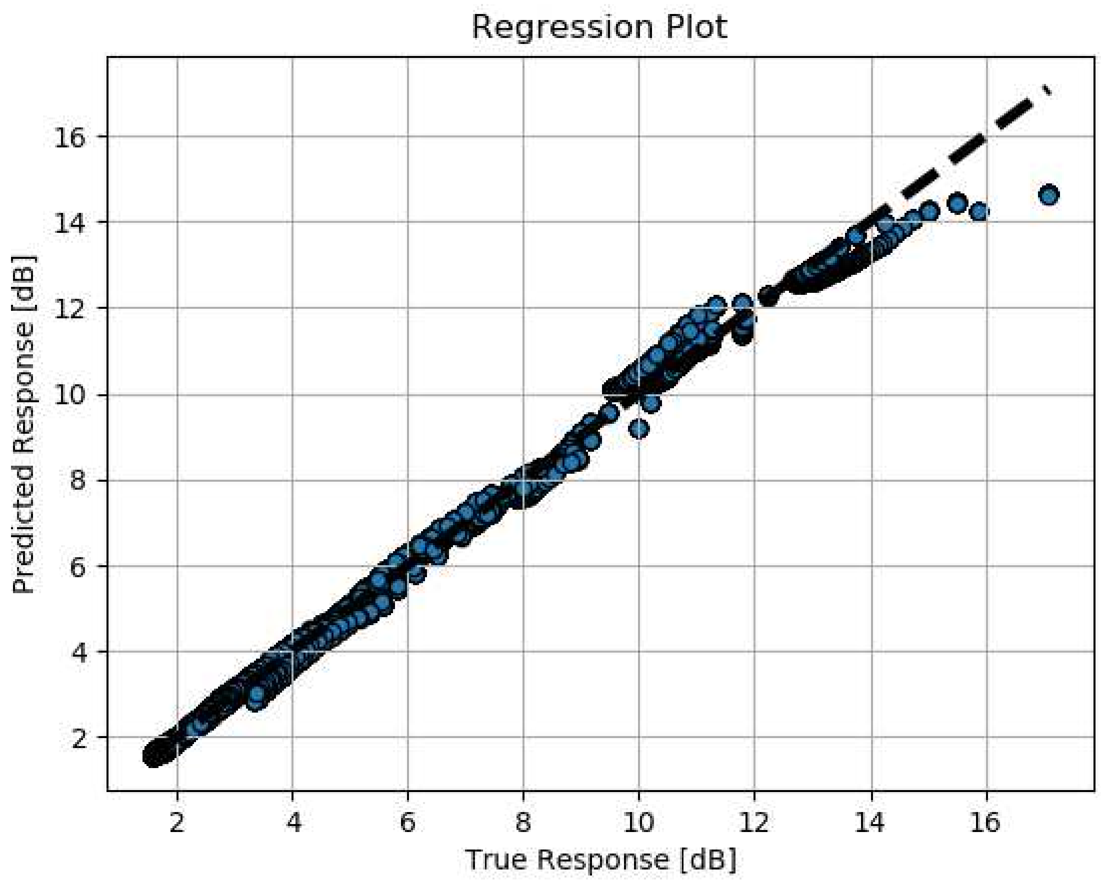
*final results also shown for Standalone ANN

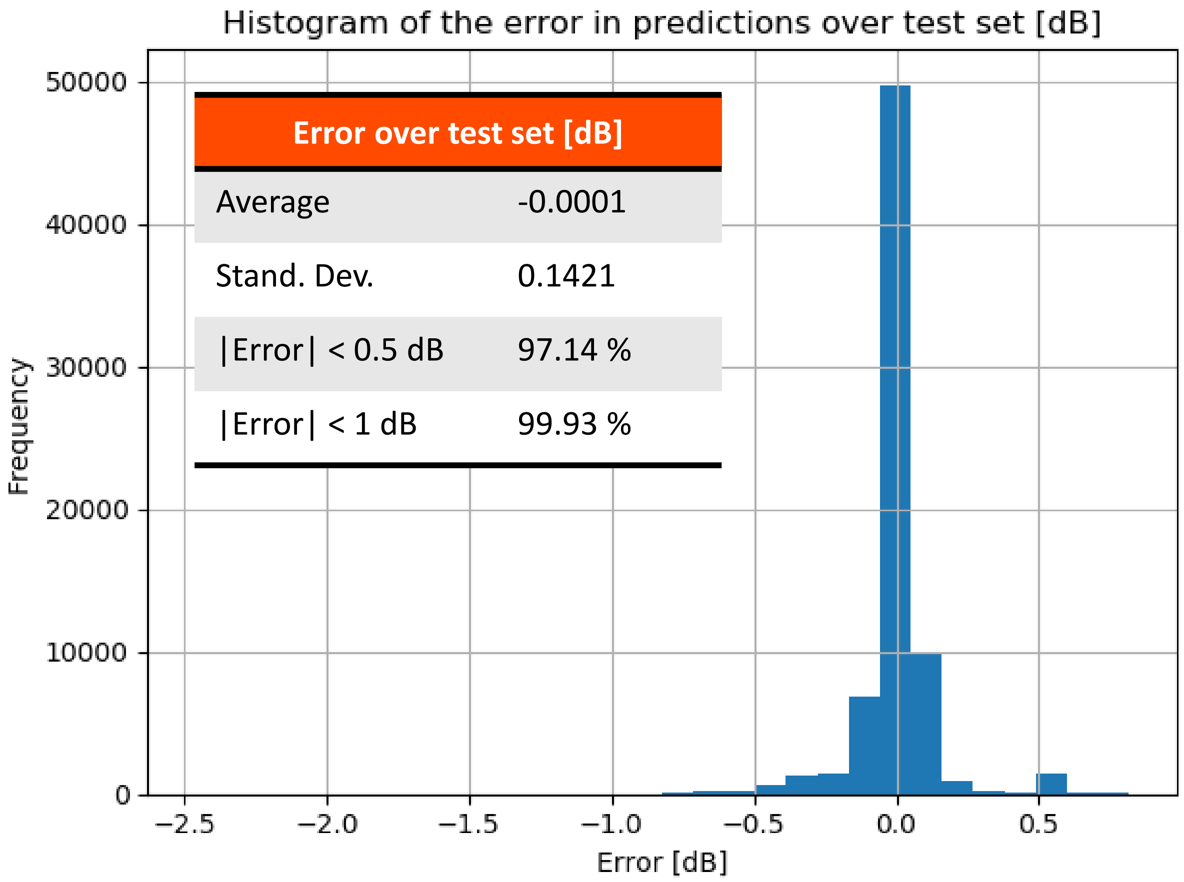
- Data set characterization
  - 73,440 examples: 45 span lengths x 17 launch powers x 96 frequencies
  - Two sets for training
    - » Large set: 51,500 examples for training (70% of total set)
    - » Small set: 12,000 examples for training (optimized selection)
- ANN characterization
  - Hyperparameters are optimized using grid-search
    - » ANN with two hidden layers with 20 neurons each
    - » Output layer with one neuron (regression problem)
    - » Neurons activation function is the relu
    - » Learning method is the stochastic gradient descent
    - » Adaptive learning rate starting with 0.001 (decrease after two epochs w/o improving loss)
    - » Stop criteria is 2000 epochs or average error smaller than 0.0001



(Some) Use Cases: Non-Linear Impairments Estimation

- NL OSNR Accuracy Results: <u>Large</u> training data set
  - ANN obtained a R^2 of 0.998394



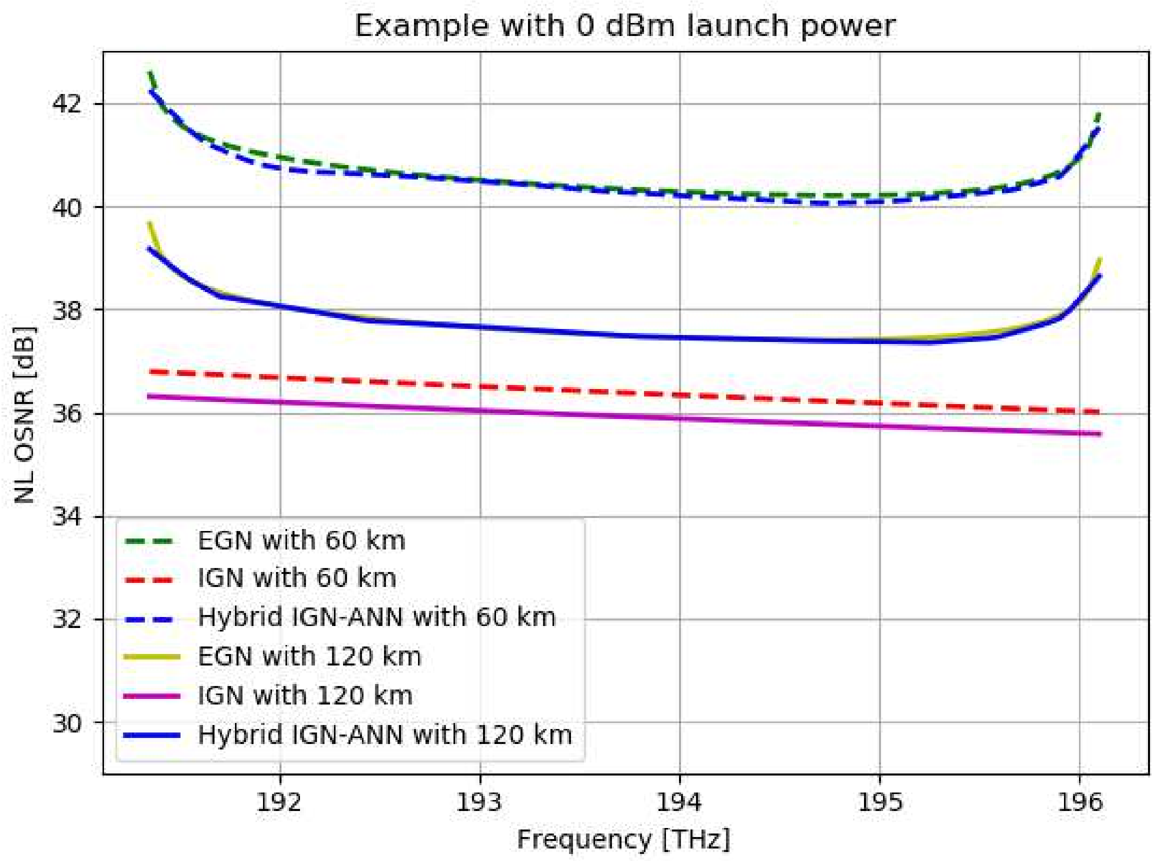


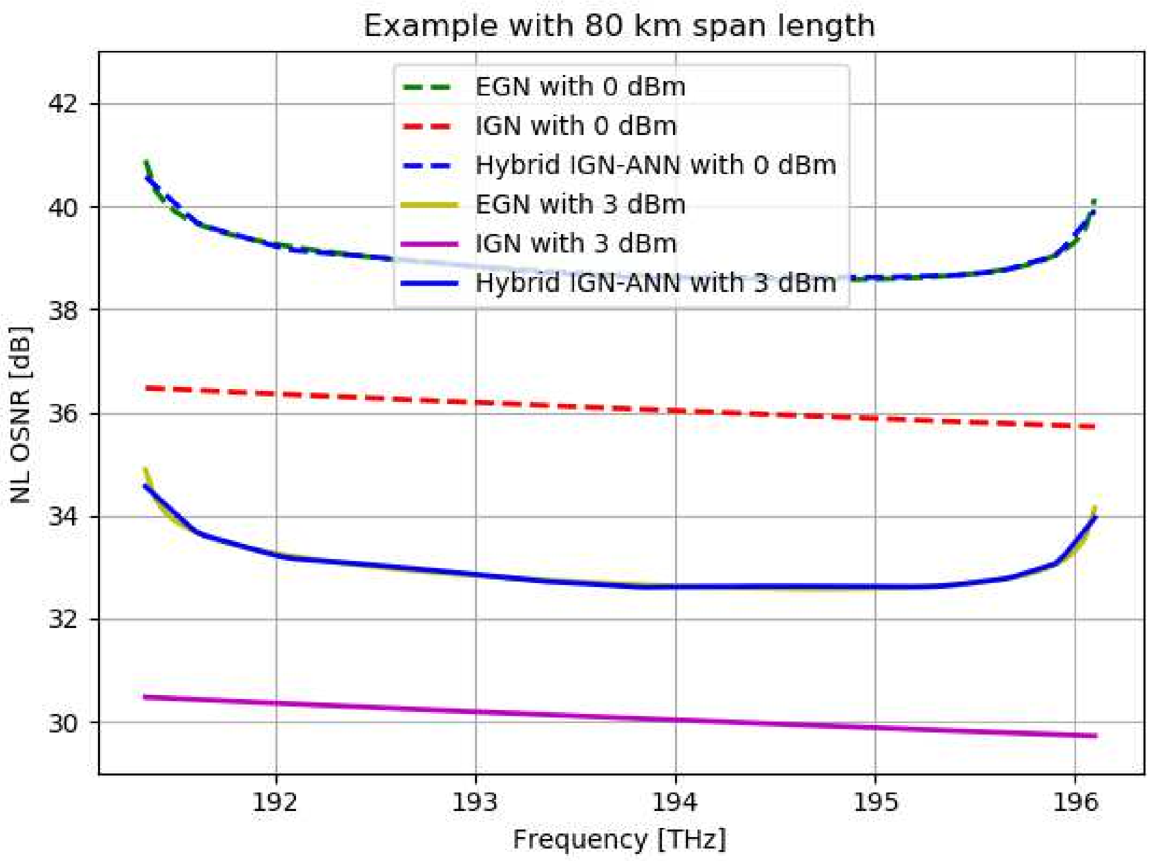
Error over test set [dB]	
Average	-0.0001
Stand. Dev.	0.1421
|Error| < 0.5 dB	97.14 %
|Error| < 1 dB	99.93 %

Infinera

(Some) Use Cases: Non-Linear Impairments Estimation

- NL OSNR Accuracy Results: Large training data set
  - Example of target and estimated NL OSNR as function of frequency for three span lengths (60, 80 and 120 km) and two launch powers (0 and 3 dBm)

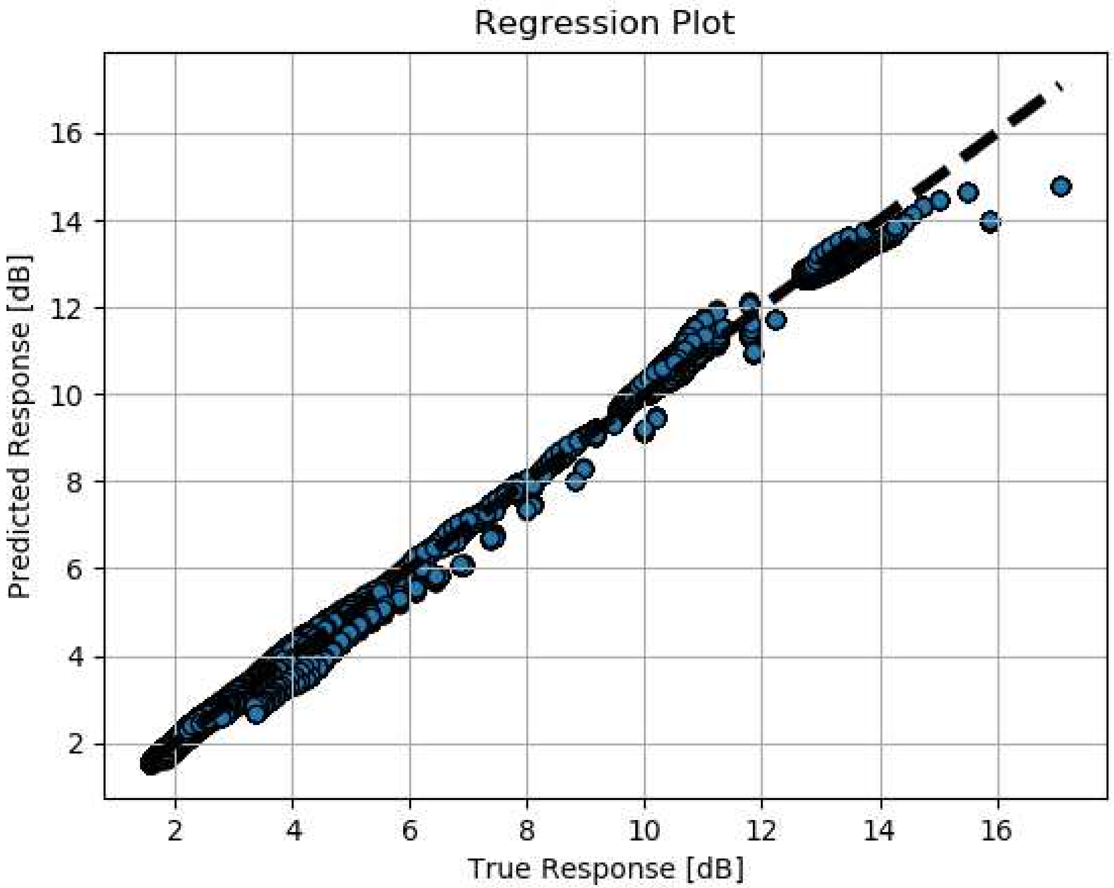


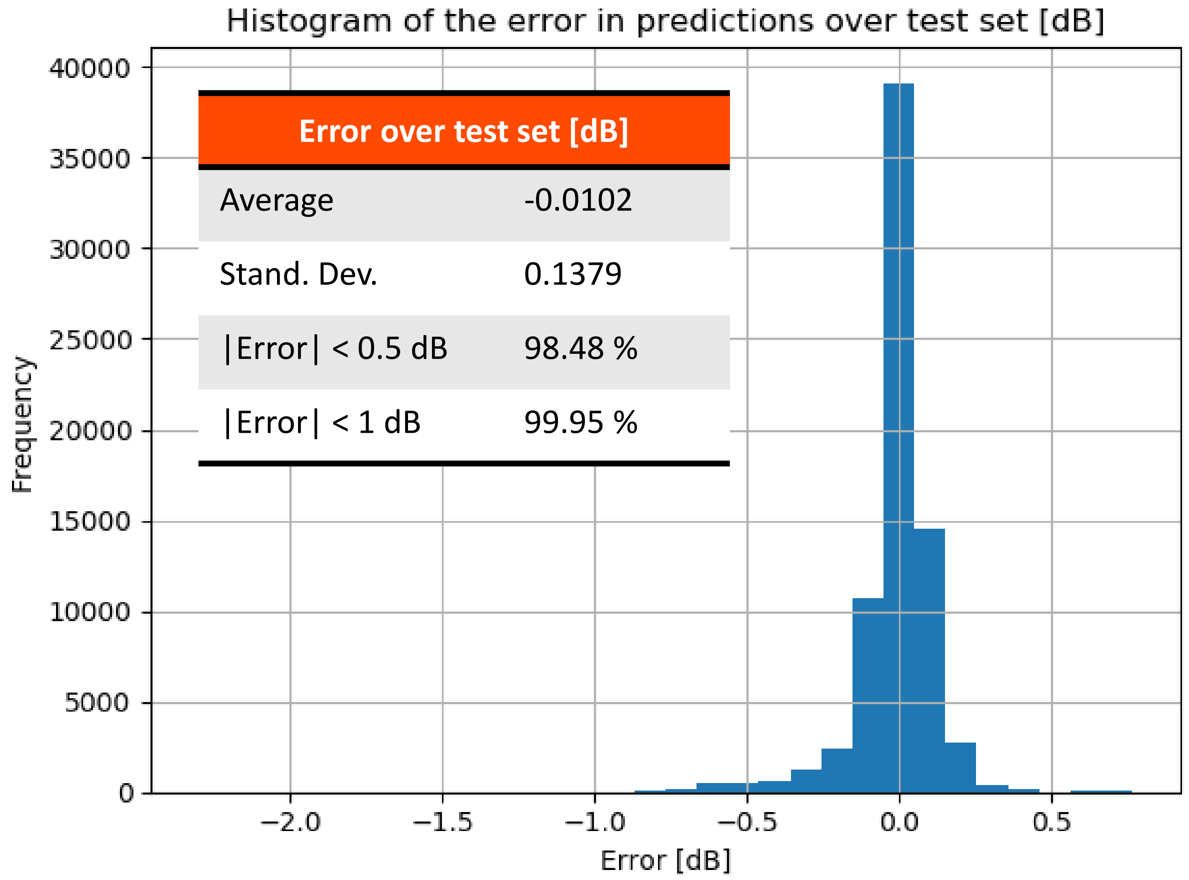


Infinera

(Some) Use Cases: Non-Linear Impairments Estimation

- NL OSNR Accuracy Results: Small training data set
  - ANN obtained a R^2 of 0.998479



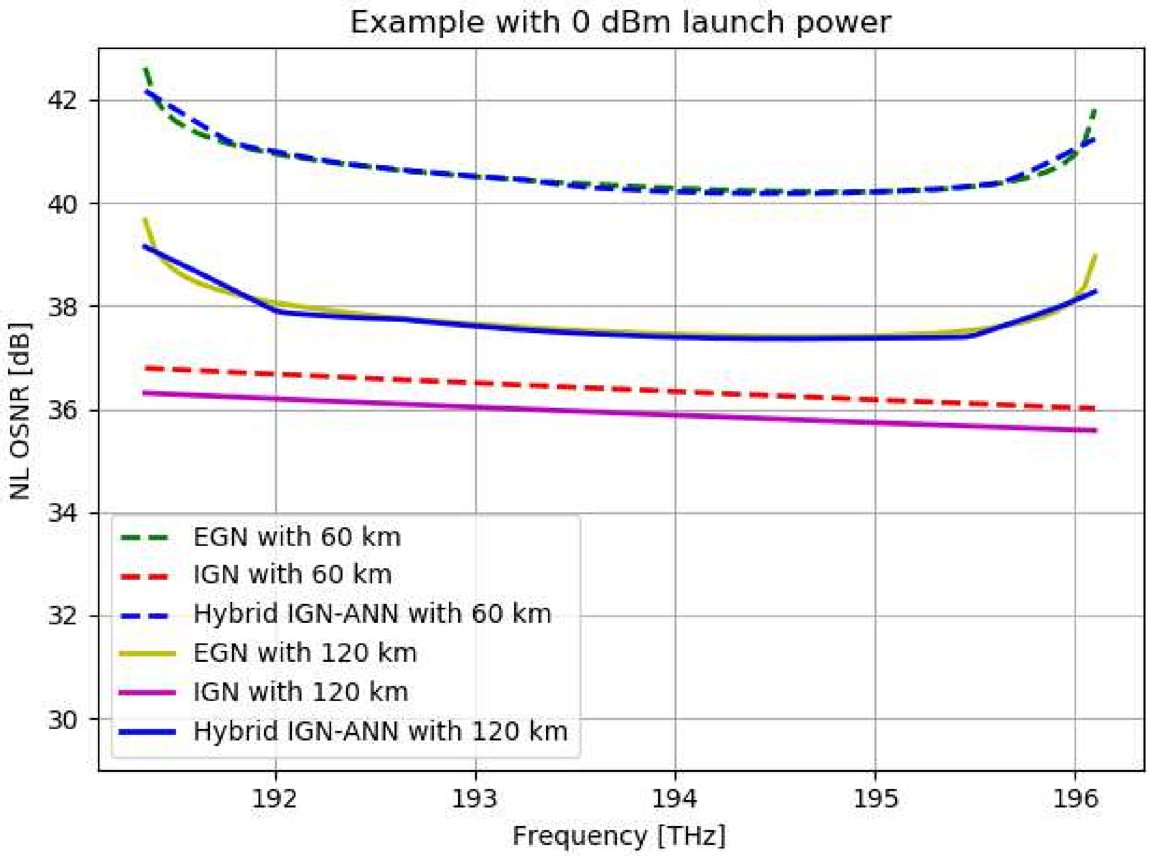


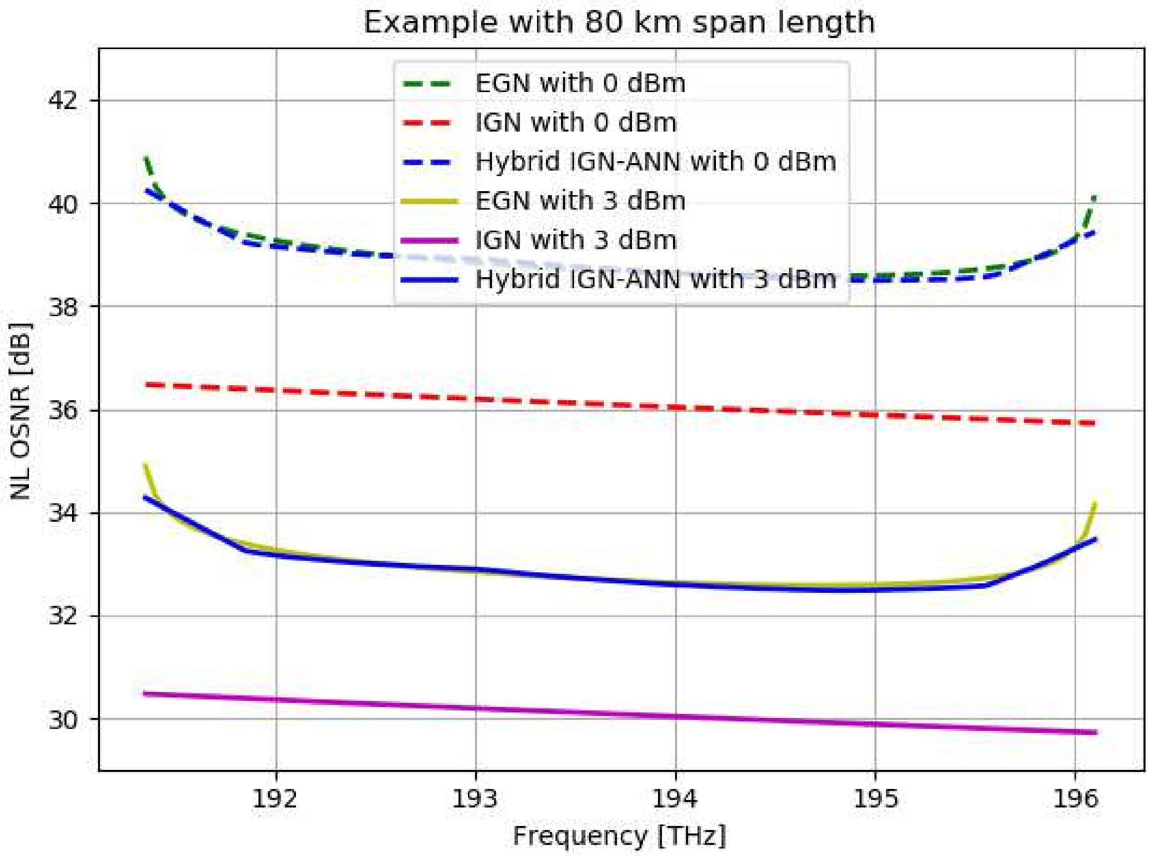
Error over test set [dB]	
Average	-0.0102
Stand. Dev.	0.1379
|Error| < 0.5 dB	98.48 %
|Error| < 1 dB	99.95 %

Infinera

(Some) Use Cases: Non-Linear Impairments Estimation

- NL OSNR Accuracy Results: Small training data set
  - Example of target and estimated NL OSNR as function of frequency for three span lengths (60, 80 and 120 km) and two launch powers (0 and 3 dBm)





(Some) Use Cases: Non-Linear Impairments Estimation

- Conclusions of Proof of Concept
  - Both **Standalone ANN** and **Hybrid iGN-ANN** show high accuracy

	Standalone ANN		Hybrid iGN-ANN	
	Large data set	**Smal data set**	**Large data set**	**Smal data set**
R^2	0.999685	0.999195	0.998394	0.998479
Average	0.0225	0.0174	-0.0001	-0.0102
Stand. Dev.	0.1244	0.2013	0.1421	0.1379
|Error| < 0.5 dB	98.93 %	95.96 %	97.14 %	98.48 %
|Error| < 1 dB	99.95 %	99.66 %	99.93 %	99.95 %

  - Both **Standalone ANN** and **Hybrid iGN-ANN** compute NL OSNR orders of magnitude faster than EGN

	EGN	Standalone ANN	Hybrid IGN-ANN
Normalized Execution Time	100	0.15	2.51

  - …but moving to a practical application requires training the ANN with more data (multi-span scenarios, different modulation formats, symbol rates, fiber types, different channel counts and spectrum profiles, etc.)

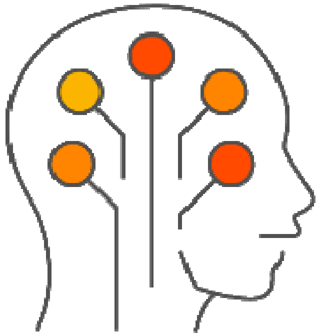
(Some) Use Cases: Non-Linear Impairments Estimation

- Extending the concept of Hybrid iGN-ANN
  - Detailed performance model can comprise
    - More accurate analytical models
    - Numerical simulations and/or
    - Measurements from a monitoring system
  - Externally defined boundary conditions can be applied to regulate the ANN contribution
    - Increase/decrease the weight of the ANN contribution according to level of confidence for specific input conditions
    - Selectively discard ANN contribution (e.g. if measurement data corrupted or not trustworthy)
- Physical impairments aware routing engine can be kept simple and fast

Infinera

(Some) Use Cases: Advanced Failure Prediction

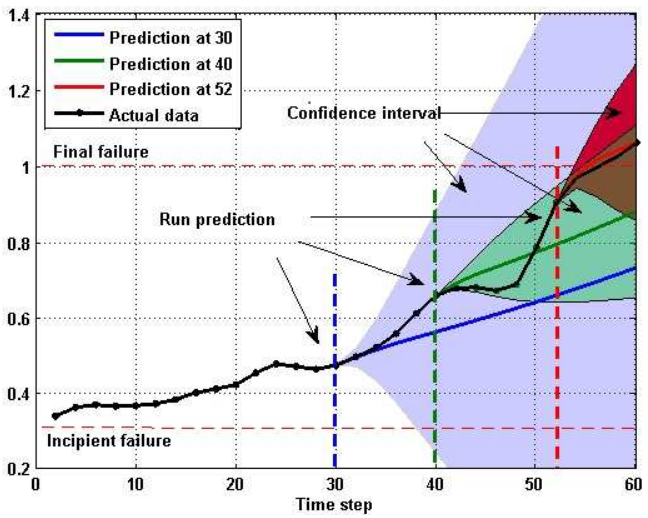
- How to operate optical networks as to improve metrics such as scalability and reliability while decreasing OpEx?
  - Automate
    - More decisions made and enforced by machines
    - Abstract complexity to facilitate the remaining human decisions
  - Act instead of react
    - Monitor the evolution of key parameters across the network
    - Continuously analyze trends and trigger preventive measures (e.g. to avoid failures)
    - Correlate data from different sources to further improve the process
  - Adapt
    - Factor in the specifics of each network deployment
    - Improve the overall process as more data is collected (i.e. learn from experience)



ML / AI

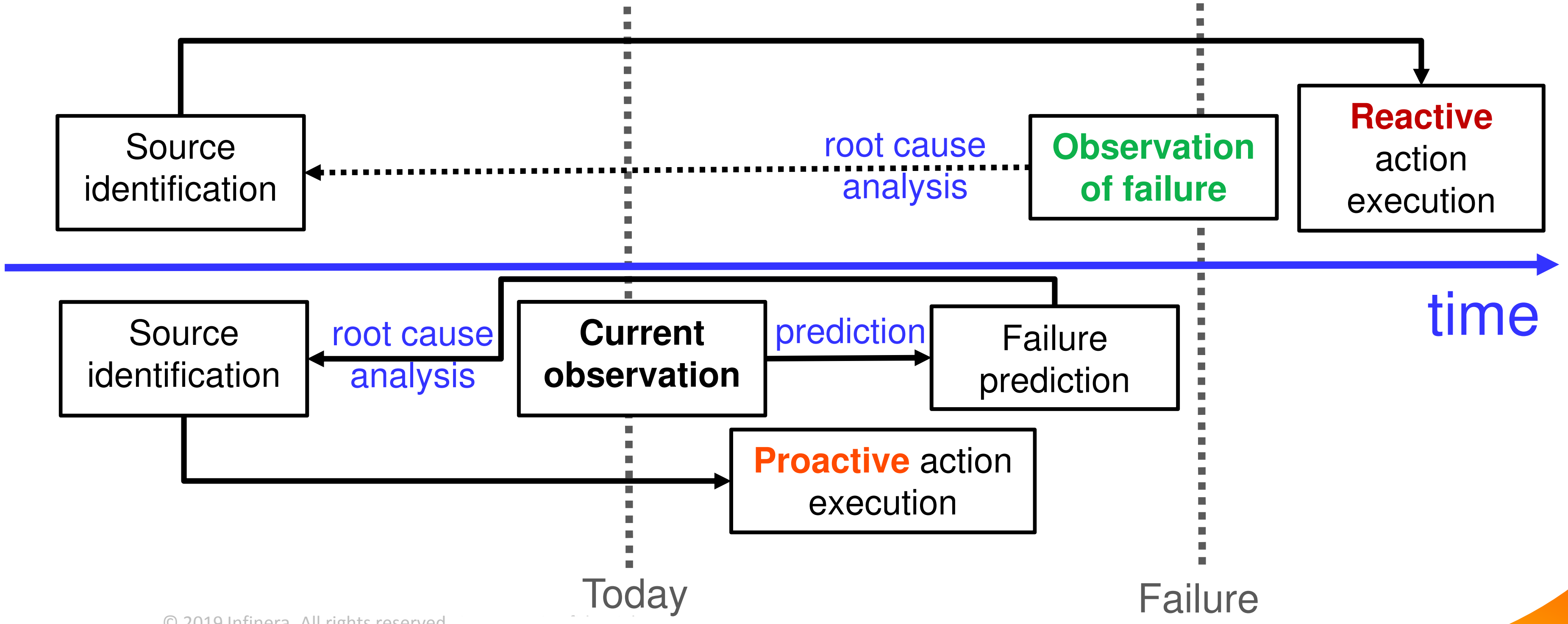
(Some) Use Cases: Advanced Failure Prediction

- Avoiding failures or mitigating their impact
  - Failures can occur as a result of processes occurring in different timescales
    - Slow degradation of devices and components (e.g. aging) should be possible to predict
    - Preventive replacement of devices / components (predictive maintenance) and proactive measures to circumvent them (self-healing) during network operation can enhance network reliability, reduce OpEx and improve customer experience
    - Sudden failures can be defined as failures whose predicted time to failure is smaller than the react time



Previous MoO

Future MoO



Conclusions

Infinera®

Conclusions

Is Machine Learning Needed* in Optical Communication Systems & Networks?

* i.e. does it add enough value for what it will cost?

- In short: “Yes…”
  - … if use case fits the stated conditions to successfully employ ML models (*data, data, and more data!*)
  - … if improvements are shown in key metrics and deployment cost is reasonable
  - … if moving to production environment can be done while meeting a range of requirements that can include fairness, robustness and assurance
- Looking ahead
  - ML applications expanded fast, time to consolidate and focus on fewer but high potential cases
  - We are just at the beginning: need to break barriers and work together to find ways to “standardize” data sets, securely share data and models, etc.

Thank You!

JPedro@Infinera.com