

WSC-2

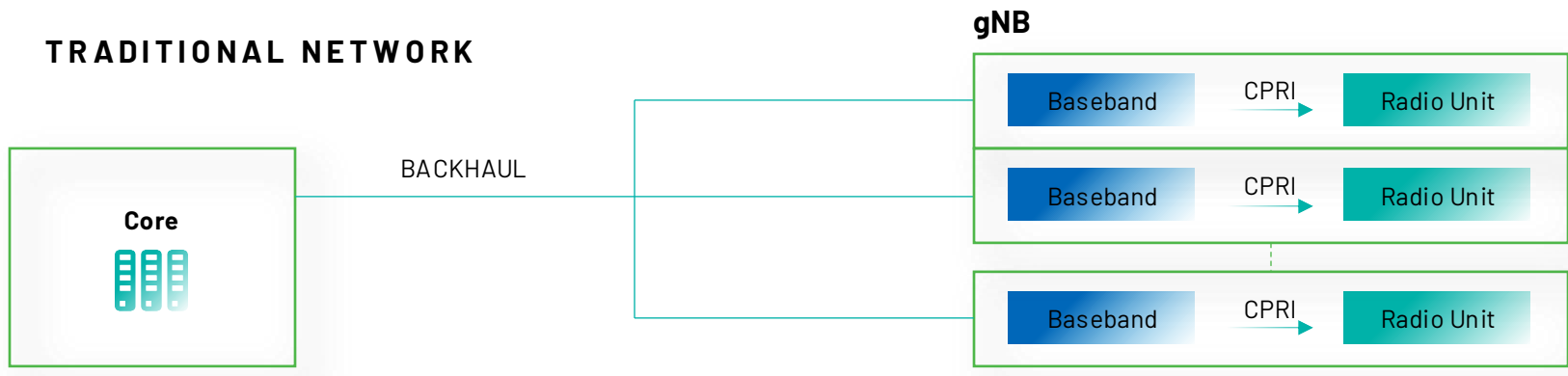
Physics-Informed Machine Learning-Based Digital Predistortion of RF Power Amplifiers

Tao Yu

Analog Devices Inc.

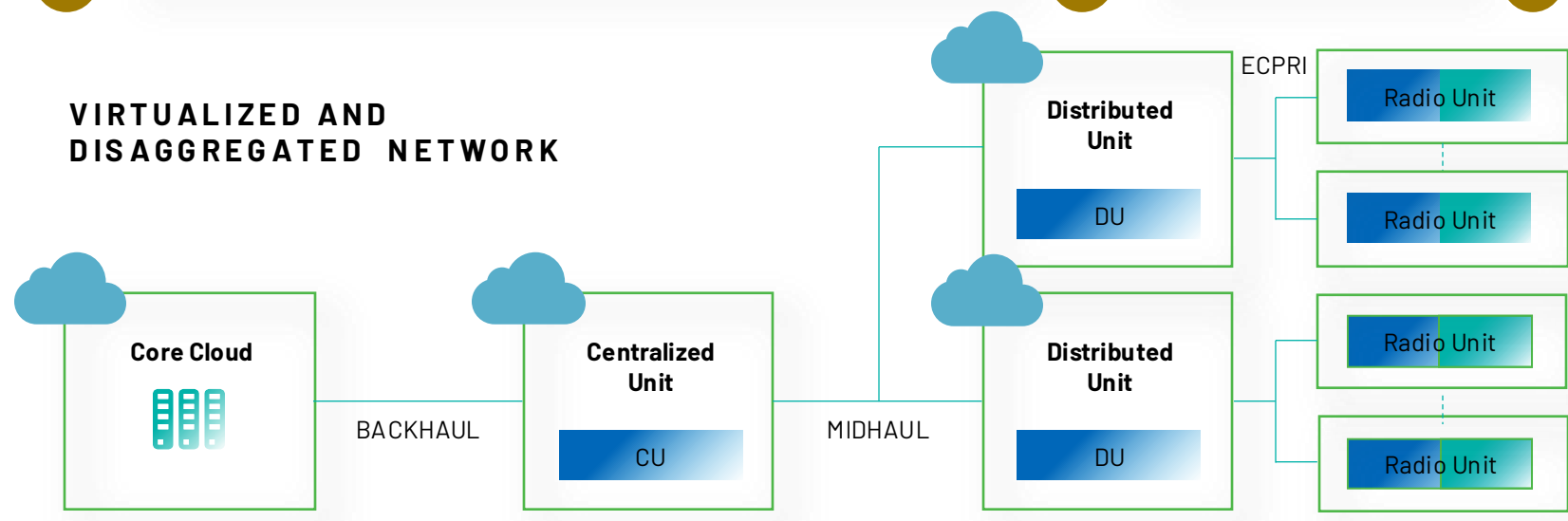
- Challenges in PA Linearization
- Machine Learning Techniques for PA Linearization
- Productization of ML-based Solutions
- Future Directions

TRADITIONAL NETWORK



- High use of Custom Silicon
- Distributed Capacity

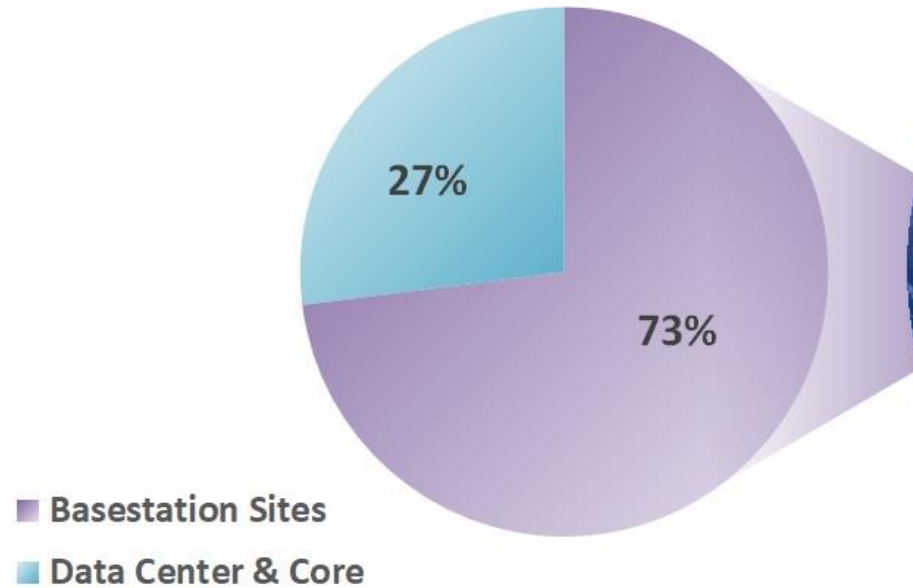
VIRTUALIZED AND DISAGGREGATED NETWORK



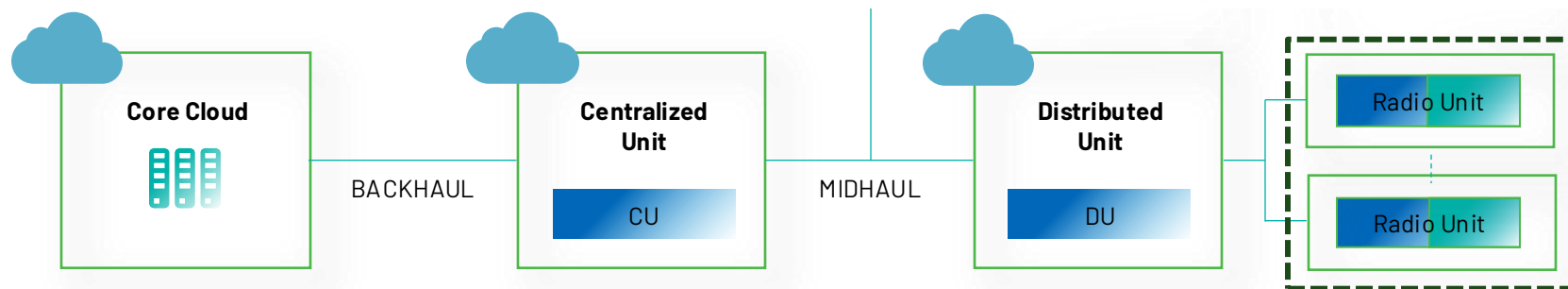
- Virtualized network running on COTS Servers
- Pooled Capacity
- Tight Coordination
- Software Defined

CPRI = Common Public Radio Interface
ECPRI = Enhanced CPRI

Challenges in Sustainability



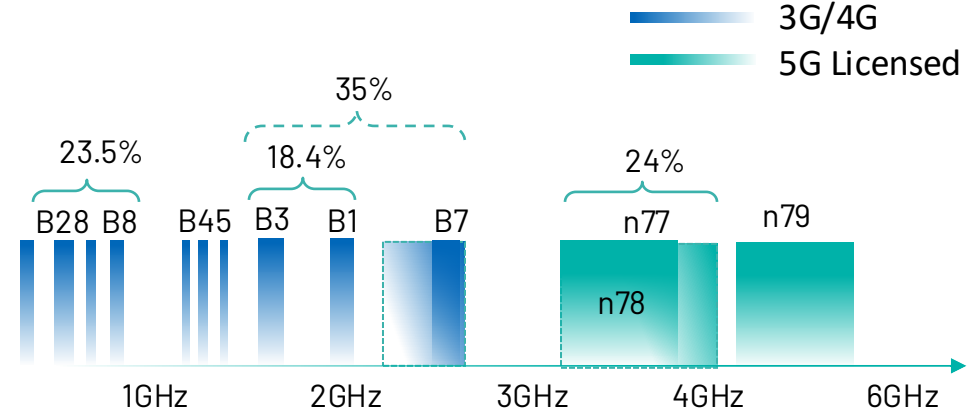
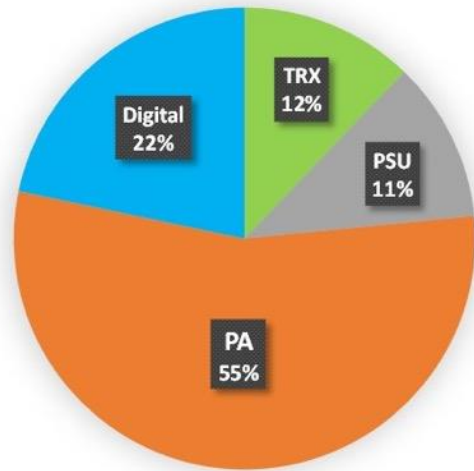
- Radio Unit of a Radio Access Network consumes the most energy
- 73% vs 27%
- Energy Savings are Imperative



CPRI = Common Public Radio Interface
ECPRI = Enhanced CPRI

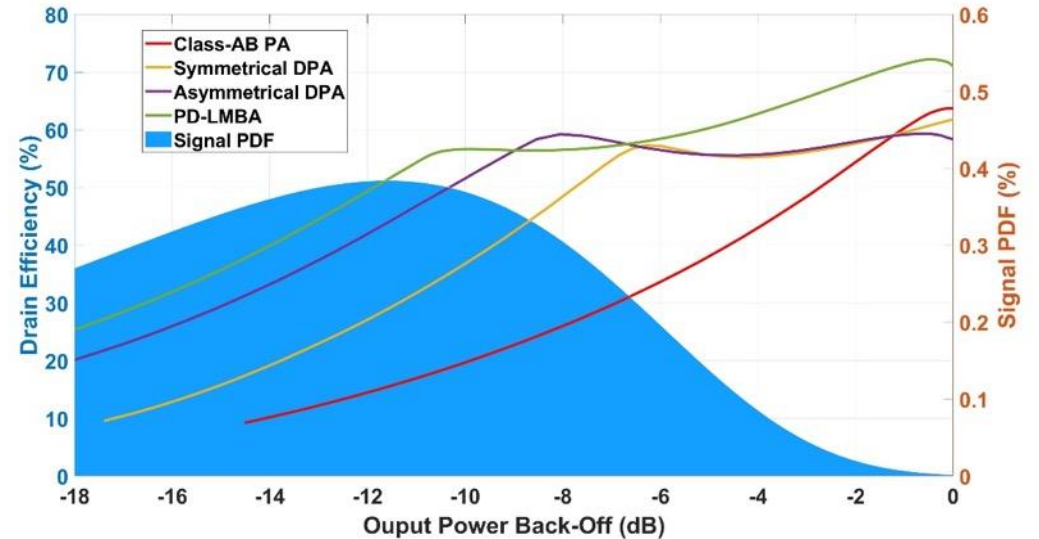
PA Efficiency

massive MIMO Key Component Power Contribution

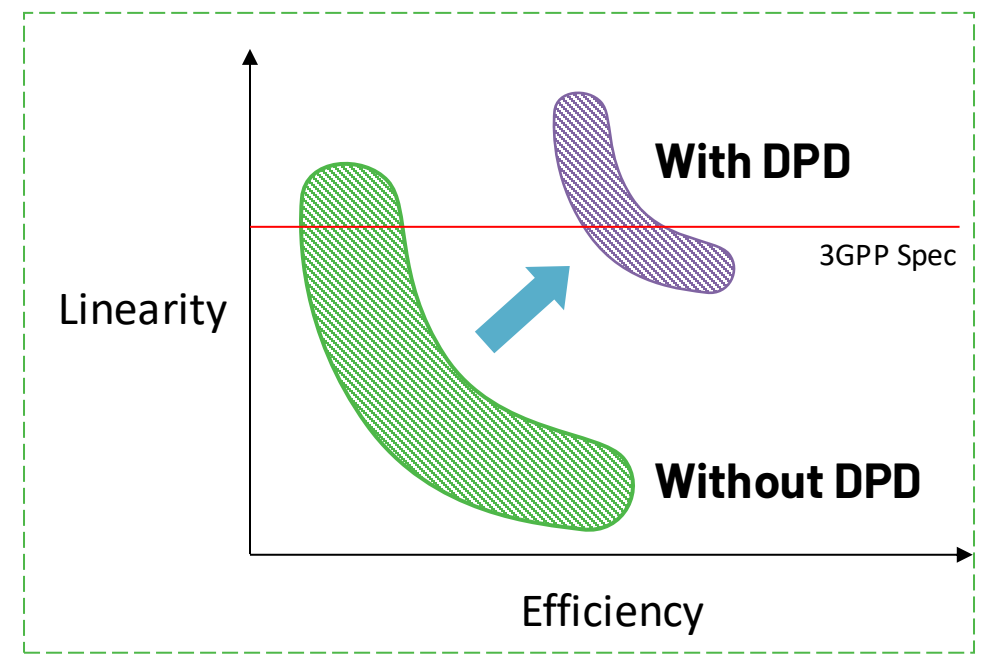
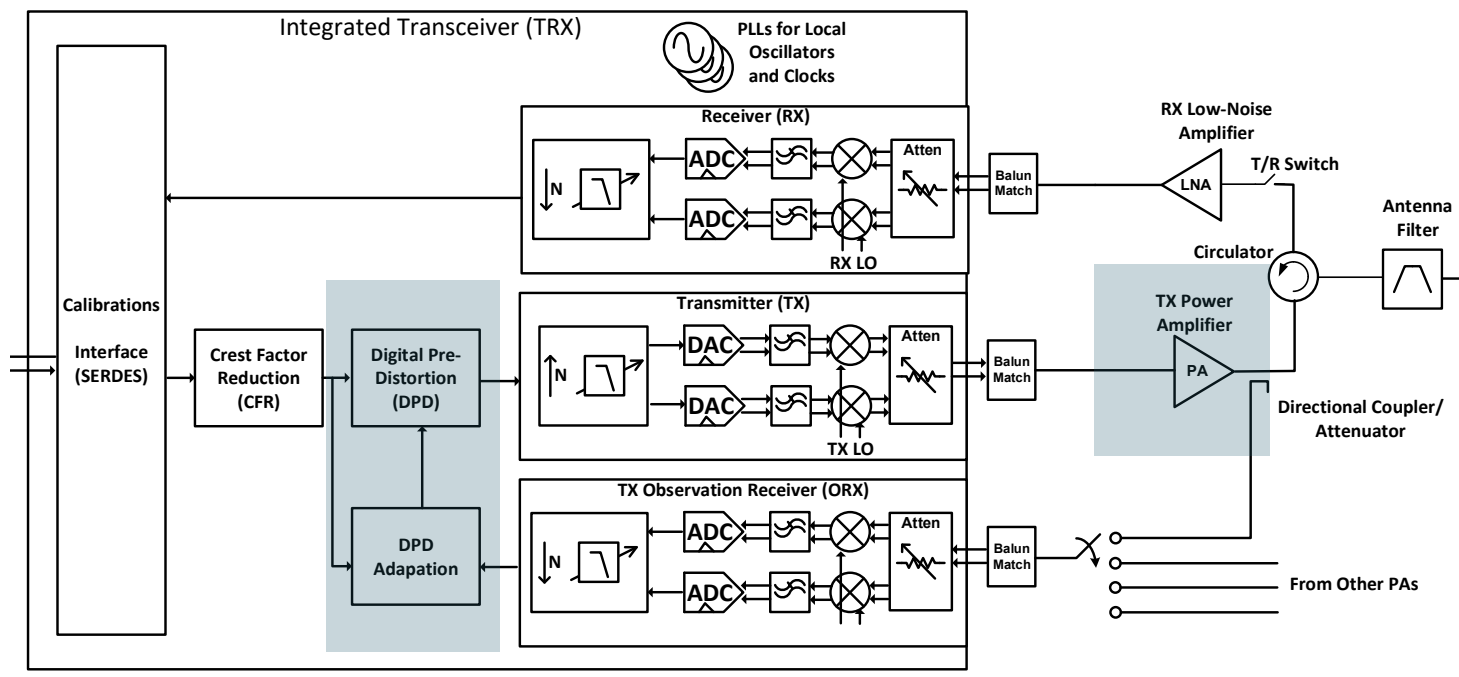


Better Spectrum Efficiency, More Bandwidth, More Efficiency

- Typical DC Power Consumption: PA > 500W!
 - 64Tx Transmit Power: 160 EAC; 2W/TX
 - 3.16W/PA – PA Eff. =40% (1.5dB PA post loss/TX)

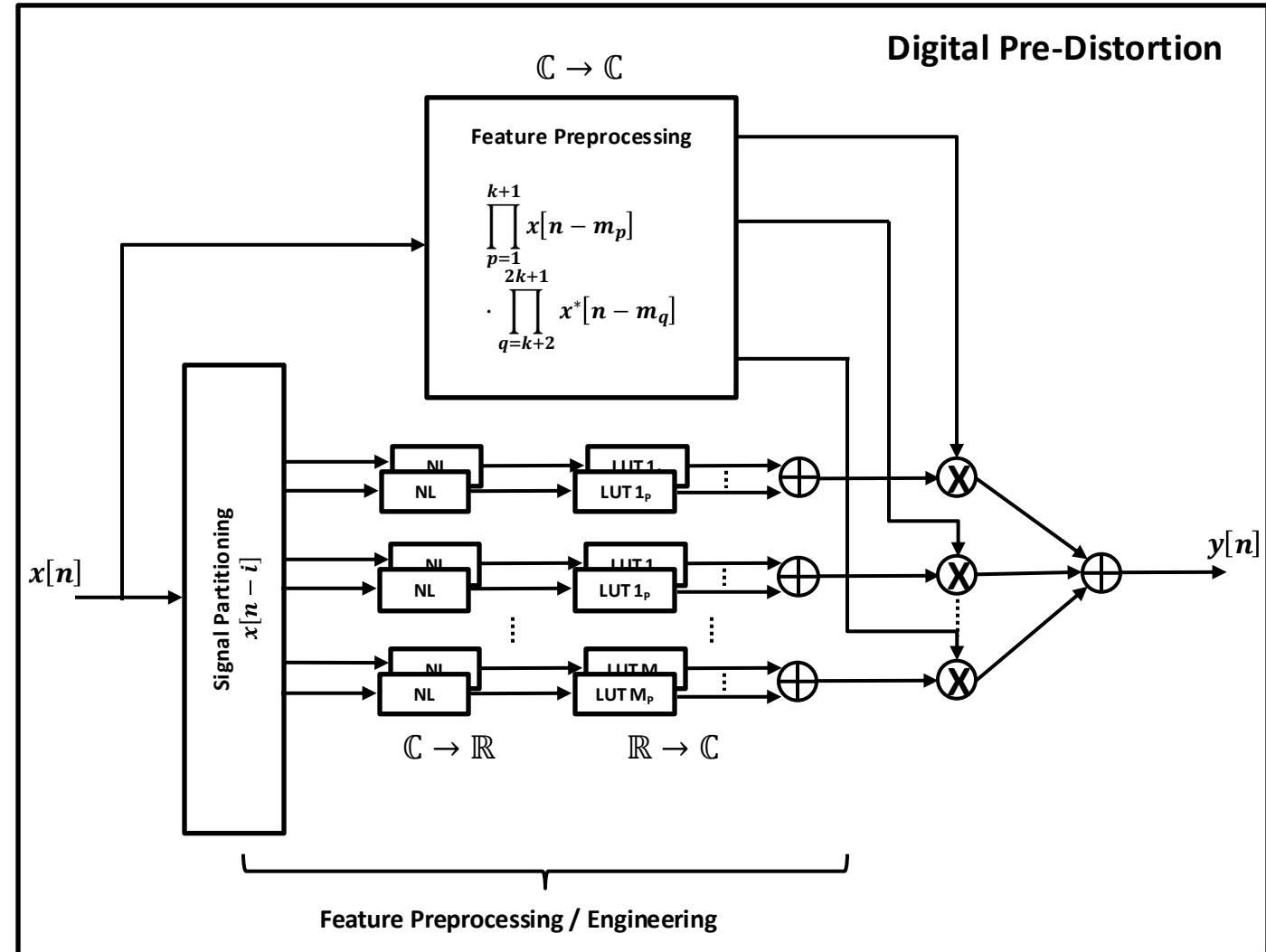
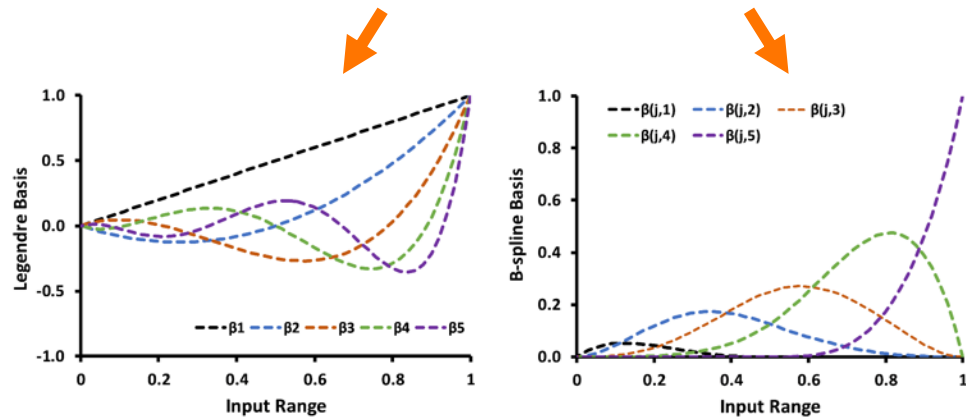


Digital Predistortion (DPD) Improves PA Linearity and Efficiency



5G driving up DPD complexity exponentially!

- Prior arts focus on feature engineering as depicted by
 - Top transformative path
 - $\mathbb{C} \rightarrow \mathbb{C}$
 - Bottom transformative path
 - $\mathbb{C} \rightarrow \mathbb{R}$ followed by $\mathbb{R} \rightarrow \mathbb{C}$
- LUT typically implements piecewise linear basis functions
 - E.g., orthogonal polynomials, B-splines





$$F(x) = \sum_{j=1}^J \alpha_j \beta_j(x)$$

Coefficient vector

Basis function
Each $j \mapsto (K, M, Q)$
 K = non-linear order
 M = memory
 Q = cross-term memory

Solution to coefficients:

$$\min_{\alpha} \left| y_n - \sum_{j=1}^J \alpha_j \beta_j(x) \right|^2 + \lambda L(\alpha)$$

$$\sum_{j=1}^J |\alpha_j|^2, \text{ L2 - norm}$$

$$\sum_{j=1}^J |\alpha_j|, \text{ L1 - norm}$$

Memory polynomial (MP) :

$$\sum_m^M \sum_k^K \alpha_{m,k} x(n-m) |x(n-m)|^k$$

Generalized Memory Polynomial (GMP) :

$$\sum_m^M \sum_k^K \sum_q^Q \alpha_{m,k,q} x(n-q) |x(n-m)|^k$$

Dynamic Deviation Reduction (DDR) :

$$\sum_k^K \sum_q^Q \alpha_{k,q} P_k(|x(n)|^k) x(n-q) +$$

$$\sum_k^K \sum_q^Q \alpha_{k,q} P_k(|x(n)|^k) x^2(n) x^*(n-q) +$$

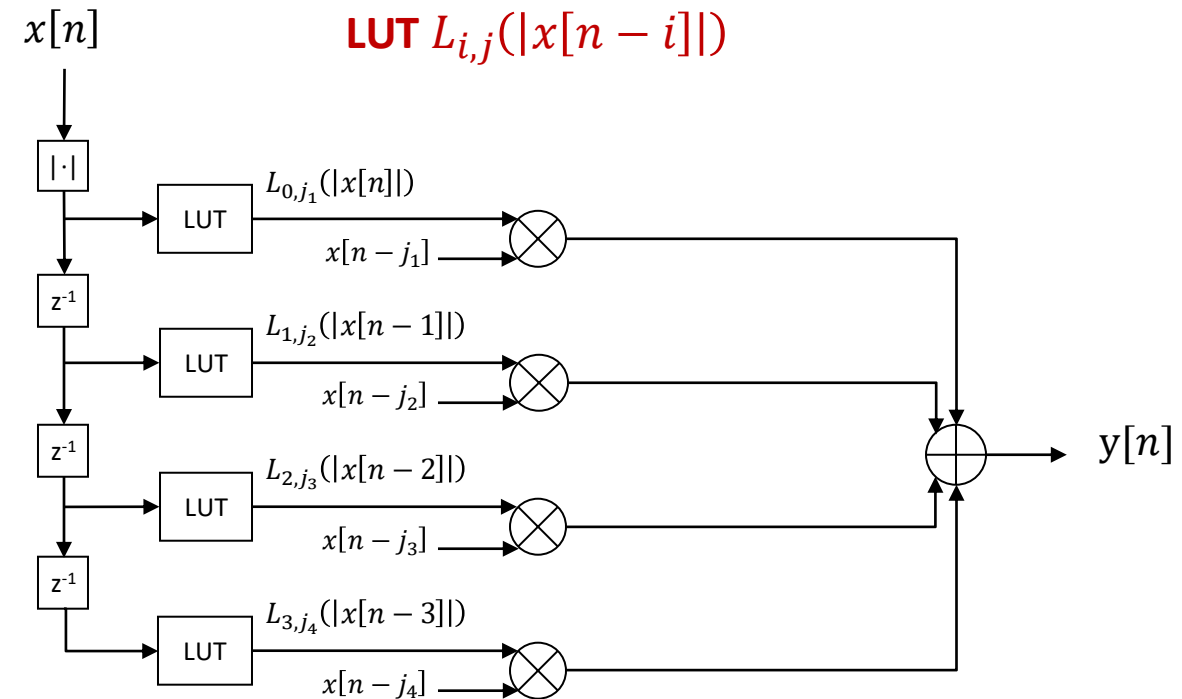
$$\sum_k^K \sum_q^Q \alpha_{k,q} P_k(|x(n)|^k) x(n) |x(n-q)|^2 + \dots$$

- DPD Model Architecture Discovery
- Neural-Network-assisted Physical Modeling for DPD
- Hybrid Model with Volterra-based Actuator and Neural Network

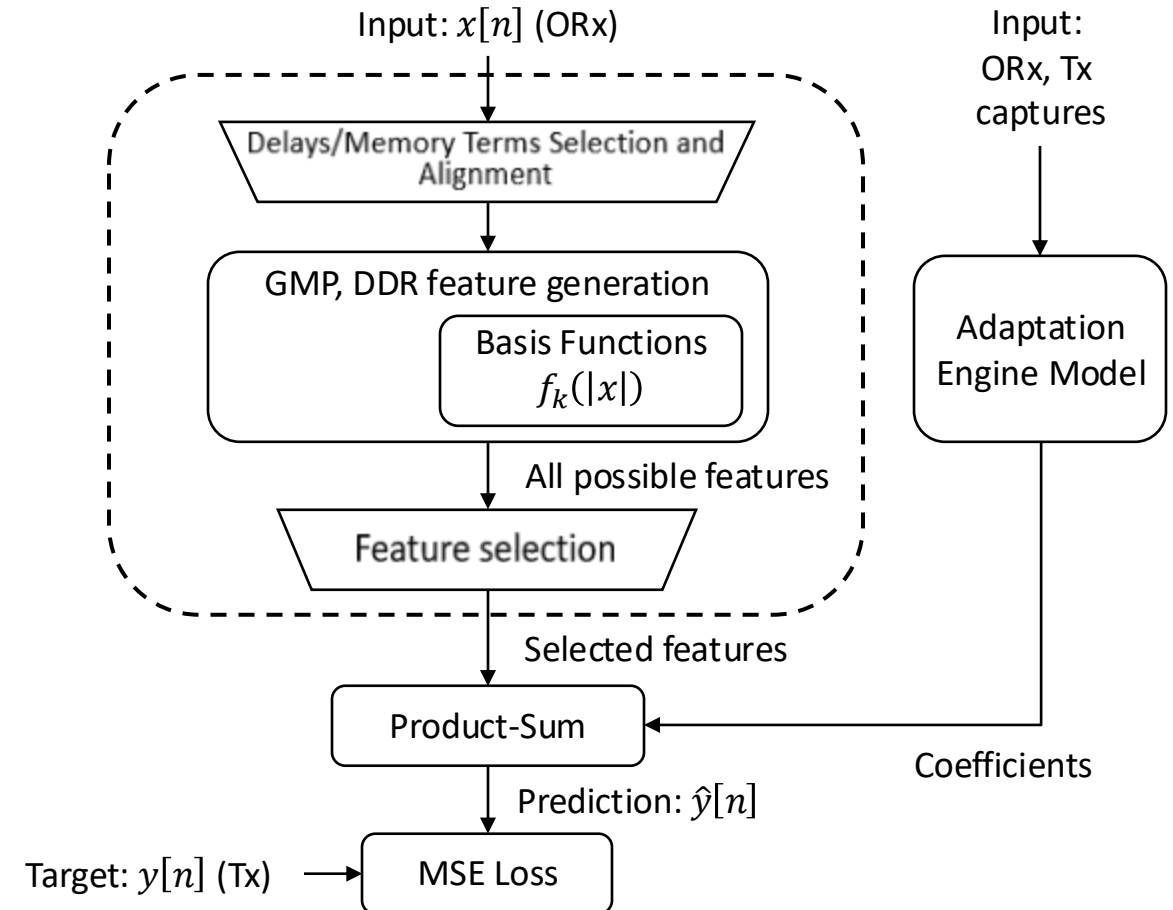
- Many parameters to be co-optimized
 - LUT and feature selection (search space size of $\sim 2^{400}$)
 - Global alignment delay
 - LUT basis function shape (f_k)
 - Model accuracy/size tradeoff
- Current optimization method:
 - Use Legendre basis functions
 - Select all diagonal IJ terms
 - Randomly select off-diagonal IJ terms
 - Check performance on a single target waveform
- **How can we best utilize the full capability of the DPD hardware?**

Hardware actuator implements:

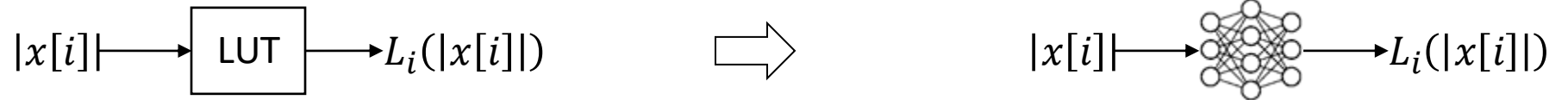
$$y[n] = \sum_{i,j} \underbrace{\sum_k c_{ijk} f_k(|x[n-i]|)}_{\text{LUT } L_{i,j}(|x[n-i]|)} x[n-j]$$



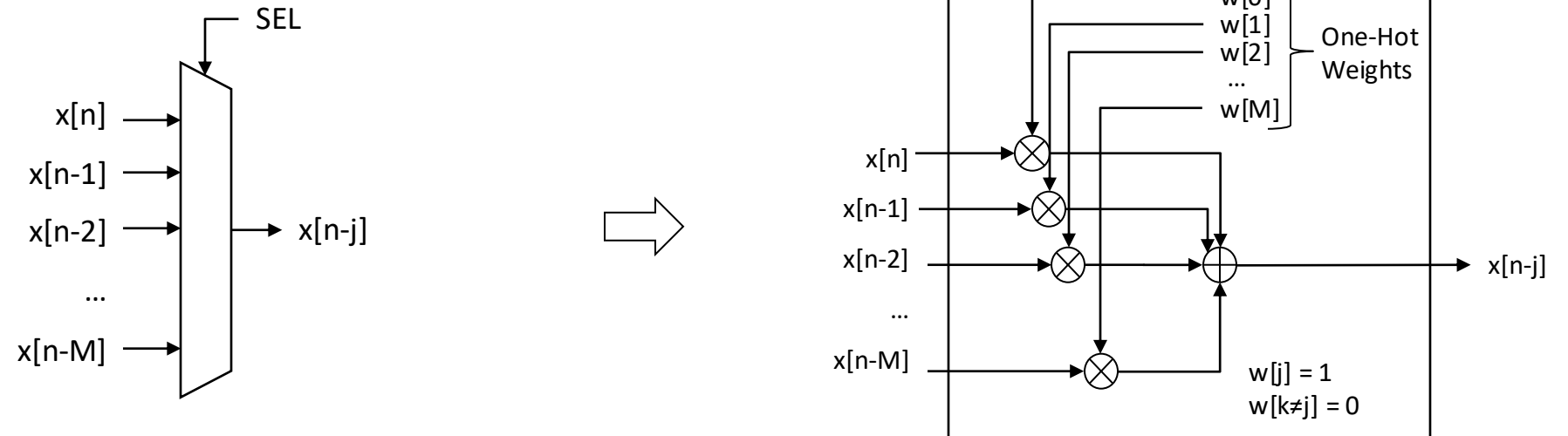
- Represent hardware engine operations as a computational graph
- Make each trainable operation differentiable and parameterized
 - Dense neural network for basis functions
- Training
 - Training data includes many waveform captures
 - Inverse or direct mode
 - MSE or MAE loss function
- Use backpropagation to optimize parameters
 - **All parameters jointly optimized over full model space and all waveforms of interest**



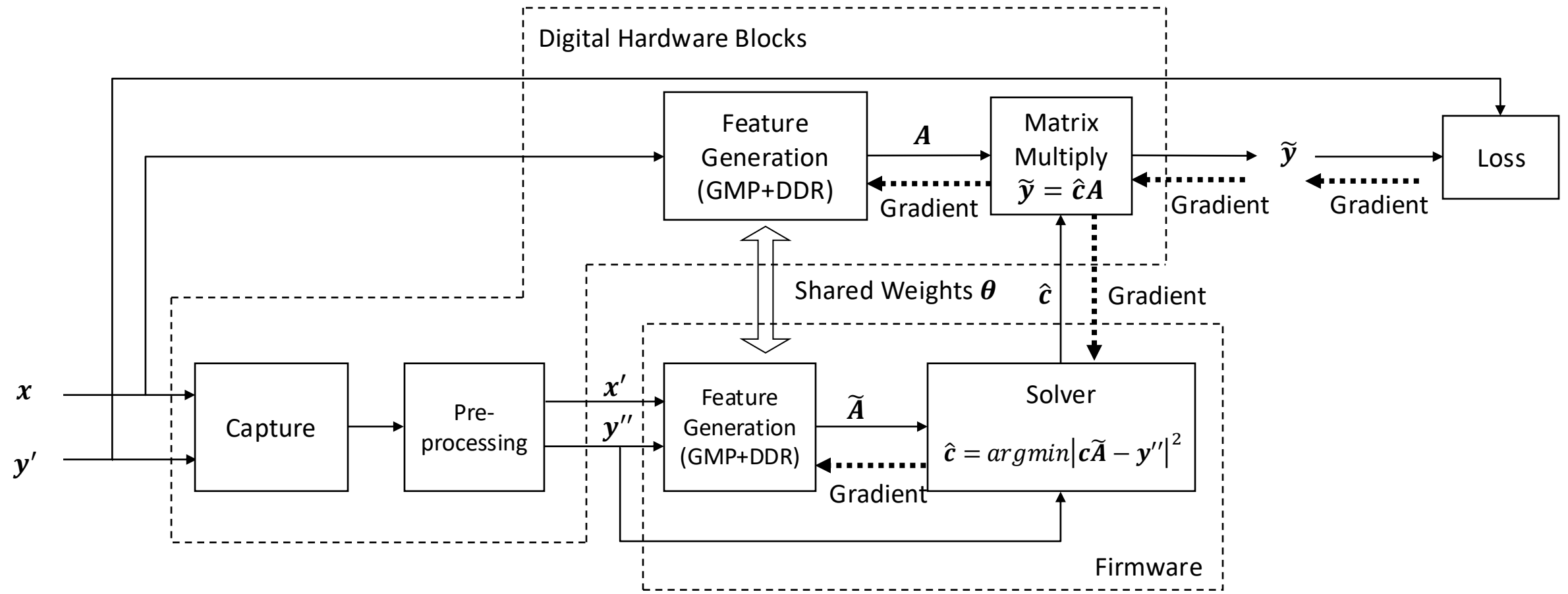
Single-input Mapping



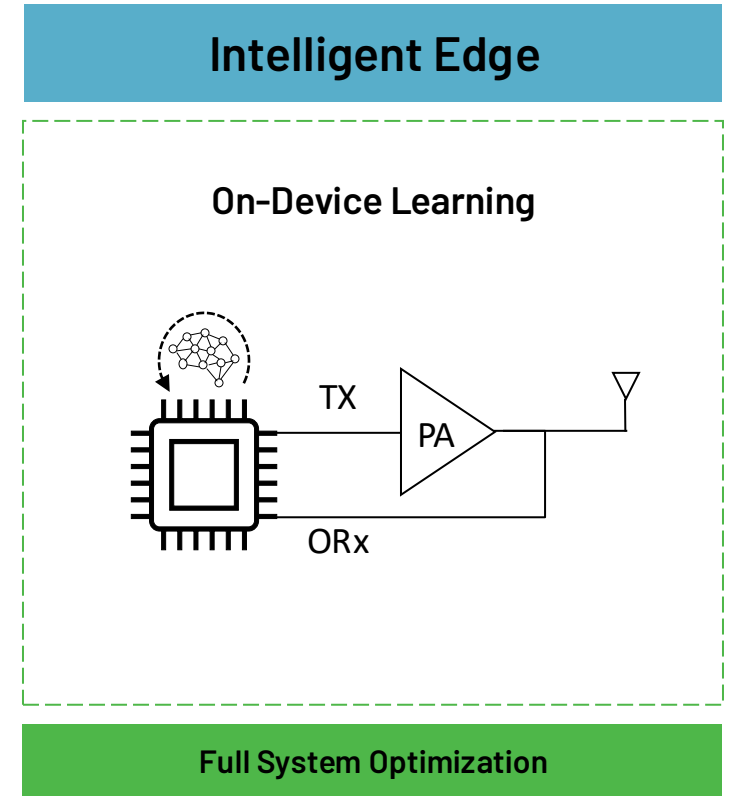
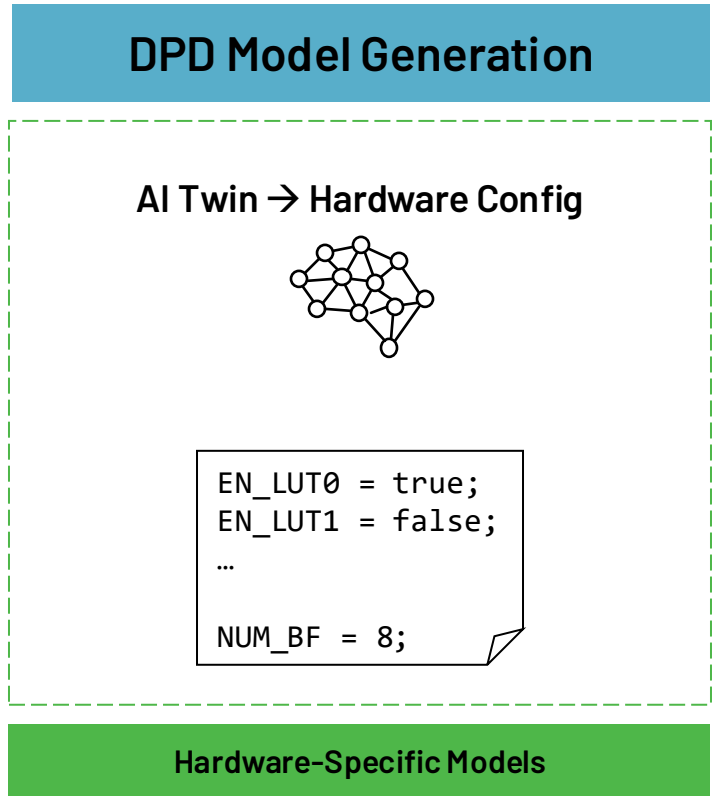
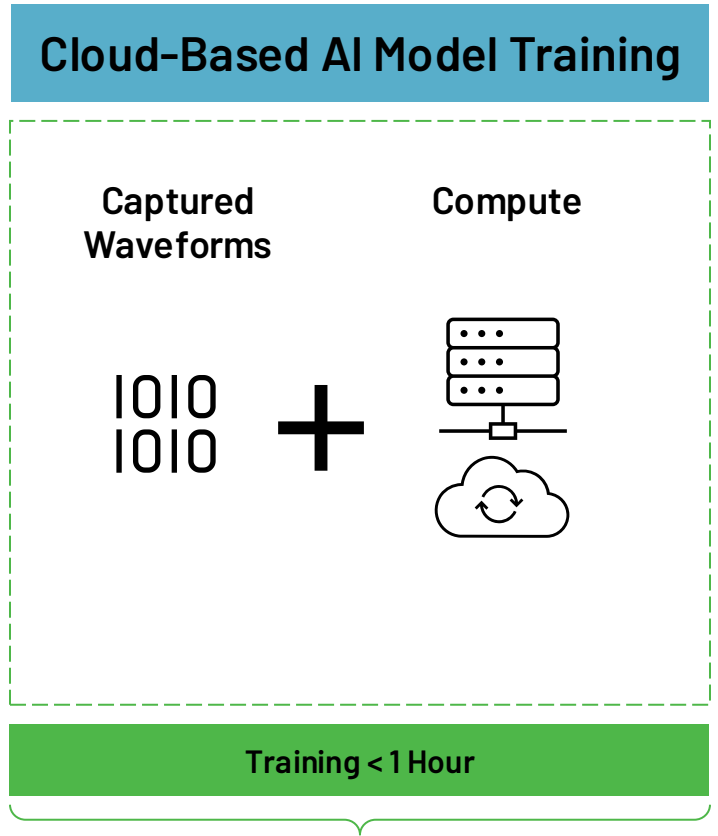
Muxing / Selection



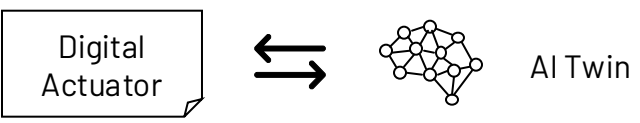
Gradient-Aware Adaptation Engine



DPD Wizard

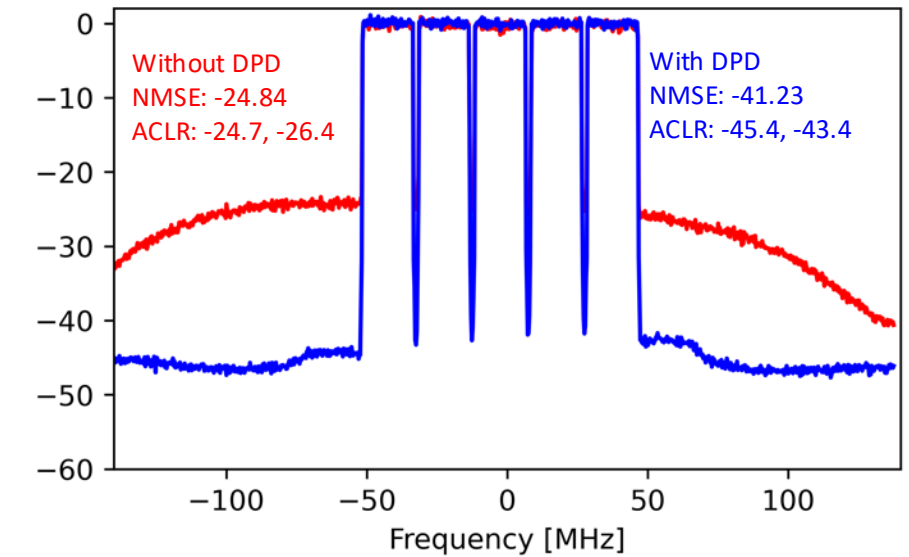
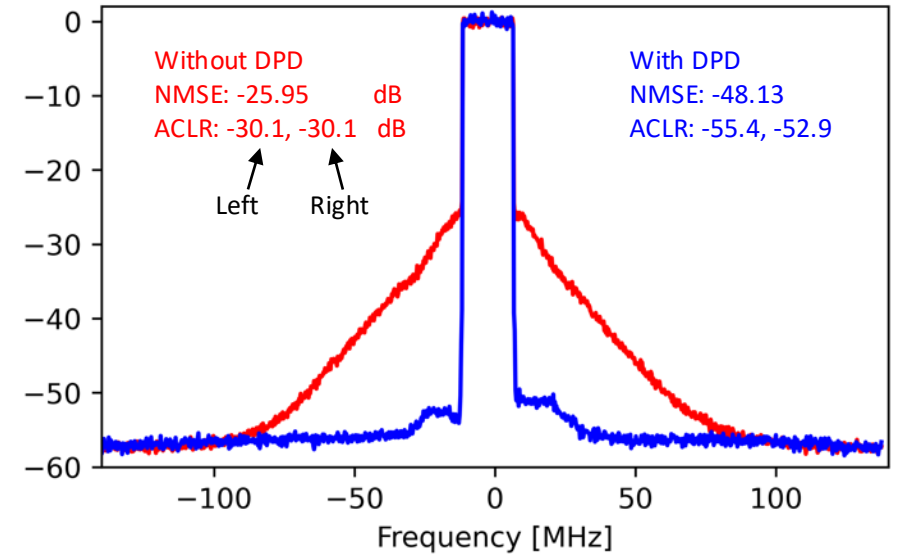
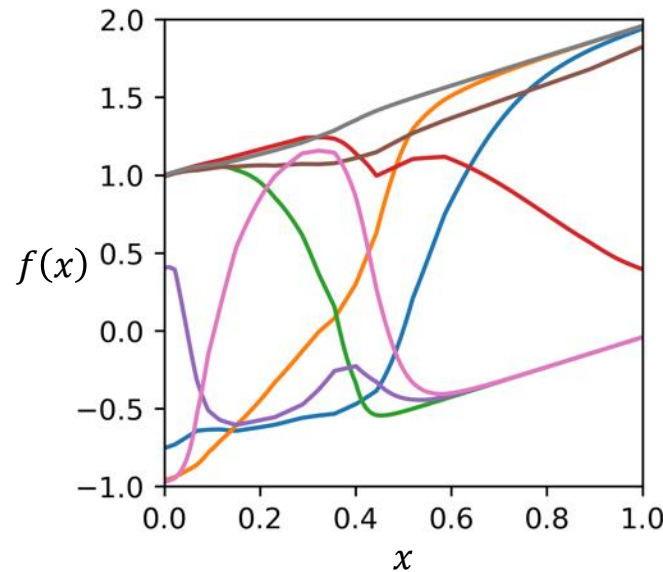
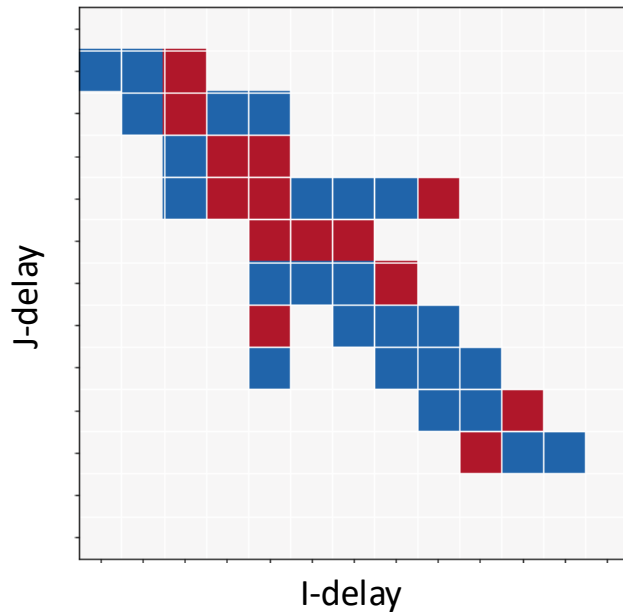


Cover All Waveforms of Interests



Reduce Model Optimization from Days to Hours

- Small DPD model on QPA2705
 - Strong regularization resulting in 14 LUTs + 39 Terms
 - Good DPD performance confirmed with spectrum analyzer
 - Significant activity of off-diagonal memory terms
 - Nonlinear behavior of basis function around RMS magnitude of envelope



Wolfspeed 1A2639 PA (ADI Gen5 DPD)

- Conventional search: - 50.3 dBc ACLR
- ML Discovered: - 51.1 dBc ACLR

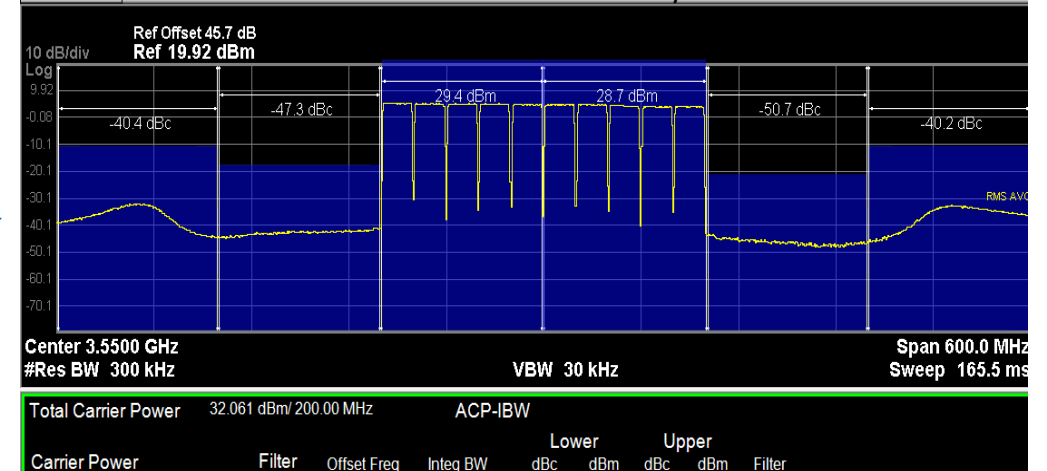
Skyworks 66520 (ADI Gen4 DPD)

- Target Performance - 48 dBc ACLR
- Conventional search: - 47.3 dBc ACLR
- ML Discovered: - 48.1 dBc ACLR

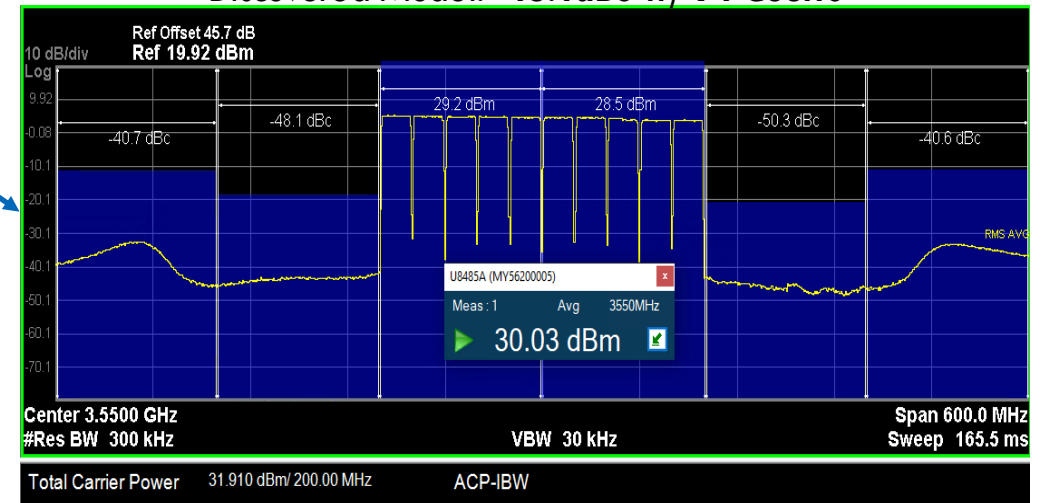
Key advantages:

- Models discovered with ML consistently outperform conventional search
- Reduced power consumption due to smaller NAS models
- Models for different hardware developed with the same building blocks and workflow
- Streamlined data capture and model search allow for quick turnaround of highly optimized models

Handcrafted Model: **-47.3dBc w/ 190 Coeffs**

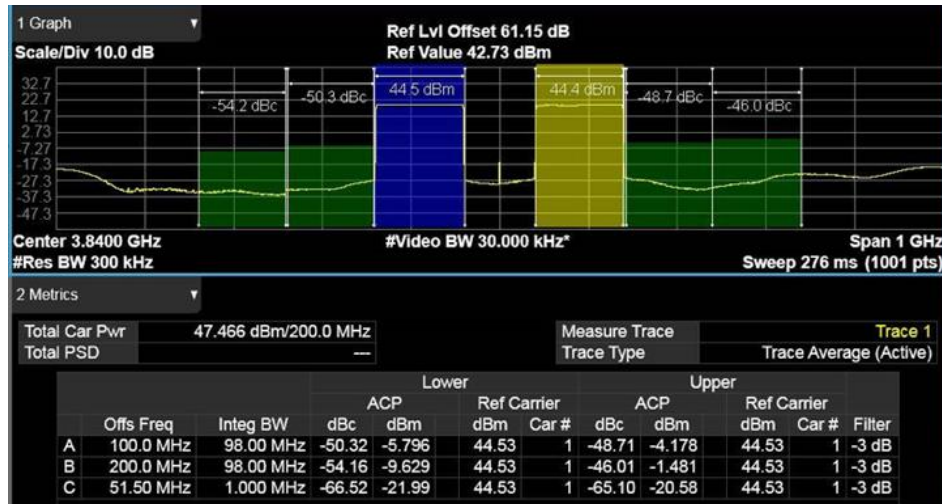
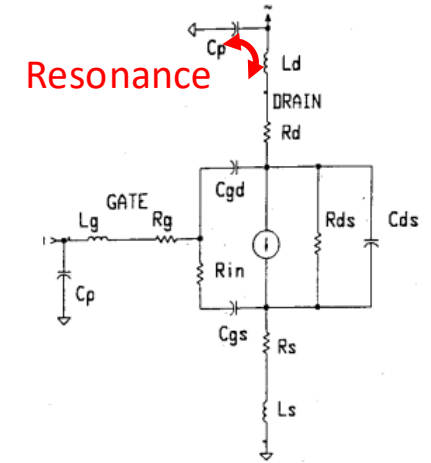


Discovered Model: **-48.1dBc w/ 94 Coeffs**

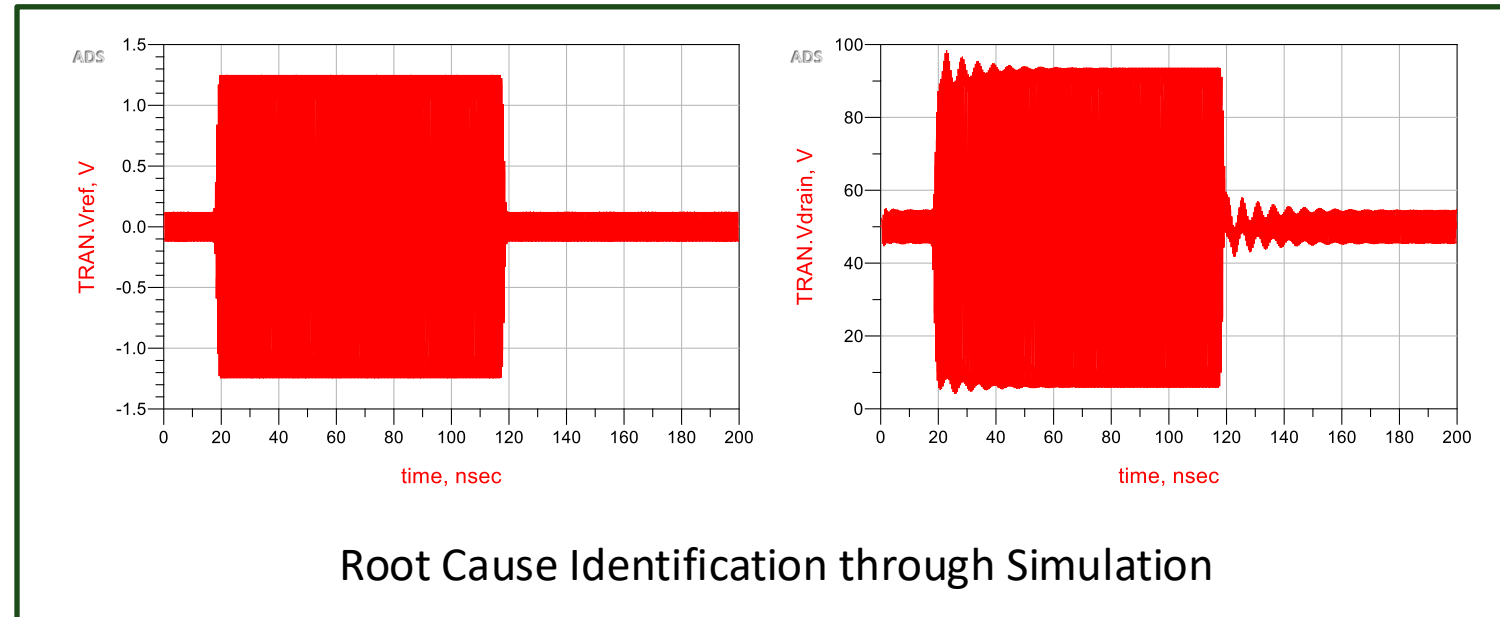


- DPD Model Architecture Discovery
- Neural-Network-assisted Physical Modeling for DPD
- Hybrid Model with Volterra-based Actuator and Neural Network

- Unique challenge in high power (40W) PAs
 - Resonance of drain voltage exhibit medium-term (100ns – 1us) memory effects
 - FCC SEM limit cannot be satisfied on a 40W final stage
 - -13 dBm/MHz total emission → -21.5 dBm/MHz for 4T4R with margin



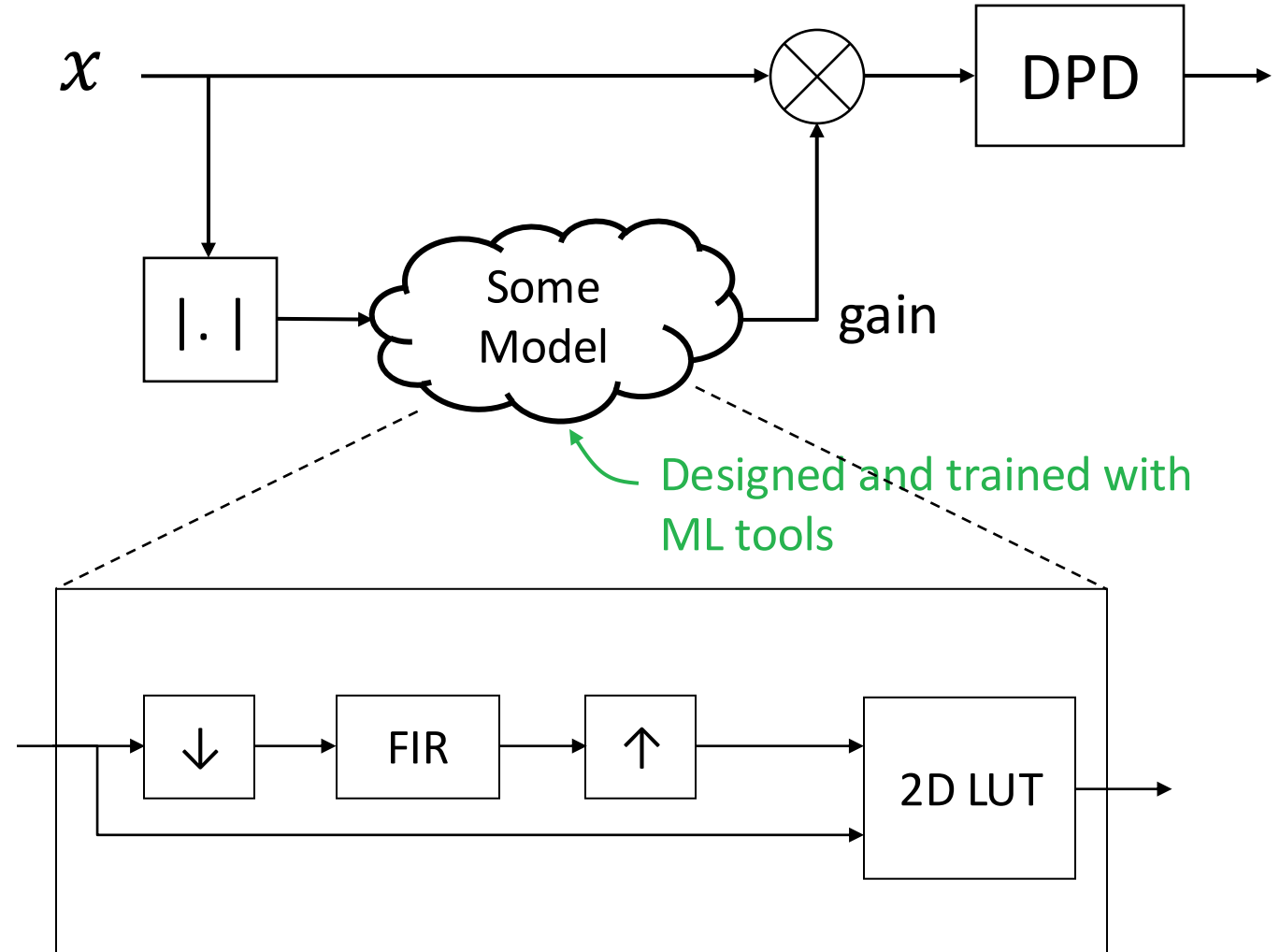
Achieved Spectral Emission: -20.58 dBm/MHz



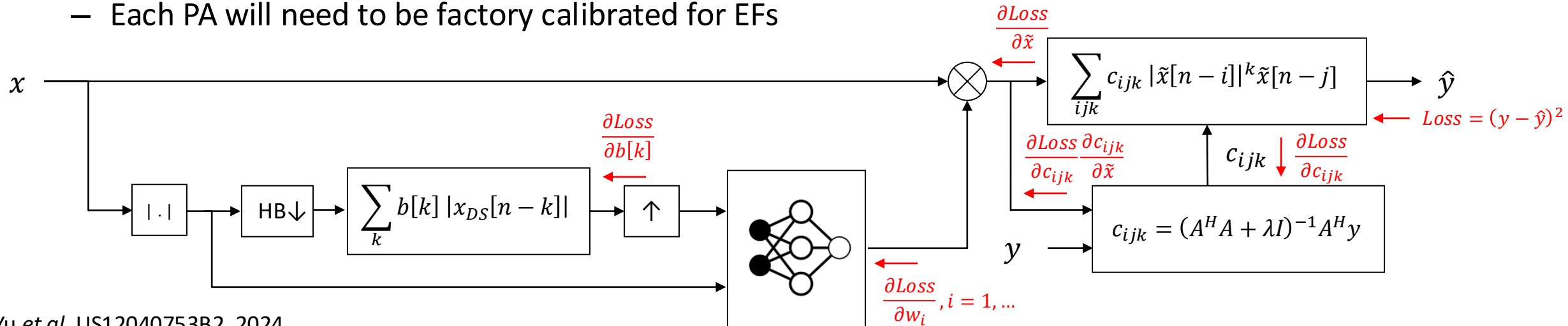
Root Cause Identification through Simulation

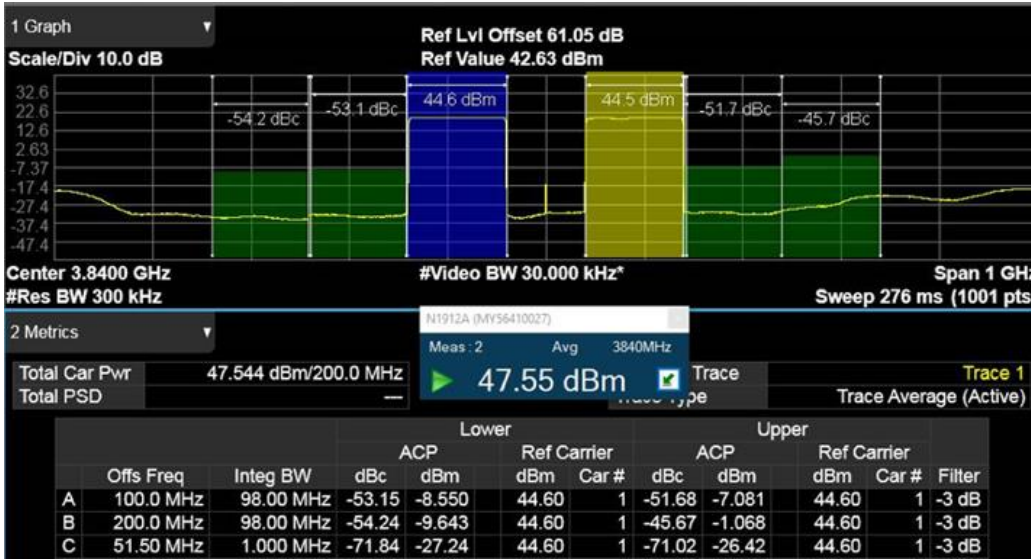
Medium-Term DPD (MT-DPD)

- Physical understanding: waveform envelope induced drain voltage variation over long duration
- Drain voltage resonance model: filter response (FIR)
- Unknown quantity: resonance to complex gain mapping
 - NN to learn the mapping and FIR filter coefficients together

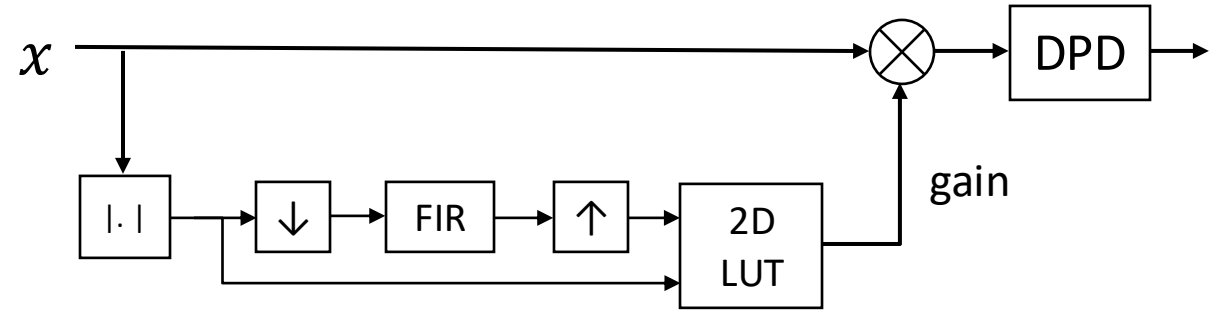
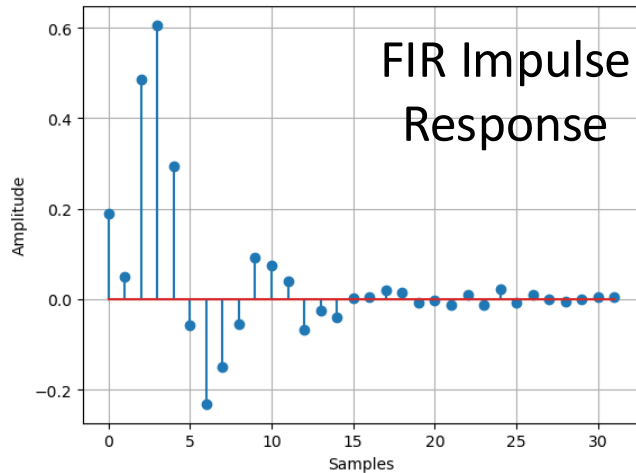


- Required Data
 - A single TX-ORx pair on widest BW at full power (TM3) waveform: Open loop
- Optimization Tool
 - PyTorch (ML framework)
 - Automatic differentiation to backpropagate modeling error (through LS fitted DPD model) to compute gradients on learnable parameters (NN weights & FIR coefficients)
 - SGD-based optimizer (ADAM) to perform mini-batch updates over captured dataset
 - Each PA will need to be factory calibrated for EFs

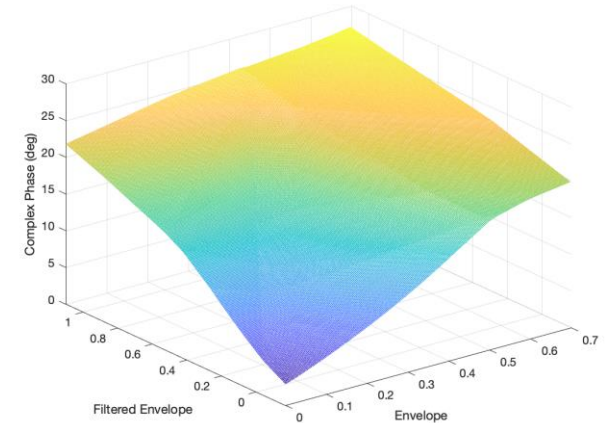
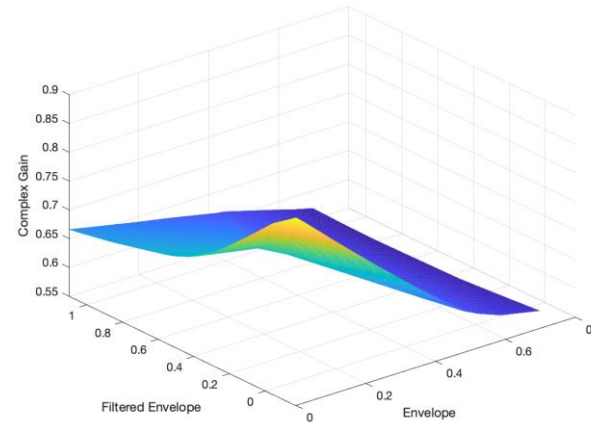




OBUE: -20.58 dBm/MHz \rightarrow -26.42 dBm/MHz

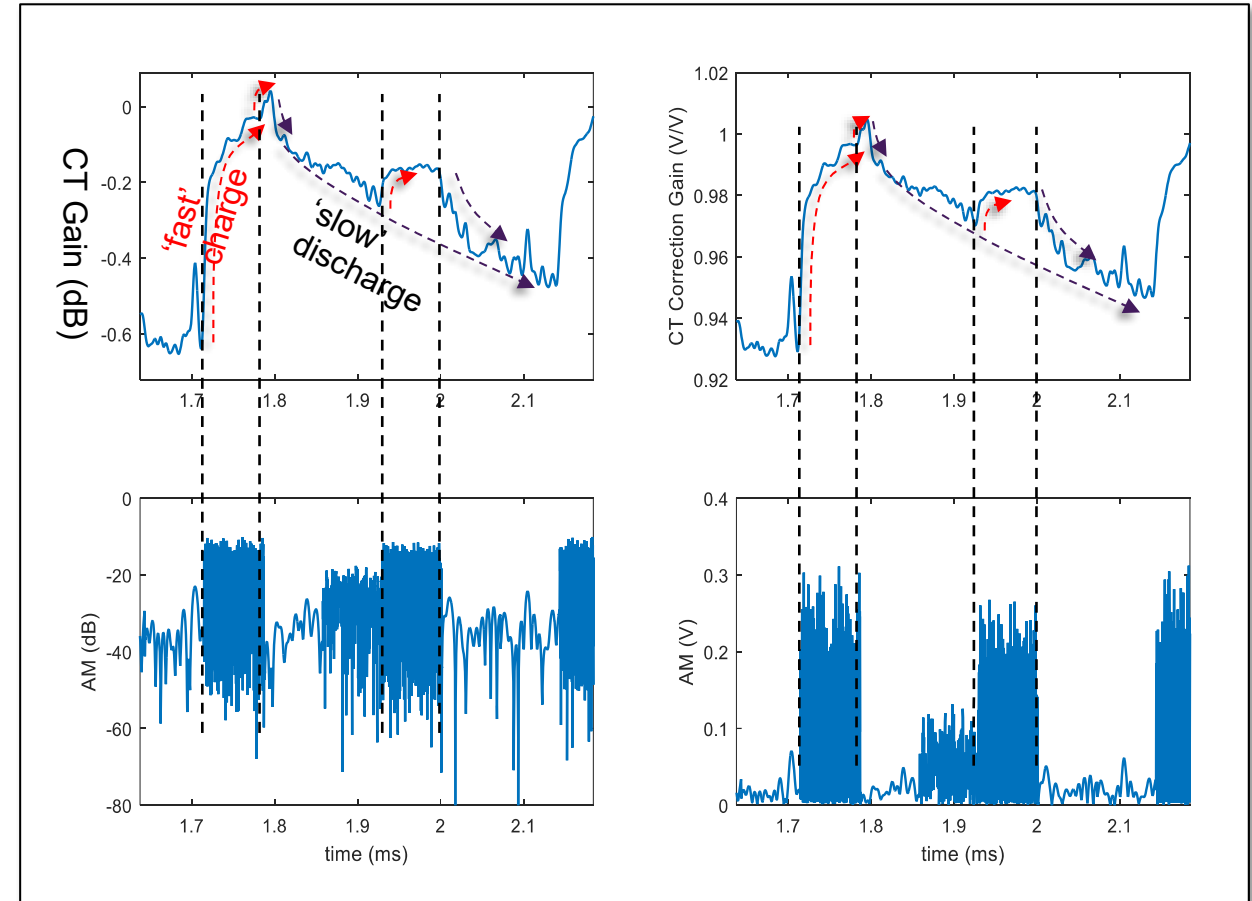


2D LUT @ 4-bit resolution



- DPD Model Architecture Discovery
- Neural-Network-assisted Physical Modeling for DPD
- Hybrid Model with Volterra-based Actuator and Neural Network

- Charge trapping in GaN HEMT based PAs
 - Interface traps in GaN HEMT exhibit long-term memory due to charge trapping/detrapping
 - Severe in-band error vector magnitude (EVM) in frames of varying power profile
 - Time Constant $> 500\mu\text{s}$
 - DPD with LT memory require sophisticated handcrafted models

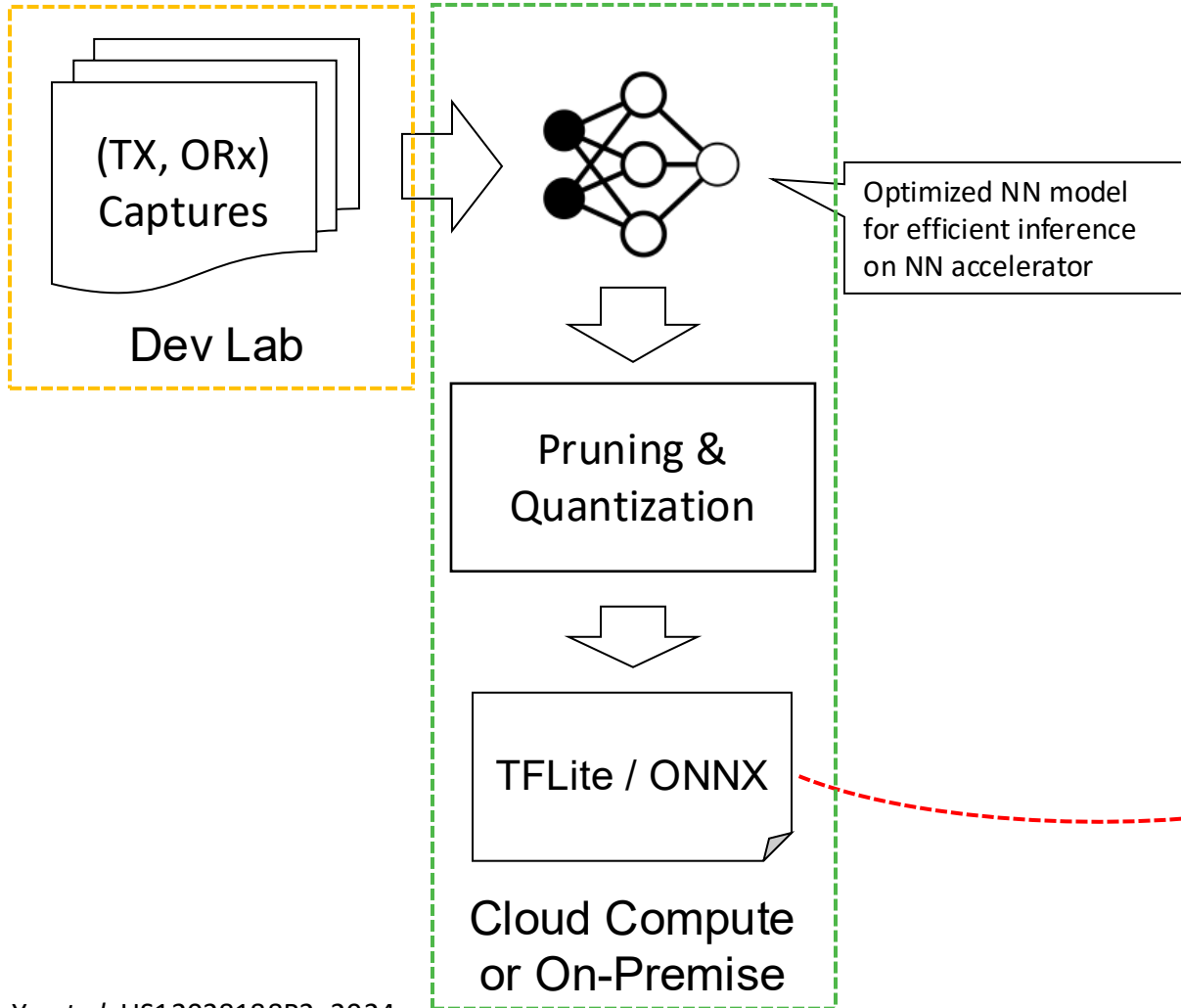


*Z. Yu *et al*, PAWR 2024

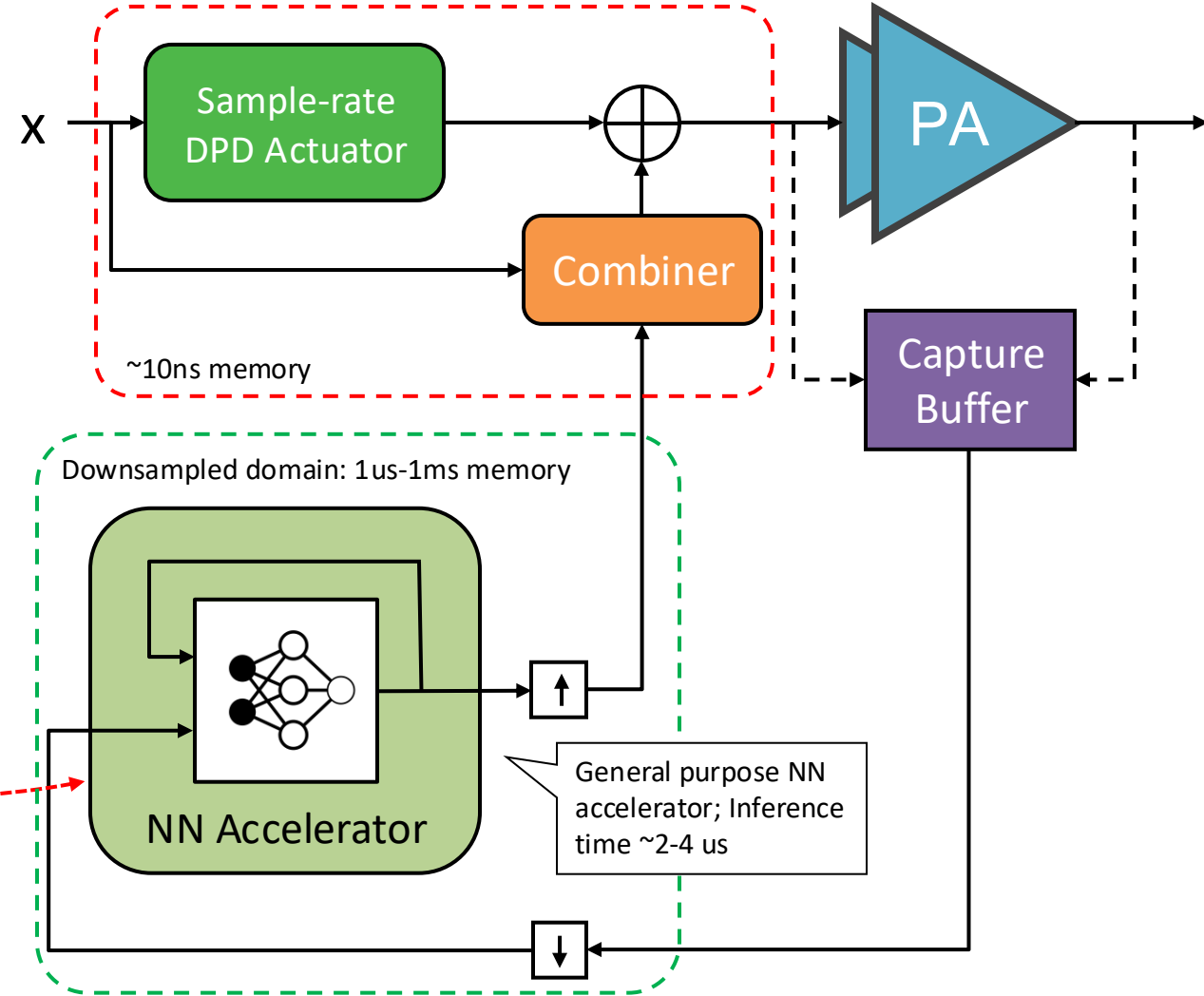
Z. Yu *et al*, PAWR 2025

A. Venkatasubramani *et al*, US11563409B2

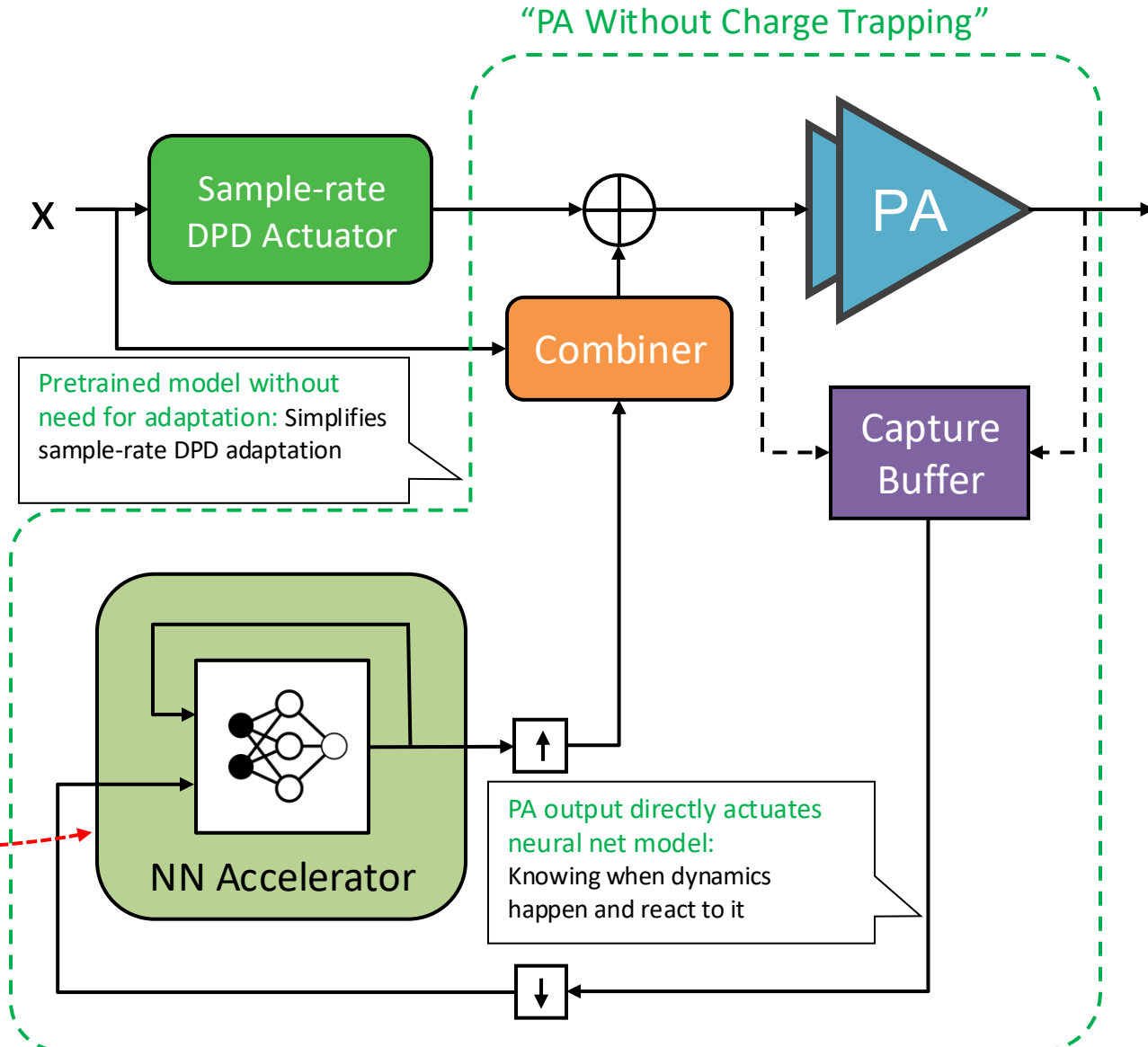
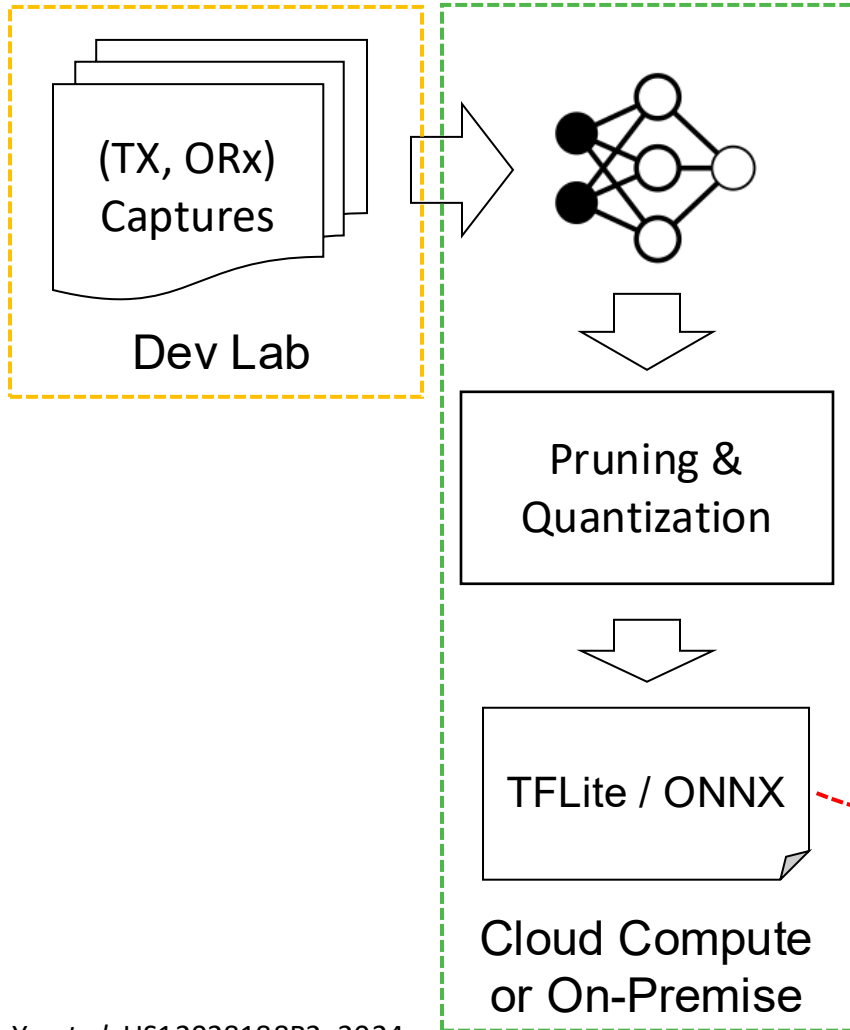
Offline Training



Online Actuation

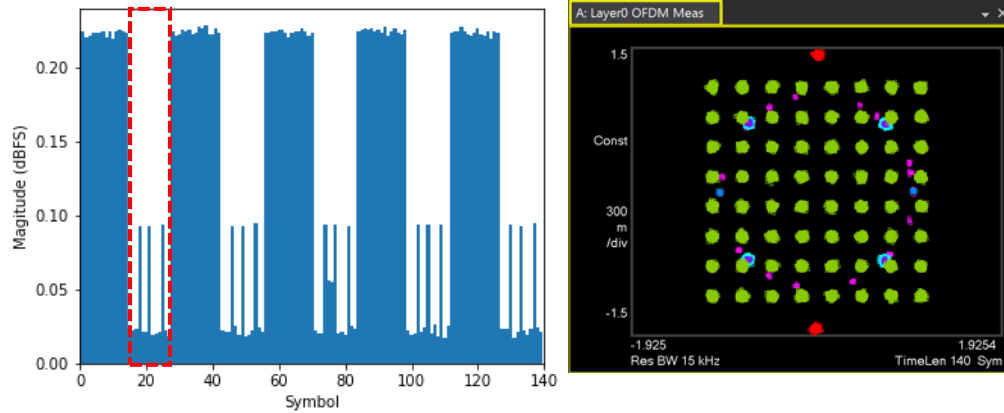


Offline Training

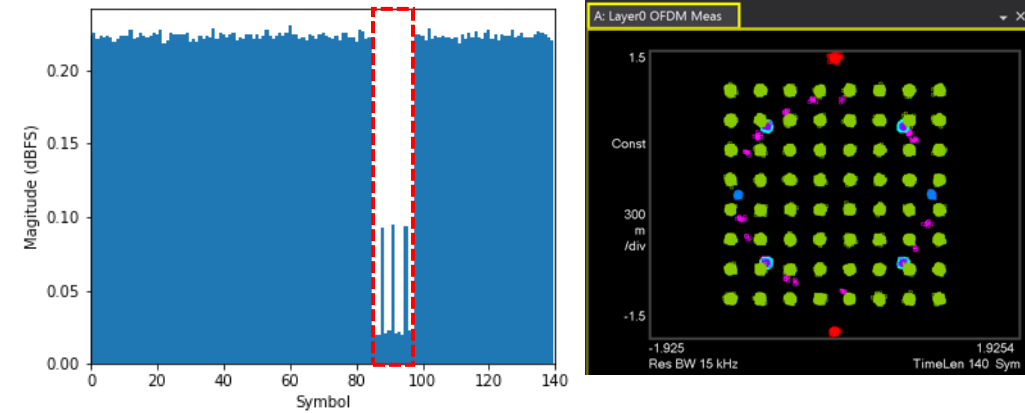


T. Yu et al, US12028188B2, 2024

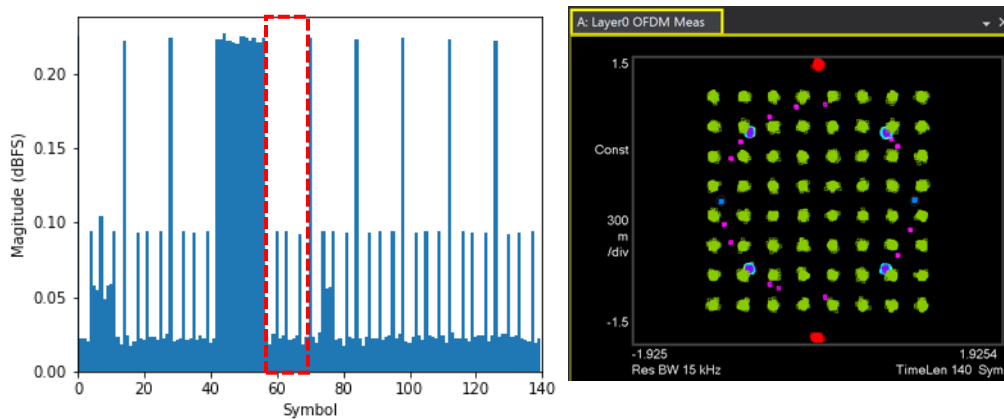
- QPA2705 GaN Doherty PA



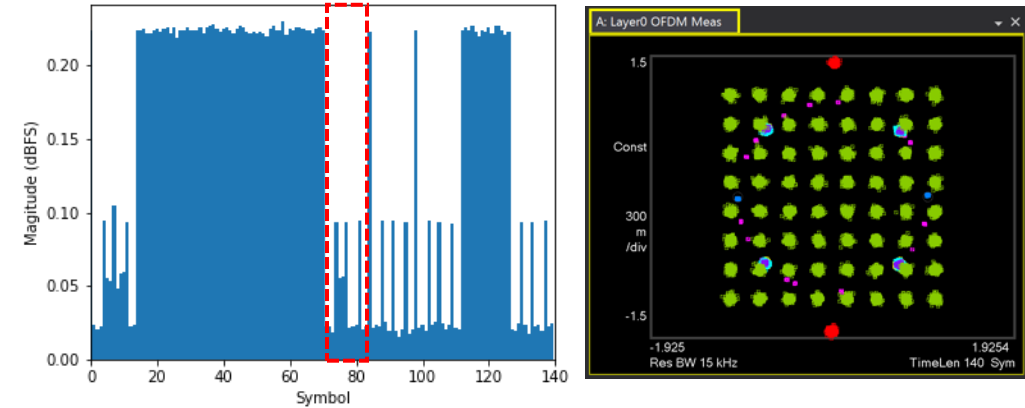
50/50 TM3/TM2, 3GPP 64QAM EVM = 1.80%
Subframe #1 EVM: **3.9%** → **1.7%** with Neural DPD



90/10 TM3/TM2, 3GPP 64QAM EVM = 1.96%
Subframe #6 EVM: **4.3%** → **2.3%** with Neural DPD

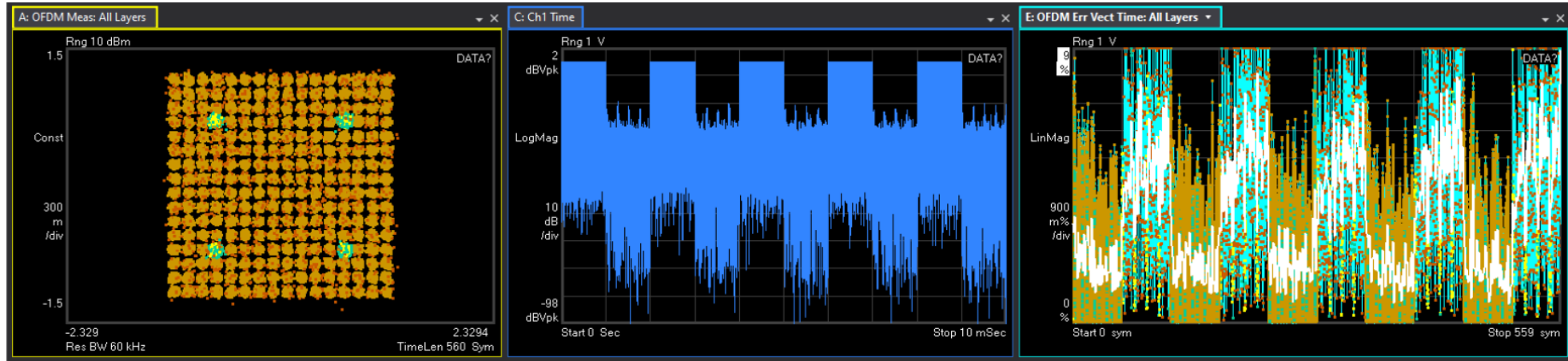


10/90 TM3/TM2, 3GPP 64QAM EVM = 1.86%
Subframe #4 EVM: **3.95%** → **0.8%** with Neural DPD

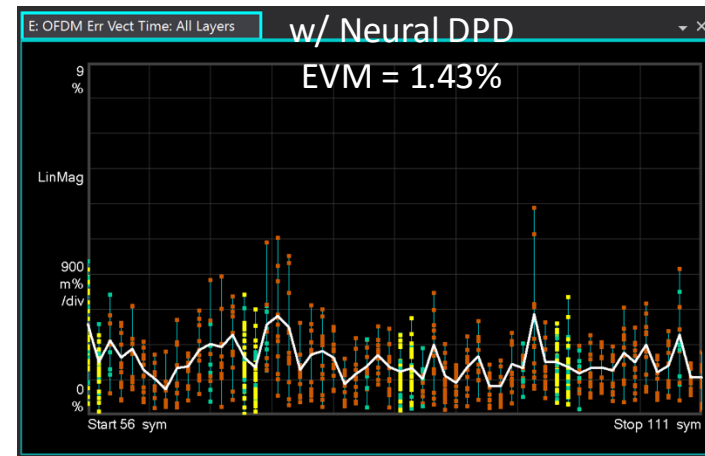
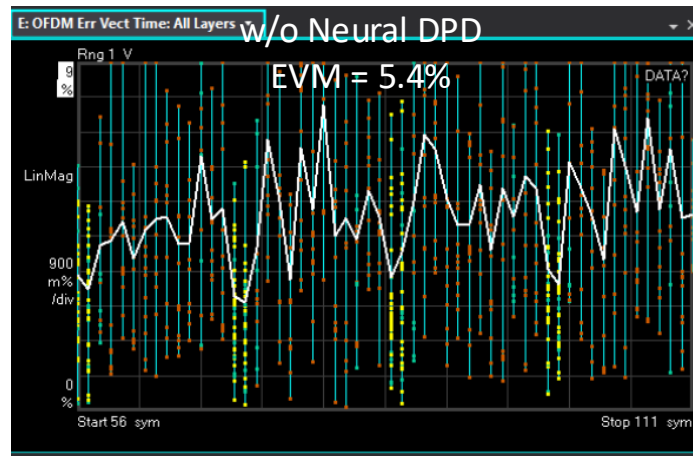


Random Allocation, 3GPP 64QAM EVM = 2.30%
Subframe #5 EVM: **4.7%** → **2.5%** with Neural DPD

- 40W GaN Doherty Macro PA
 - NR 100MHz, 60KHz SCS, 50/50 135/1 RBs



EVM of Subframe #1



Samana – Monolithic 4T4R/8T8R Split 7.2a Platform

Minimizes Dev and System Cost

- Native 4T4R Basestation on a chip!
- Adaptability: 400MHz iBW, 400MHz oBW, 400MHz F_c <math>< 7.125\text{GHz}</math>
- End-to-end SW integration eases ORU development



5G TDD
Small Cell



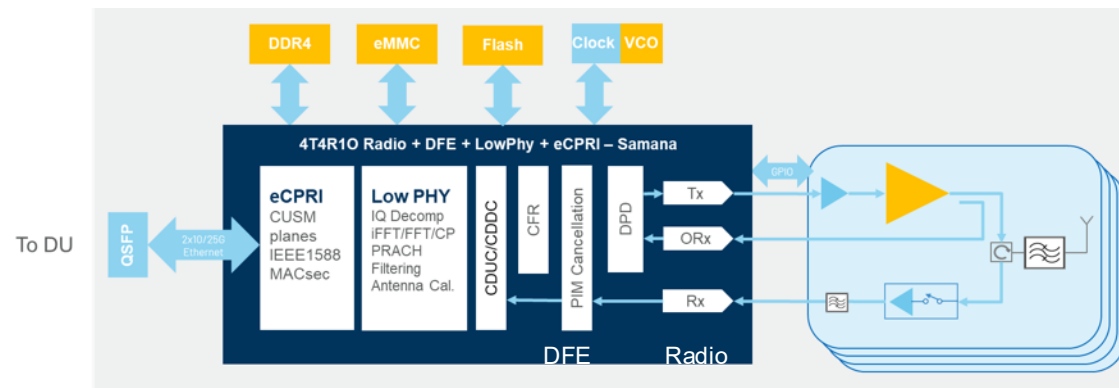
4G FDD
Small Cell

New Industry Benchmark in Features & Power

- PIMc, MACsec, compression, DPD6
- Significant power reduction vs FPGAs
- Extensive programmability options enable customer differentiation

Integrating Proven Technology Minimizes Risk

- ADI 3rd Gen development on 16nm
- Integrating proven Radio + DFE technology powered by AI/ML
- >3 Million ADI DFE Transceivers deployed

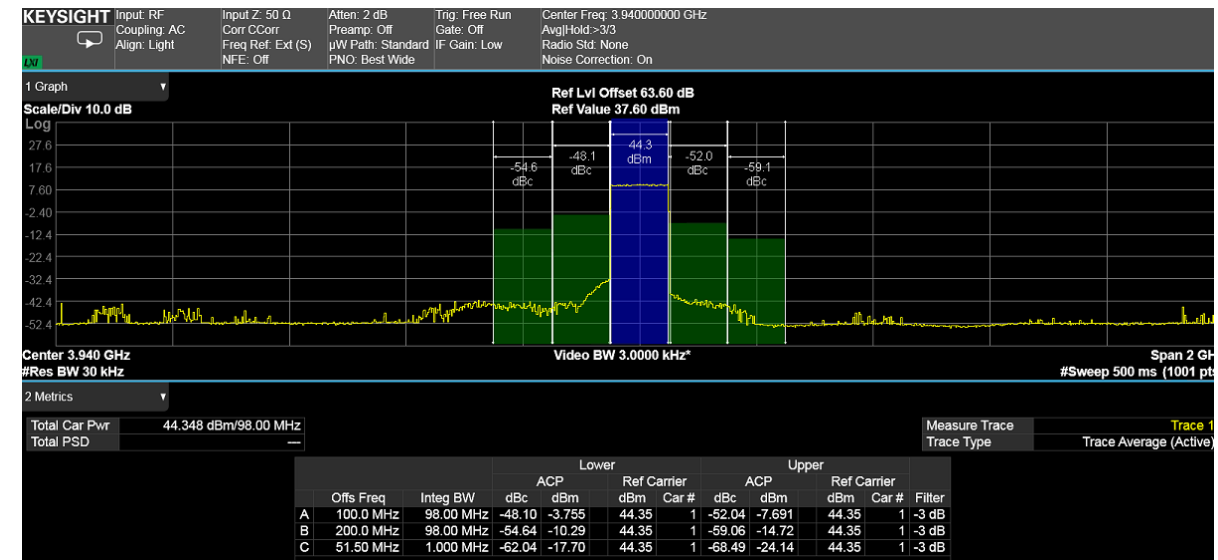
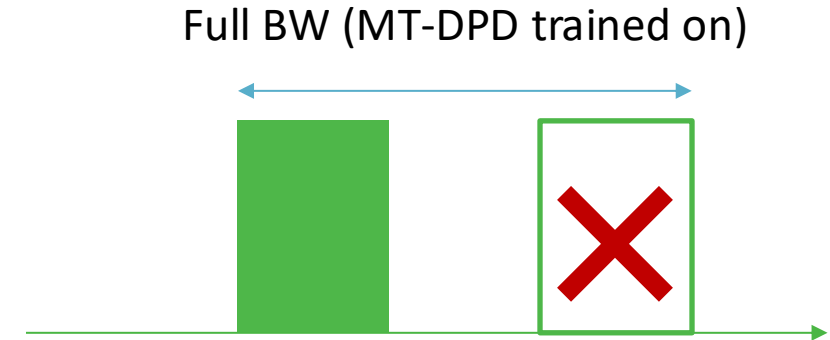


TDD Macro



Multi-band
FDD Macro

- High Power Macro PAs: 40-50W → 100W (400W 4T4R Cat. B)
 - OBUE: < -19 dBm/MHz @ 1.5MHz offset
 - High Efficiency > 43%
- Robustness across Wide Range of Use Cases
 - Extensive stress tests, repeated and reproduced



- Machine Learning with Physical Insights
 - Efficient tool to reduce system complexity
 - Computationally efficient incorporating physical insights
- Paradigm Shift & Design Considerations
 - “Oversizing” hardware for flexibility
 - Value of AI/ML tooling
- Opportunities & Future Directions
 - Incorporation of more AI-native & AI-driven design
 - More research to be done in larger scale transmitters