



# A Transformer-Based Foundation Model to Capture Generic Device Physics

MOS-AK workshop, Silicon Valley  
Dec. 14, 2023

# How AI impact engineering

AlphaGo (2016)



**Message:**  
"AI will win every well defined optimization game"

**Limitation:**  
"In most cases, engineering problems cannot be mathematically defined, relying on human knowledge"

ChatGPT (2022)



**Message:**  
"Every human knowledge can be pre-trained"

**What Knowledge?**

# Pre-trained semiconductor knowledge

H

Explain the current flowing mechanism of the I-MOS



The I-MOS, or Impact-Ionization Metal-Oxide-Semiconductor, is a specialized type of Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) designed to take advantage of the impact ionization phenomenon for amplification and switching purposes. To understand the current flowing mechanism of the I-MOS, it's essential to grasp the concept of impact ionization and how it operates in this device.

1. **Structure:** The I-MOS has a unique structure compared to traditional MOSFETs. It typically consists of a heavily doped drain region, a lightly doped channel region, and a gate electrode separated by an insulating oxide layer.
2. **Off-State (No Applied Voltage):** In the off-state (when no voltage is applied to the gate), the I-MOS control electrode is reverse-biased, and no current flows through the channel.
3. **Impact Ionization:** When a high drain voltage is applied, the electric field in the channel becomes strong enough to cause impact ionization. This process occurs within the channel region, where electrons gain enough energy to create additional electron-hole pairs.
4. **Avalanche Effect:** Once impact ionization begins, it triggers an avalanche effect, causing the newly generated electrons to gain even more energy and create additional electron-hole pairs. This results in a rapid multiplication of charge carriers in the channel.
5. **Current Amplification:** The avalanche effect leads to a significant increase in the drain current. This is the unique feature of the I-MOS. It can amplify the input signal by a large factor due to the impact ionization process, making it suitable for applications where high gain or amplification is required.
6. **Gate Control:** The gate voltage can still influence the channel's conductivity, but in the I-MOS, the primary mechanism of current flow is dominated by the impact ionization process.

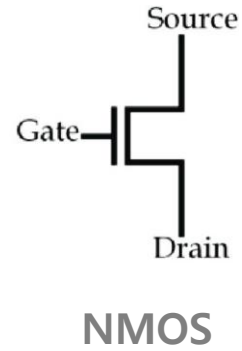
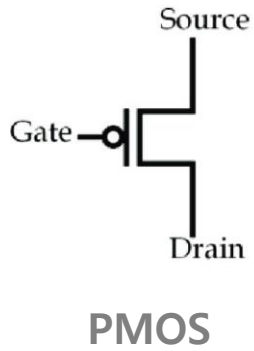
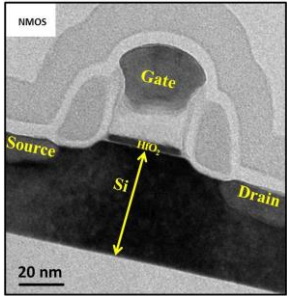
In summary, the current flowing mechanism of the I-MOS relies on the impact ionization process in the lightly doped channel region. When a high drain voltage is applied, impact ionization generates additional charge carriers, leading to an avalanche effect and a significant increase in the drain current. The gate voltage controls the extent of this process, allowing for amplification of the input signal. The I-MOS is primarily used in specialized applications where high gain or amplification is required.

Verbal knowledge on semiconductor is already pre-trained by chatGPT

Can we train non-verbal knowledge?

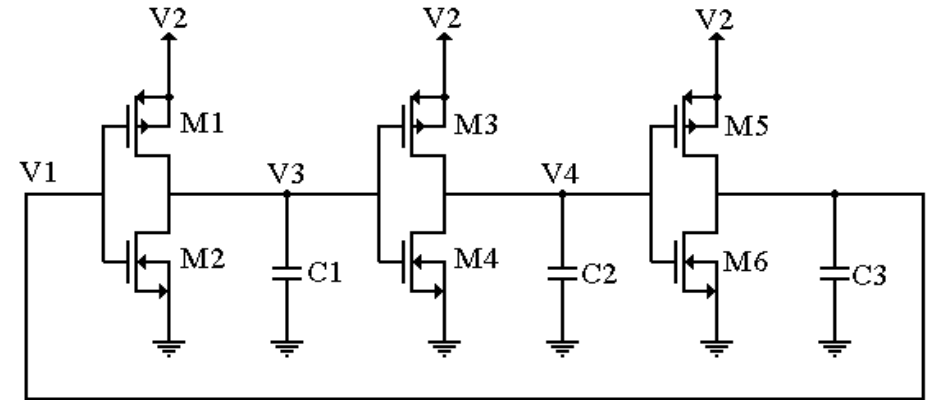
Such as device physics

# Compact model



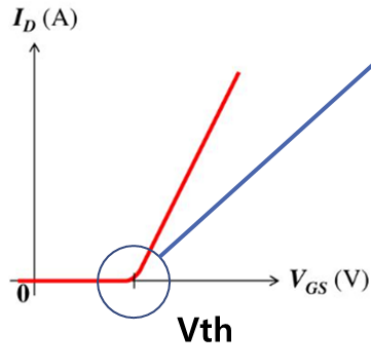
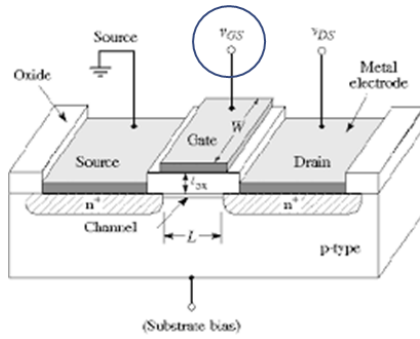
Compact Modeling  
→

$$[\{I\}, \{Q\}] = f(V_g, V_d, V_s, V_b)$$



Circuit Simulation

# BSIM equations for planar MOSFET (threshold voltage model)

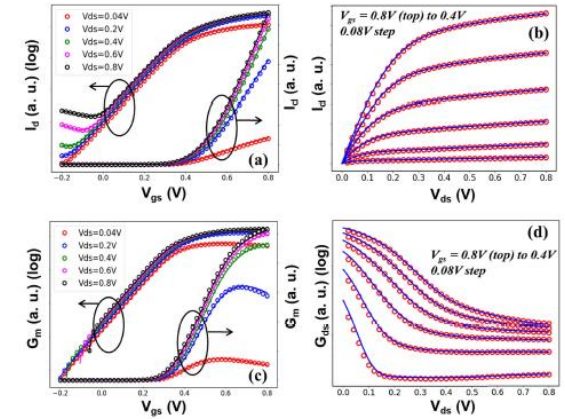
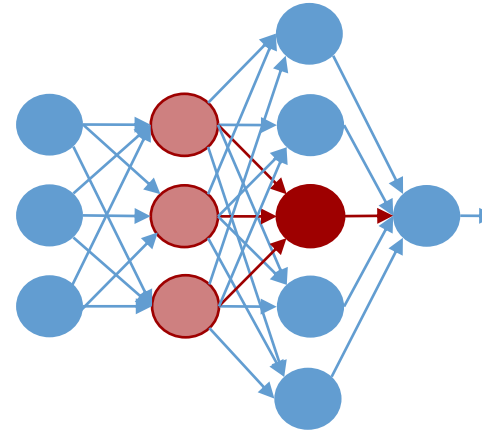
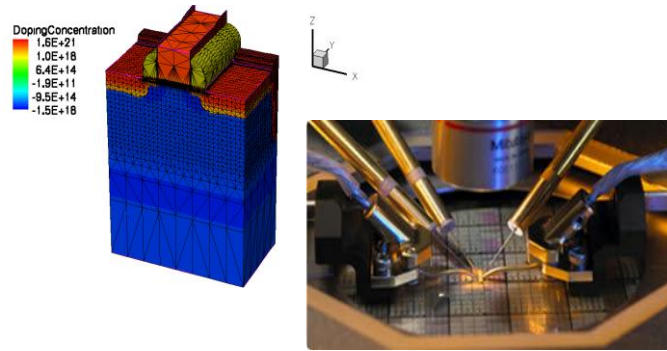


$$\begin{aligned}
 V_{th} = & V_{TH0} + \left( K_{1ox} \cdot \sqrt{\Phi_s - V_{bseff}} - K_{1l} \cdot \sqrt{\Phi_s} \right) \sqrt{1 + \frac{LPEB}{L_{eff}}} - K_{2ox} V_{bseff} \\
 & + K_{1ox} \left( \sqrt{1 + \frac{LPE0}{L_{eff}}} - 1 \right) \sqrt{\Phi_s} + \left( K_3 + K_{3B} \cdot V_{bseff} \right) \frac{TOXE}{W_{eff}' + W_0} \Phi_s \\
 & - 0.5 \cdot \left[ \frac{DVT0W}{\cosh\left(DVT1W \frac{L_{eff}W_{eff}'}{l_{re}}\right) - 1} + \frac{DVT0}{\cosh\left(DVT1 \frac{L_{eff}}{l_t}\right) - 1} \right] (V_{bi} - \Phi_s) \\
 & - \frac{0.5}{\cosh\left(DSUB \frac{L_{eff}}{l_{t0}}\right) - 1} (ETA0 + ETAB \cdot V_{bseff}) \cdot V_{ds} - nv_t \cdot \ln \left( \frac{L_{eff}}{L_{eff} + DVTP0 \cdot (1 + e^{-DVTP1 \cdot V_{ds}})} \right) \\
 & - \left( DVTP5 + \frac{DVTP2}{L_{eff}^{DVTP3}} \right) \cdot \tanh(DVTP4 \cdot V_{ds})
 \end{aligned}$$

## Too complex to model modern devices

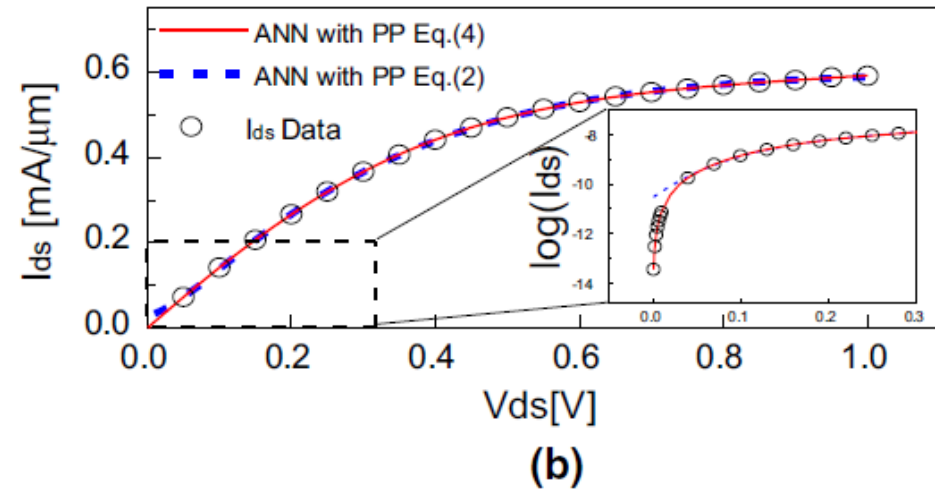
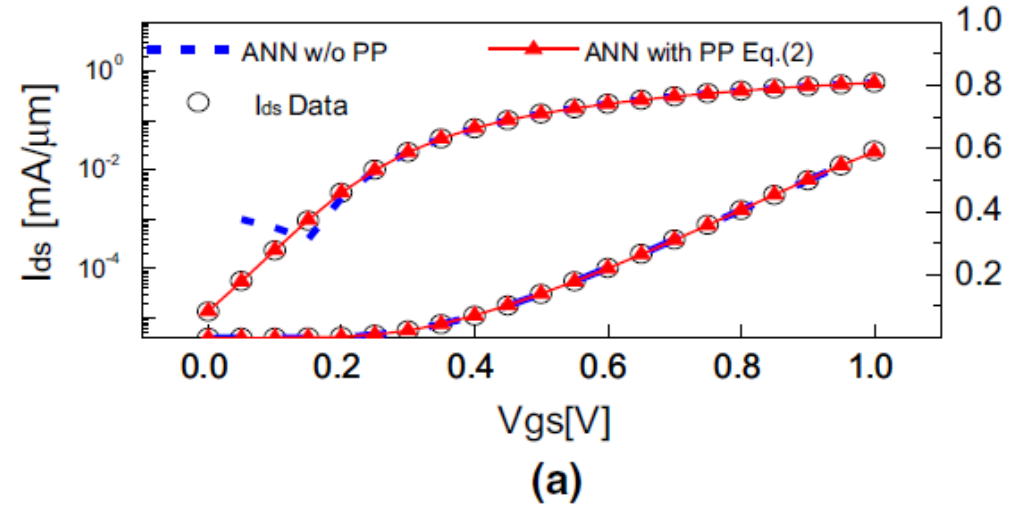
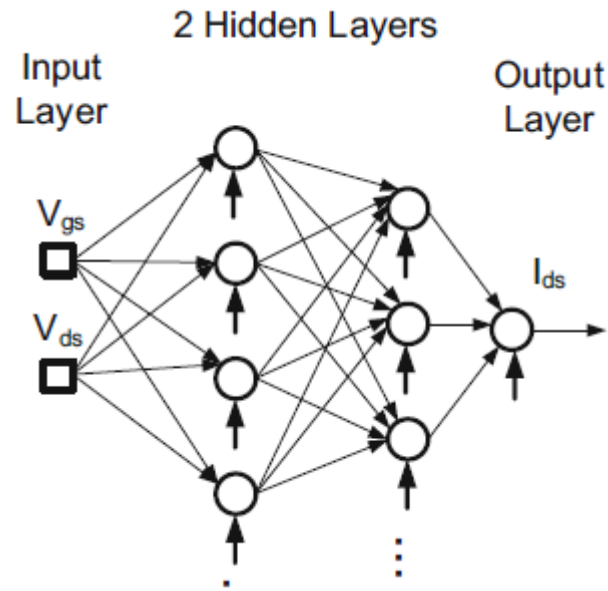
- ✓ It takes time to develop model and extract parameters
  - ✓ Only experts can do that.
- ✓ It is extremely hard to edit industry compatible model
- ✓ People rely more on data, rather than physics. (binning model)

# Neural compact modeling

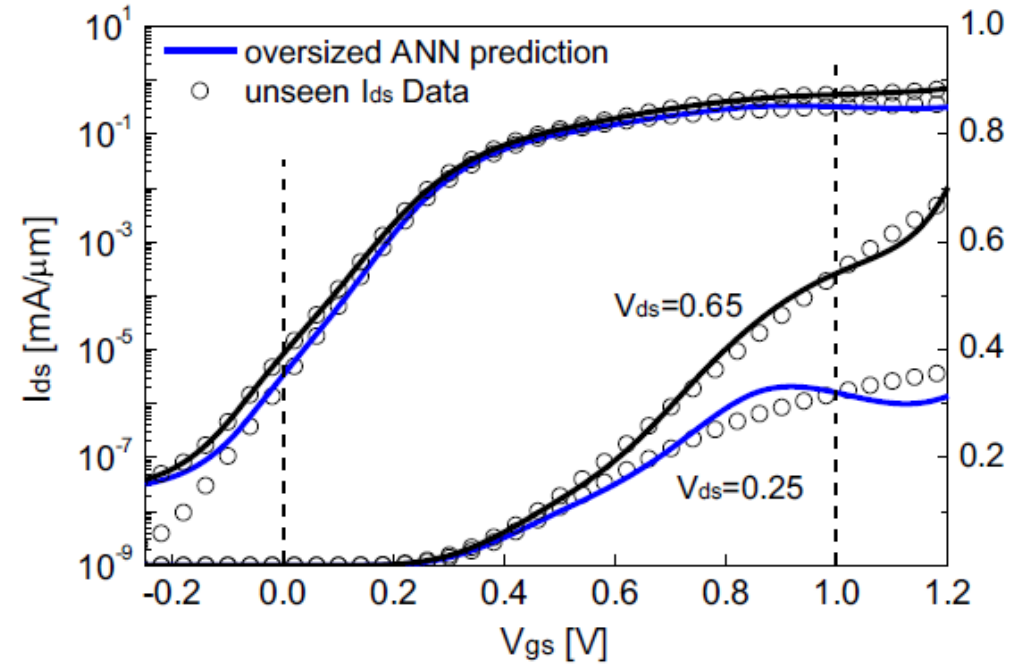
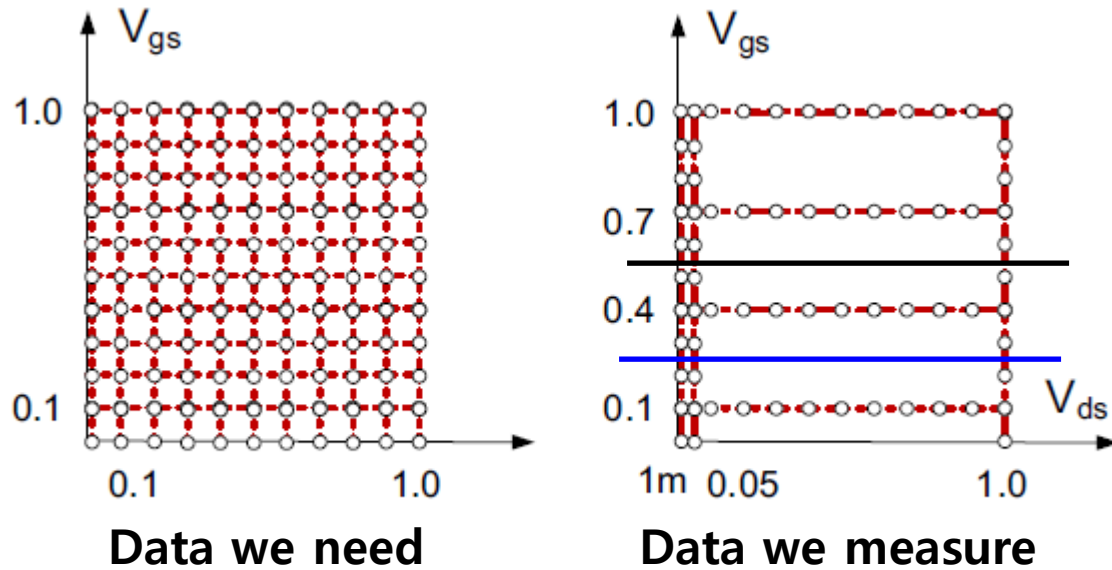


- ✓ ANN, instead of equation sets, to model the behavior of the device.
- ✓ Train ANN with the data you measured or simulated.

# Simple regression model



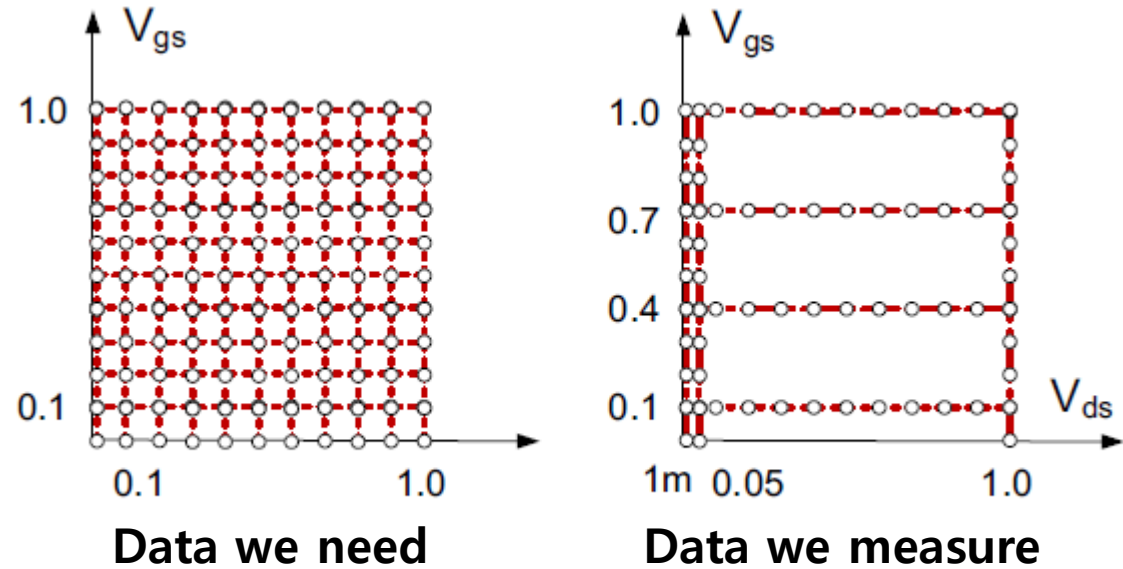
# Simple regression model (limitation)



**Inaccuracy when data not provided**



# Neural compact modeling with limited measured data



Not enough IV measurement, and even less CV measurement for transient simulation

Interpolation/extrapolation is not enough, because we need additional information to infer empty region.

## Solution to tackle data scarcity

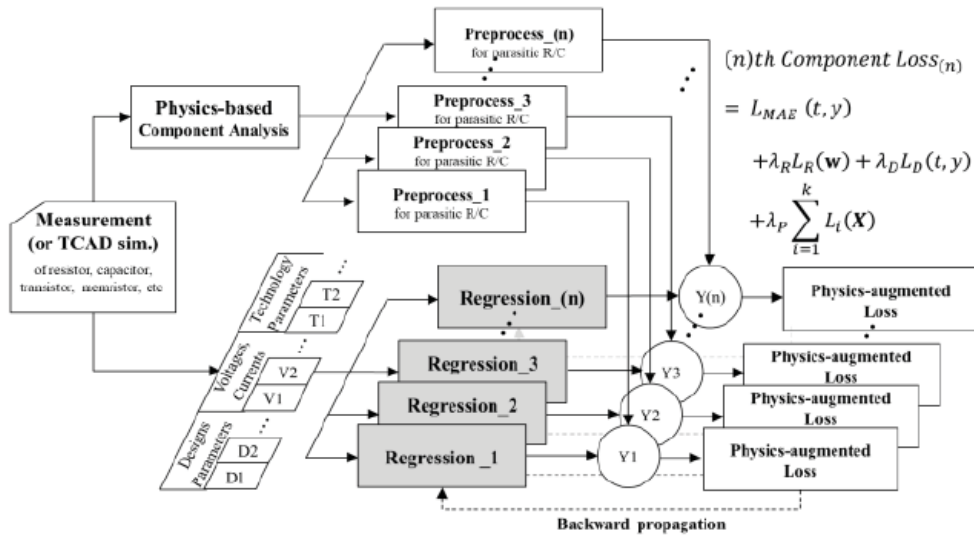
### 1. Physics embedded in NN architecture

- ✓ Edit NN architecture to incorporate physics in the model
- ✓ Most of the prior works use this method

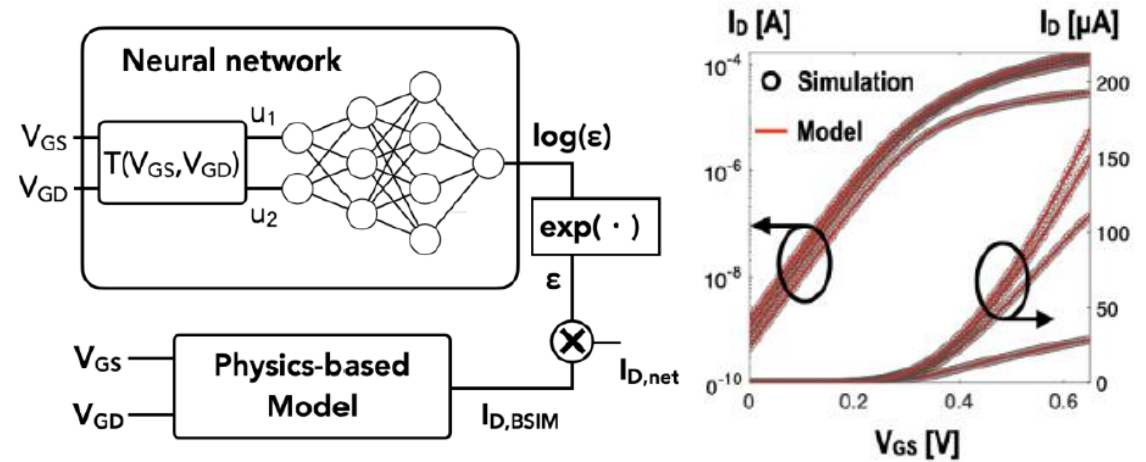
### 2. Using pre-trained physics

- ✓ Learn physics from larger dataset, which shares same physics with target device

# Physics embedded NN architecture



Design complex NN architecture



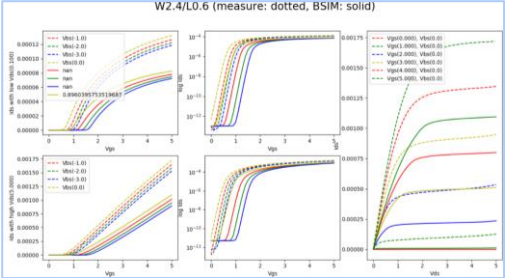
Mixing BSIMCMG with NN

## Limitation of physics embedded NN architecture

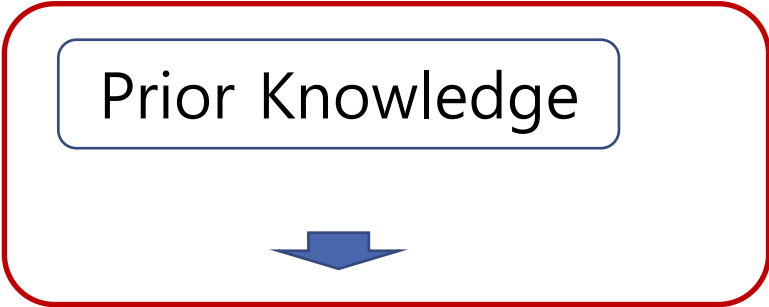
- ✓ Physical aspects of the model can be blurred by statistical learning of the NN architecture.
- ✓ It is very hard to design architecture every time, esp. input/output dimension of the model increases.

# Using Pre-trained Physics

$$I = \mu_0 \frac{W}{L} (V_{gs} - V_{th}) \cdot V_{ds}$$



$$I = 0.003 \frac{W}{L} (V_{gs} - 0.25) \cdot V_{ds}$$



Can we learn this?



# Meta Learning for Scientific Function (MAML)

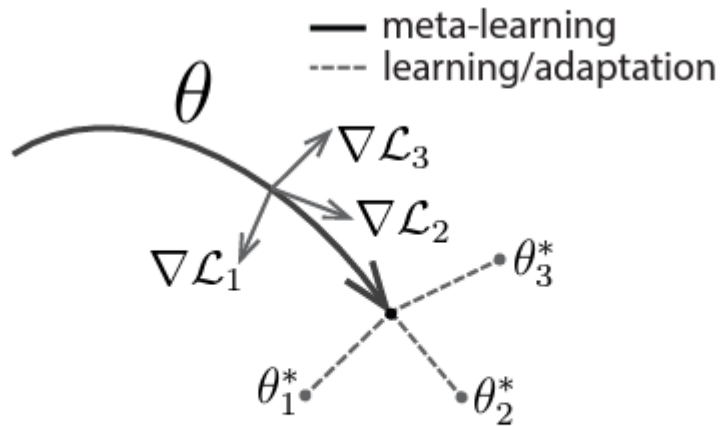
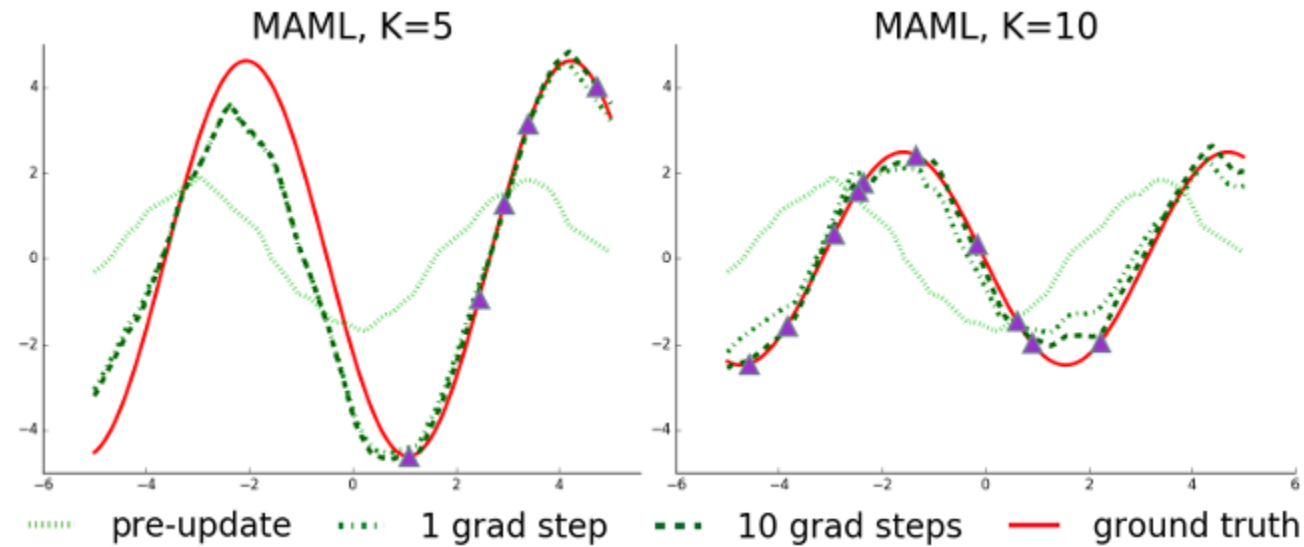


Figure 1. Diagram of our model-agnostic meta-learning algorithm (MAML), which optimizes for a representation  $\theta$  that can quickly adapt to new tasks.



- ✓ Learning sine function with different amplitude, frequency, phase.
- ✓ Once we learn sine function, we can complete sine function with a few measure points, which has not enough information.

# Meta Learning for Scientific Function (LEO)

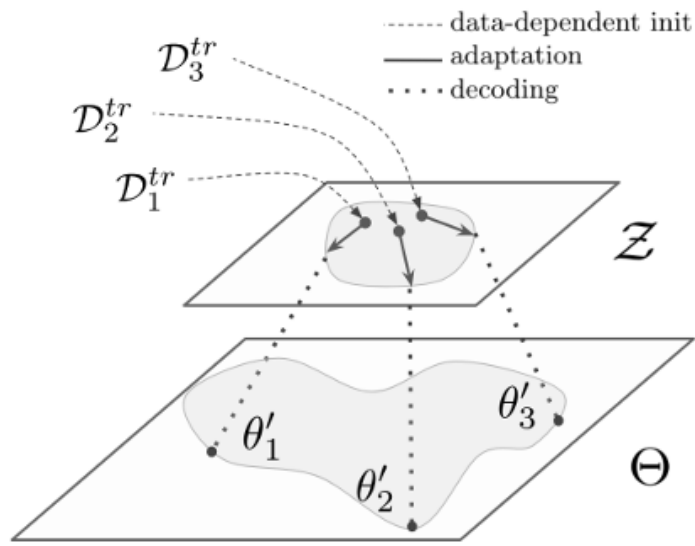


Figure 1: High-level intuition for LEO. While MAML operates directly in a high dimensional parameter space  $\Theta$ , LEO performs meta-learning within a low-dimensional latent space  $\mathcal{Z}$ , from which the parameters are generated.

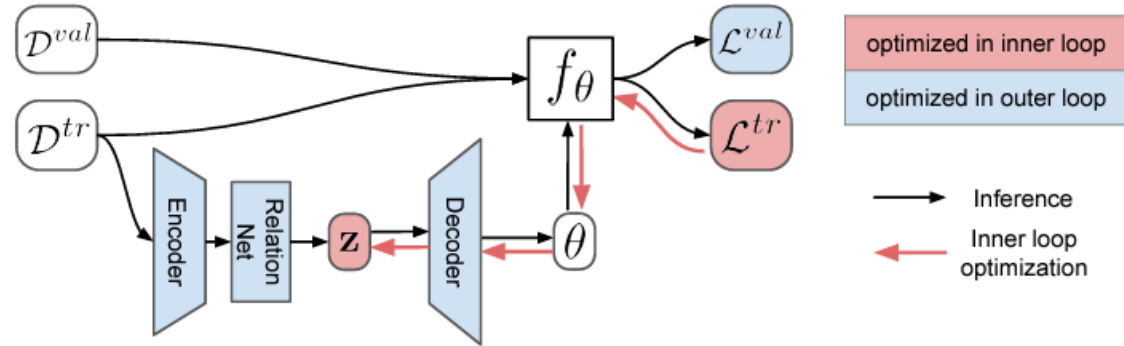


Figure 2: Overview of the architecture of LEO.

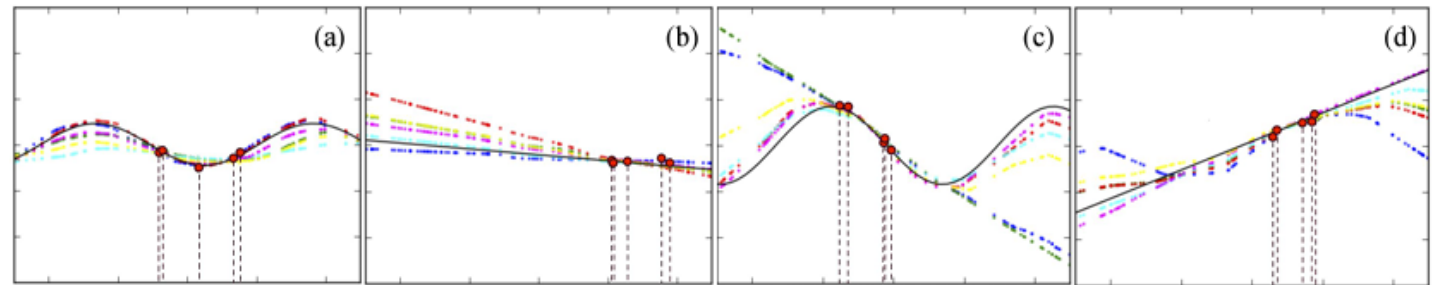


Figure 3: Meta-learning with LEO of a multimodal task distribution with sines and lines, using 5-shot regression with noisy targets. Our model outputs a distribution of possible solutions, which is also multimodal in ambiguous cases. True regression targets are plotted in black, while the 5 training examples are highlighted with red circles and vertical dashed lines. Several samples from our model are plotted with dotted lines (best seen in color).

✓ Sine, Polynomial, probabilistic distribution, better accuracy.

# Meta Learning for Scientific Function (Hamiltonian)

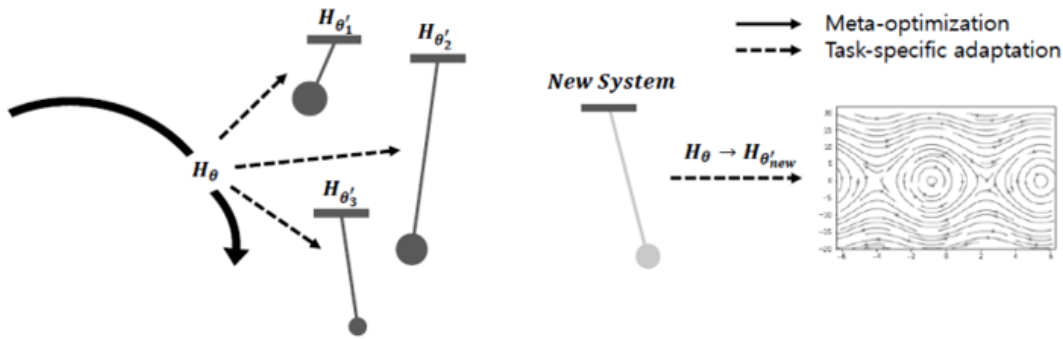
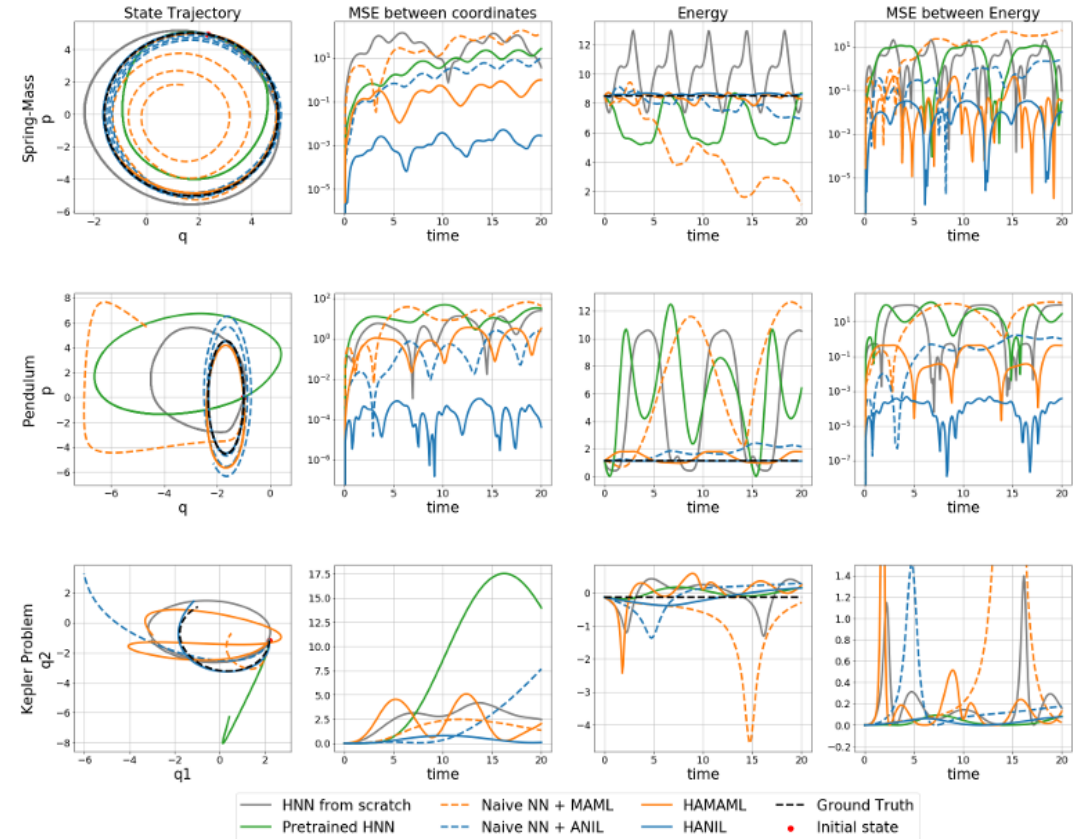


Figure 1: There is a resemblance between meta-learning and identifying the physical laws of Hamiltonian. A hypothesized governing equation of Hamiltonian, usually corrected and established by evaluating many related systems, could be learned using meta-learning as a data-driven method (left). Then, a well-established Hamiltonian can be utilized to predict new system dynamics, which could be viewed as a meta-transfer process by a well-trained meta-learner (right).

$$H_{\theta'_{new}}(q, p) = \frac{p^2}{2m_{new}l_{new}^2} + m_{new}g_{new}l_{new}(1 - \cos q)$$

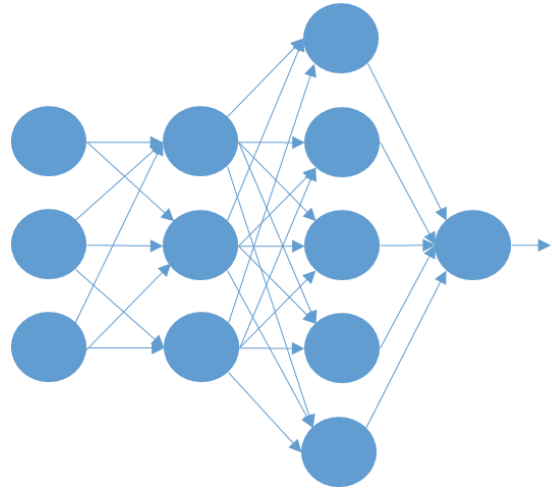
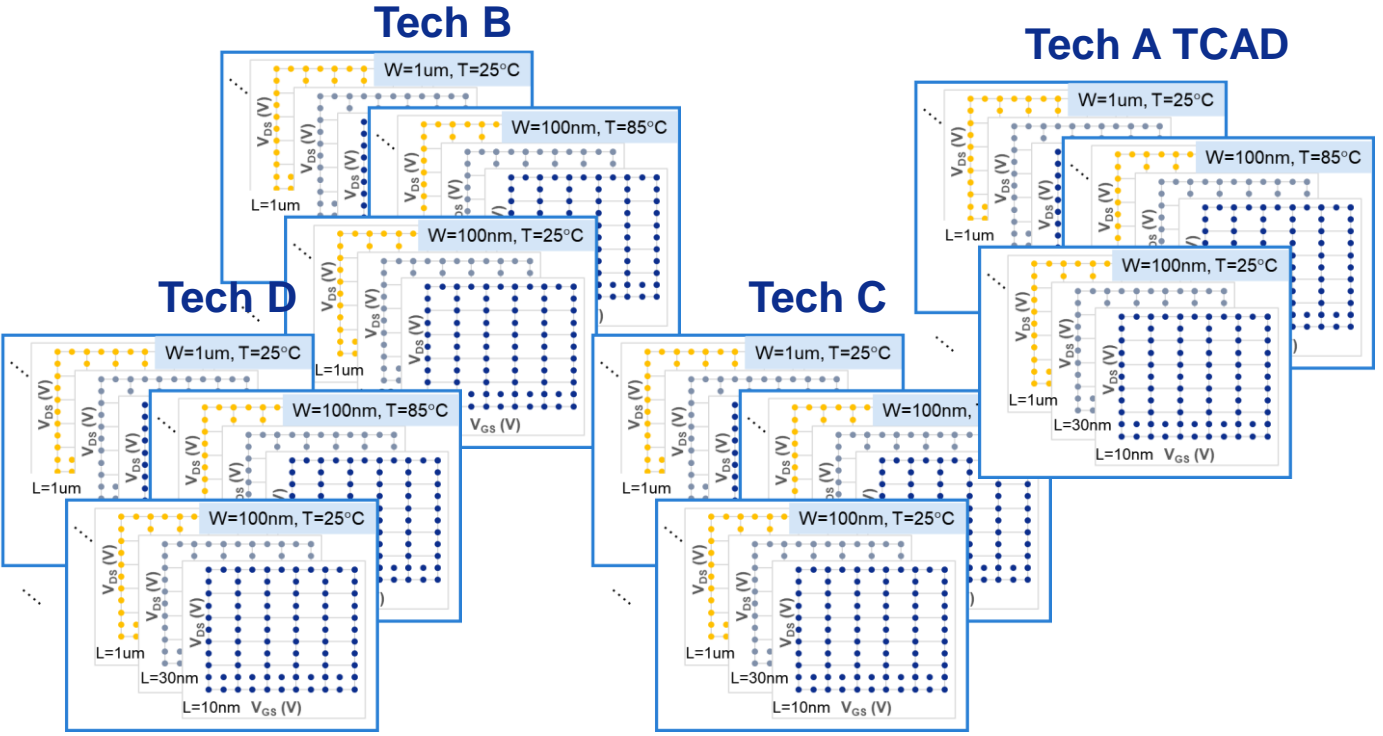


✓ Hamiltonian function (e.g. pendulum movement)

# Learning device physics from previous technology

Learn to learn tasks w/ other technology data and/or TCAD data

Quickly learn tech A





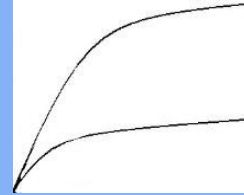
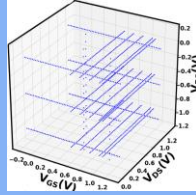
# Learning device physics – problem setup

Meta-  
Training

W/L/T  
Source  
Device

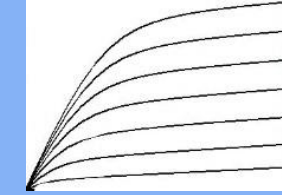
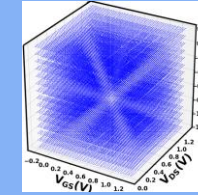
Support  
Set

45nm



Query  
Set

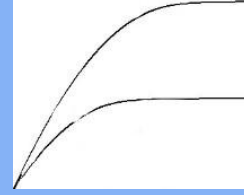
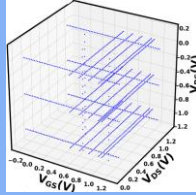
45nm



W/L/T  
Source  
Device

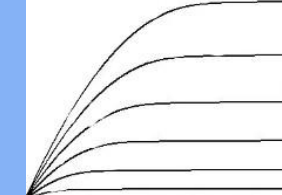
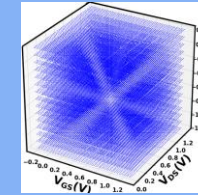
Support  
Set

45nm



Query  
Set

45nm

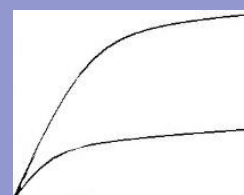
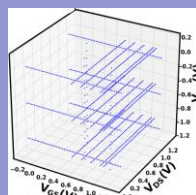


Meta-  
Testing

W/L/T  
Target  
Device

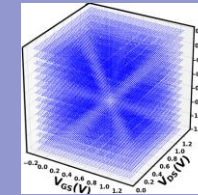
Support  
Set

32nm



Query  
Set

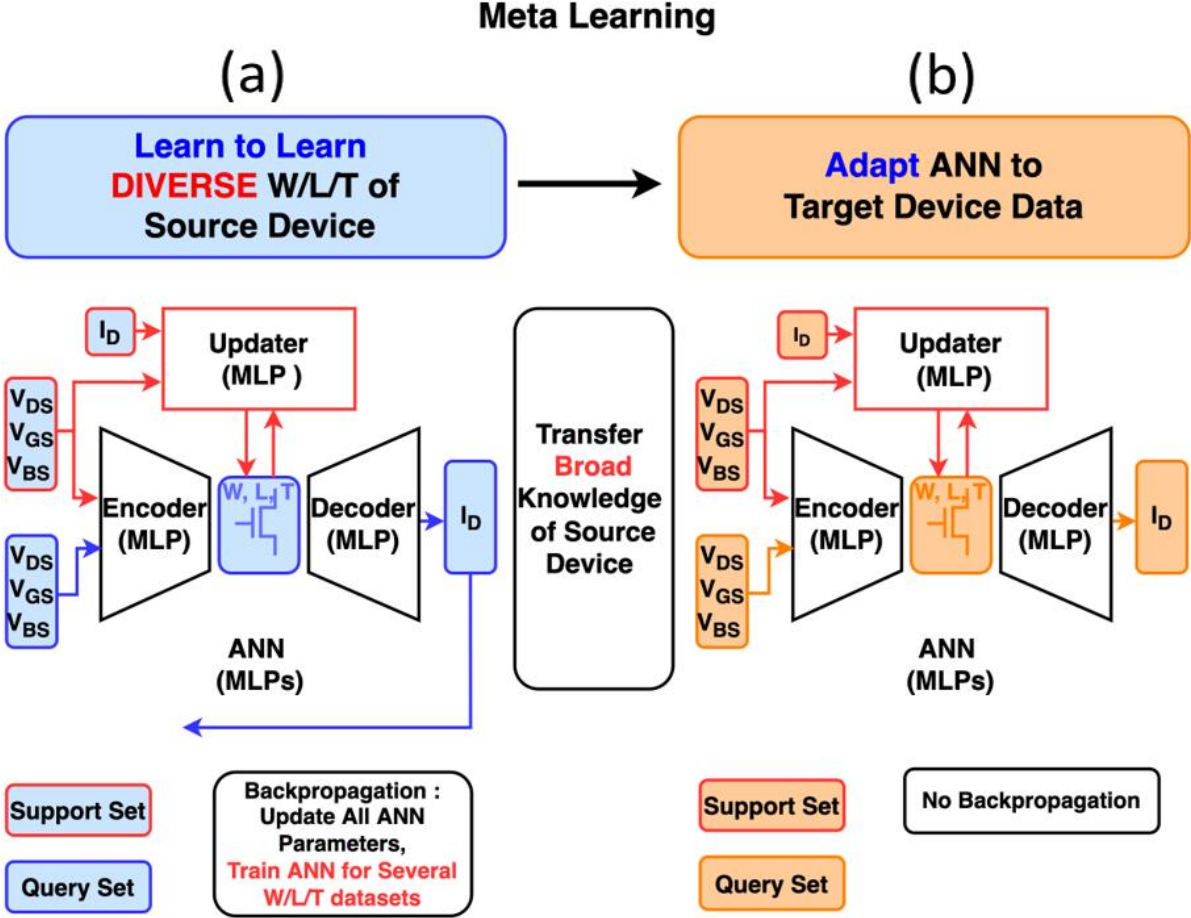
32nm



?

- ✓ Assume we have abundant data in 45nm tech node.
- ✓ With typical measure conditions, making 23nm model.

# Learning device physics – Model Architecture



"A Novel Methodology for Neural Compact Modeling Based on Knowledge Transfer", Alsemy, SISPAD, 2022

# Learning device physics – Model performance

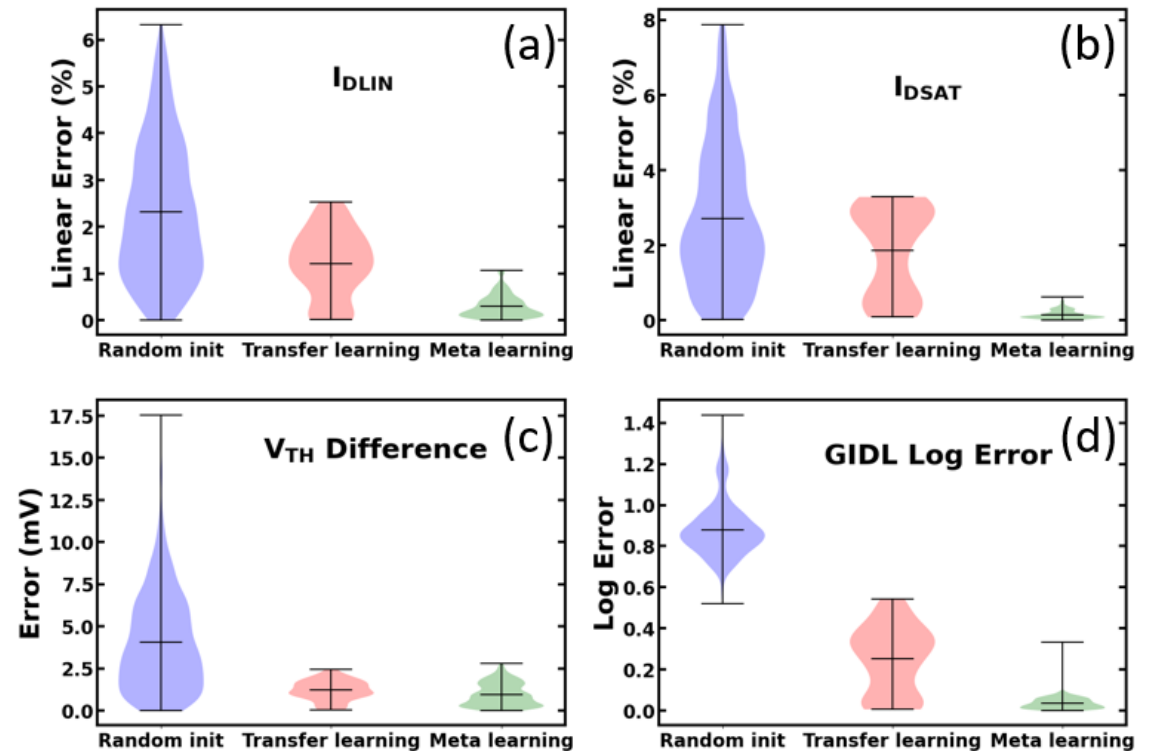
[Computational Costs and Test Errors]

- The **meta-trained ANN** shows the **lowest average relative linear and log errors**, with the shortest adaptation time.

	Random Initialization	Transfer Learning	Meta Learning
Pretraining Time	N/A	646 sec.	17 hours
Adaptation Time (per W/L/T)	538 sec.	186 sec.	1 sec.
Relative Linear Error (%)	22.9	4.3	2.3
Relative Log Error (%)	1.56	0.40	0.11

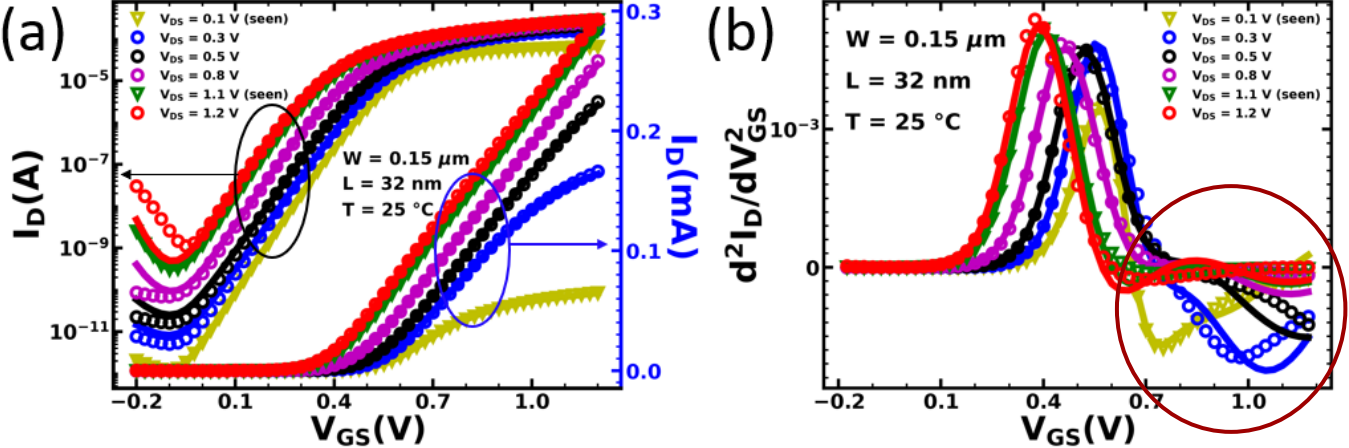
[Prediction Accuracy of Electrical Parameters]

- The **meta-trained ANN** predicts various electrical parameters for **any W/L/T** of the target device in a stable and accurate way.



# Learning device physics – Accuracy, smoothness

Random Initialization

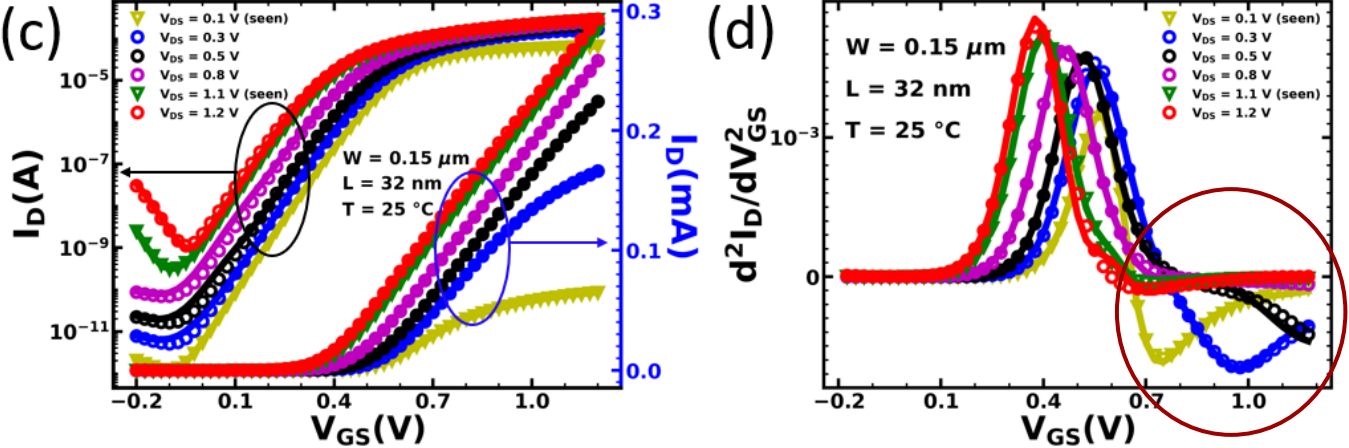


▽ Available Data (Seen)

○ Evaluating Data (Unseen)

— Model

Meta learning

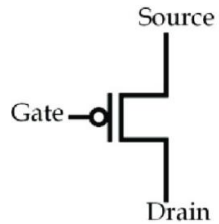


## Foundation model

- ✓ We observed that the physical aspects of the devices can be pre-trained.
- ✓ But we need to increase the capacity of the model for practical use case. (multi input for design/process parameters, CV characteristics, etc.)
- ✓ Once we learn the physics, we want to re-use it in other application. (nominal model generation, re-shaping, causal inferences)
- ✓ Transformer architecture enables it.
- ✓ Self attention on numerical relation, instead of word relation in sentence.

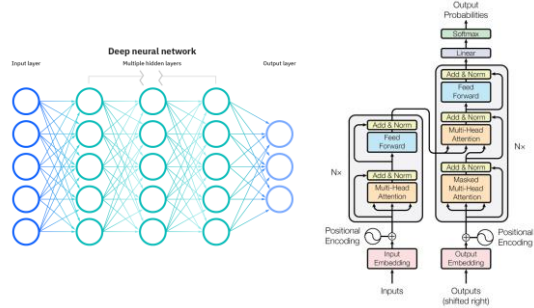
# How it works (foundation model)

## Large Data



SPICE generated high volume synthetic dataset with various technology node and process/design parameter variation.

## Foundation Model



Transformer architecture modified to process continuous scientific functions

## Tasks

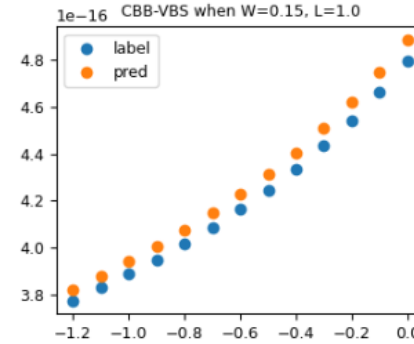
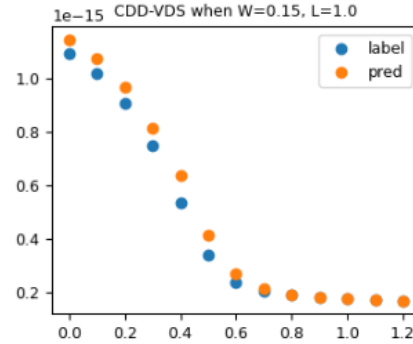
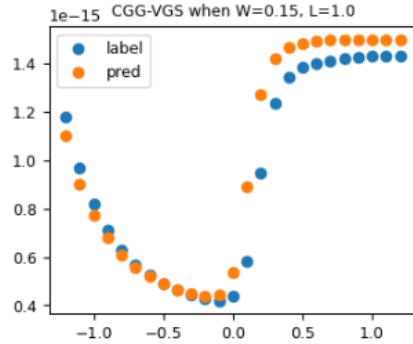
Generate compact model based on partial measurement

Re-target model function based on point target (e.g. Idsat, Idlin, Vth)

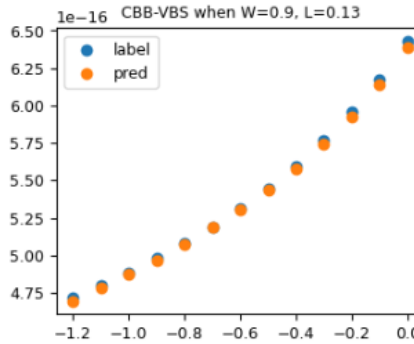
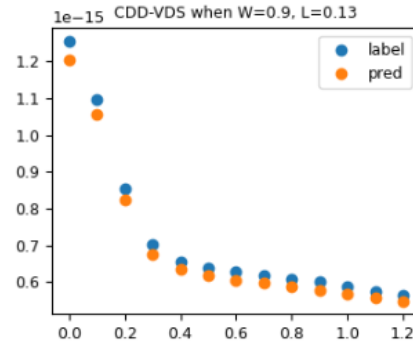
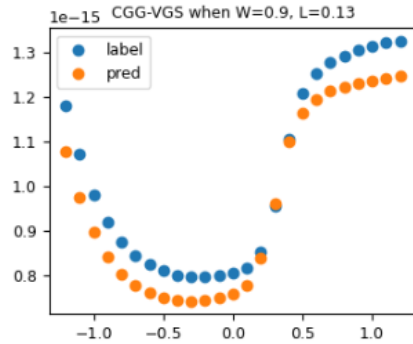
Generate variation model for corner model or Monte Carlo simulation

# Model generation with scarce data

**W=0.15 $\mu$ m,  
L=1.0 $\mu$ m**



**W=0.9 $\mu$ m,  
L=0.13 $\mu$ m**



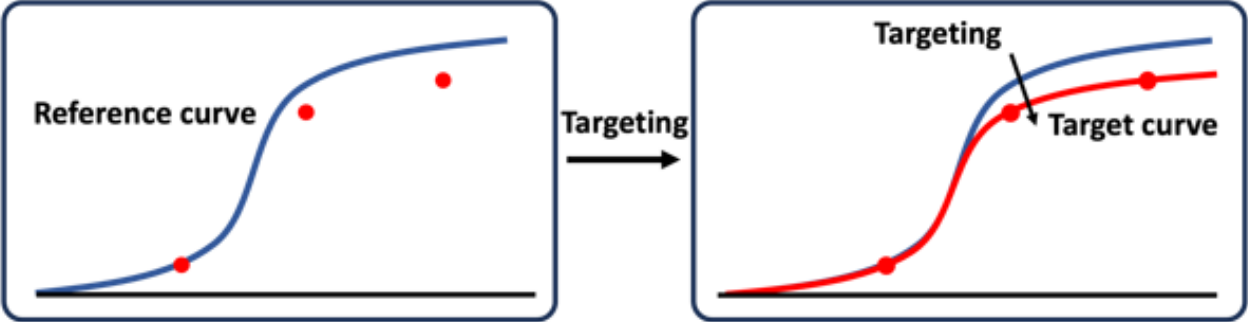
**Cgg**

**Cdd**

**Cbs**

- ✓ We showed Cgg/Cgc curve @ 10 $\mu$ m/10 $\mu$ m, and 10 $\mu$ m/0.04 $\mu$ m
- ✓ It infers every capacitance values in function domain.

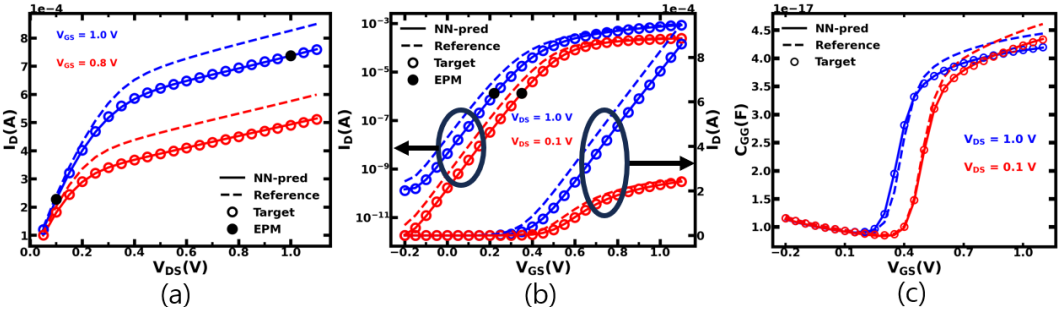
# Model re-targeting



Adjust function with target points

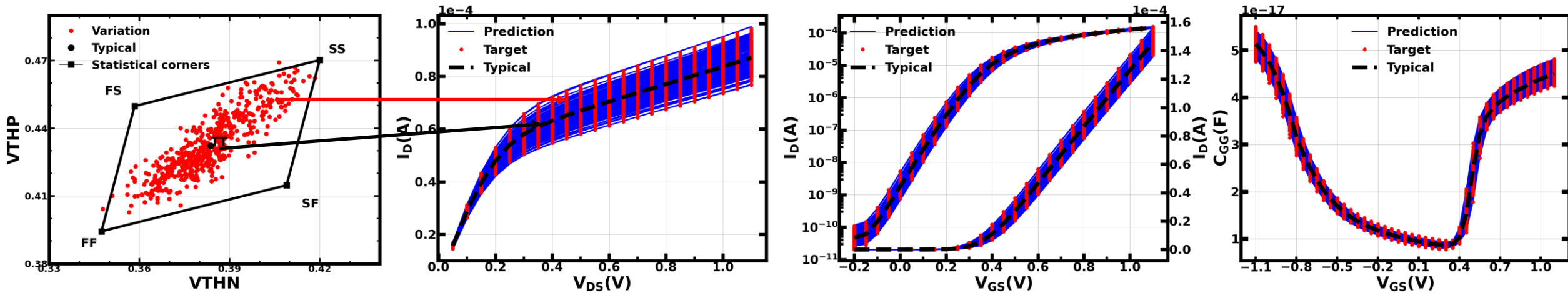
TABLE III: EPM changes

EPM	Reference	Target	Ratio
$I_{DSAT}$	0.83 mA	0.74 mA	-10.7 %
$I_{DLIN}$	0.24 mA	0.23 mA	-5.4 %
$V_{TSAT}$	0.16 V	0.22 V	36.5 %
$V_{TLIN}$	0.30 V	0.35 V	17.6 %
$V_{TEXT}$	0.37 V	0.42 V	14.0 %





# Variation model



If we can manipulate function, we can make variation model for fixed corner or Monte Carlo simulation

## Status & Plan

- ✓ We will integrate foundation model in our modeling solution, Alsis within 2024.
- ✓ Compared to equation model, modeling become easy and fast, 100% accuracy for measured data point, learned physics based inference when data is scarce.
- ✓ Interfacing 3<sup>rd</sup> party SPICE simulator through Verilog-A interface, fast and robust simulation through patented computation algorithm.
- ✓ Applicable to CMOS or other physics sharing devices.
- ✓ We are working on power devices, TFT, memristors.  
(It's a little bit tricky if model has internal states)

# How AI impact engineering

AlphaGo (2016)



**Message:**  
"AI will win every well defined optimization game"

**Limitation:**  
"In most cases, engineering problems cannot be mathematically defined, relying on human knowledge"

ChatGPT (2022)




**Message:**  
"Every human knowledge can be pre-trained"


**Device physics can be pre-trained, maybe there are many other opportunities**

**Alsemy, Inc.**

# **Thank You Q&A**

---

 Hyunbo Cho

 [hyunbo.cho@alsemy.com](mailto:hyunbo.cho@alsemy.com)