Convolutional Networks for Co-Optimization of IVR and Embedded Inductor for 2.5D Packaging

Students: Hakki M. Torun, Huan Yu
Faculty: Madhavan Swaminathan

Acknowledgements

This work was funded in part by:

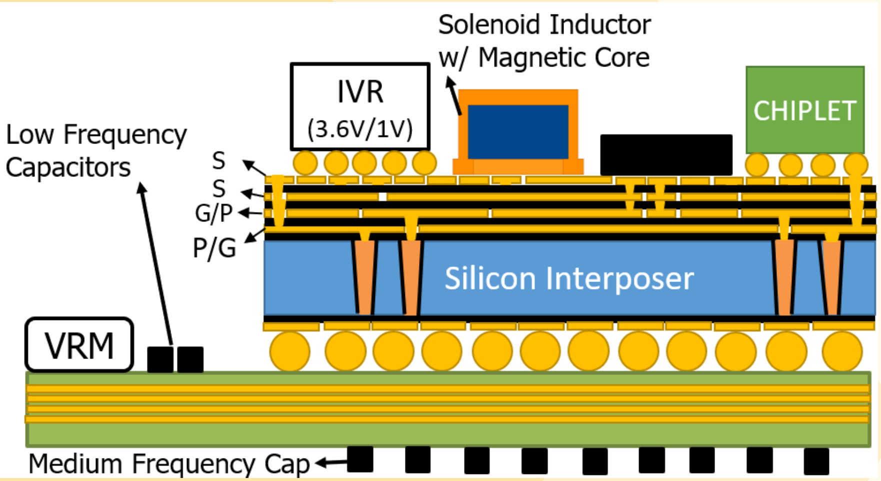
- NSF I/UCRC Center for Advanced Electronics Through Machine Learning (CAEML)
- DARPA CHIPS

Liaisons: Prof. Sung-Kyu Lim (GT), Prof. Saibal Mukhopadhyay (GT)

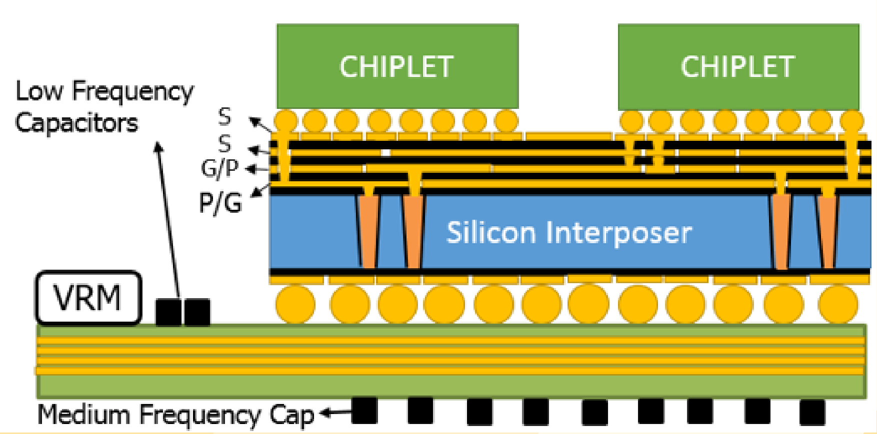
Objective

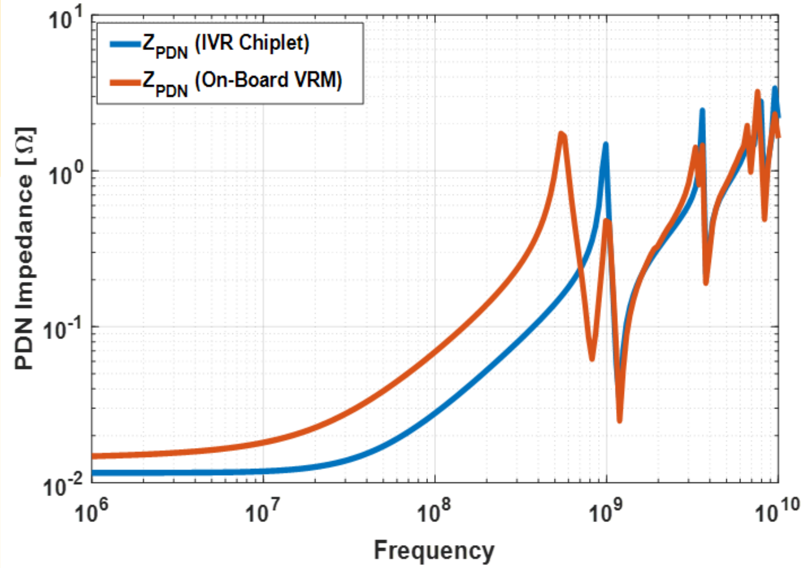
Power Delivery for 2.5D integrated systems

On-Interposer IVR Chiplet



On-Board VRM

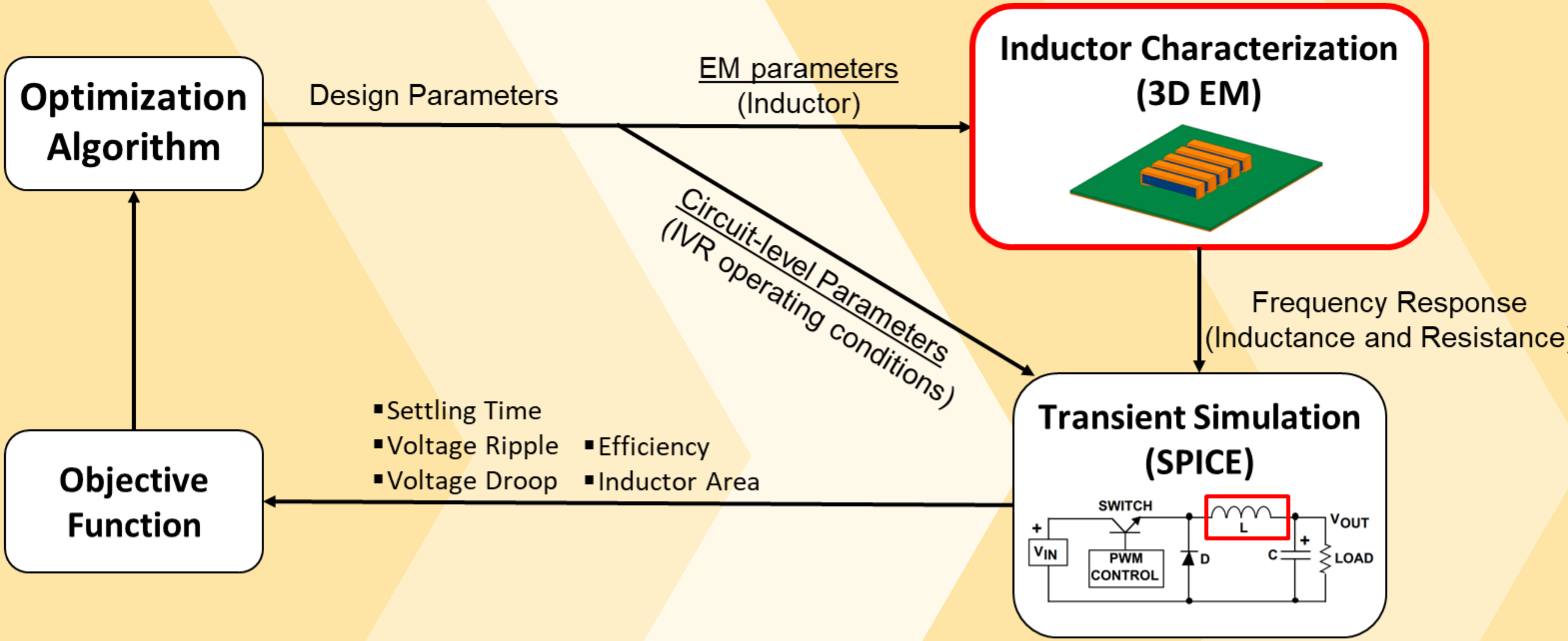




- 2 possible solutions for power delivery in 2.5D integration are: **on-board VRM and on-interposer IVR.**
- IVRs have significantly better transient performance for 2.5D packaging.
- The drawback is <u>reduced efficiency and increased area on-interposer.</u>
- **Challenge:** How to optimize IVR & VRM to increase efficiency & reduce inductor area?

Objective

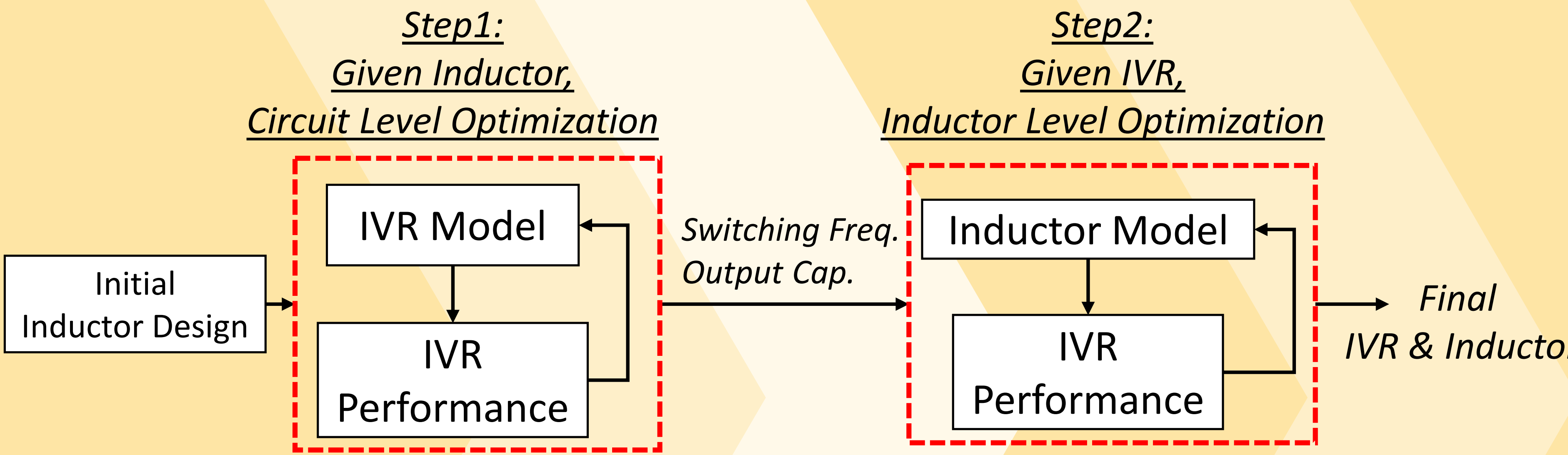
Challenges in IVR & Inductor Co-Optimization



- Inductor design trade-offs can not be determined without IVR operating conditions.
- IVR operating conditions can not be determined without inductor characteristics.
- <u>Co-Optimization is required</u>, but generally avoided due to optimization complexity.
  - Simulation = EM Characterization + Transient Analysis.
- <u>Bottleneck in optimization is CPU intensive 3D EM simulations.</u>

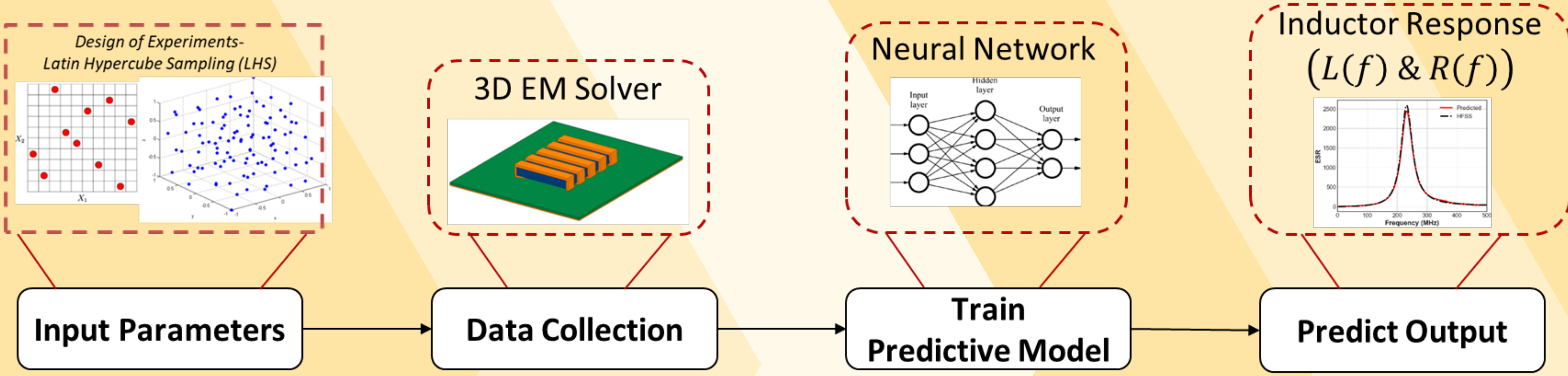
Prior Work

Two-Step Optimization for Embedded Inductor & IVR



- Inductor design trade-offs can not be determined without IVR operating conditions.
- IVR operating conditions can not be determined without inductor characteristics.
  - Different inductors can perform better with different IVR parameters.
- Two-Step optimization then becomes sub-optimal.
- Co-optimization is required, but usually avoided due to complexity.

Surrogate Model Development for Embedded Inductor

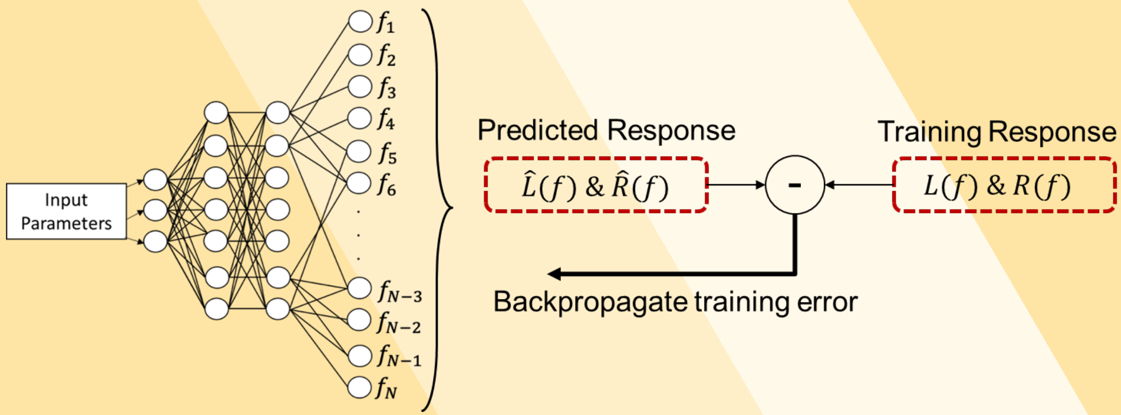


- Moderate amount of data is collected from the actual 3D EM solver.
- The data is then used to train a predictive model.
  - Learning-based models are preferred for being universal approximators.
- Trained model can then be used in any optimization loop very efficiently.
  - <u>Replace EM simulation with the trained model!</u>

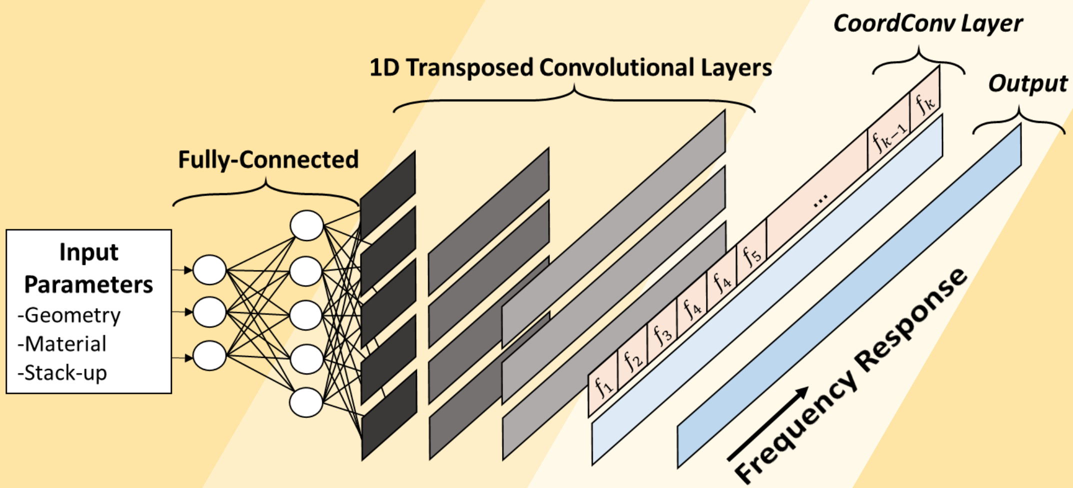
Proposed Approach

Building Learning-based Model

Conventional Fully-Connected Neural Network (FC-NN)



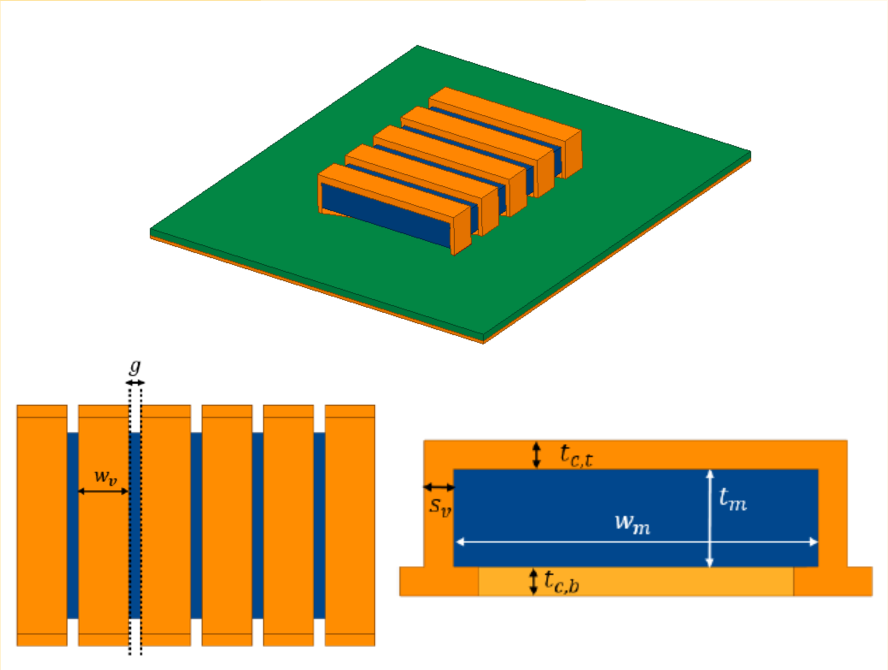
Proposed Spectral Transposed Convolutional Networks (S-TCNN)



Proposed Loss Function: $L_{freq} = \frac{1}{N}\sum_{n=1}^{N}\sqrt{\frac{1}{K}\sum_{k=1}^{K}(y_{n,k} - \hat{y}_{n,k})}$

- FC-NN is one of the most commonly used approach to predict freq. responses.
- # of learnable parameters increase exponentially when # freq. points increase.
- Proposed S-TCNN exploits spatial correlation in the frequency axis.
- Design parameters are passed through FC layers.
- The latent space is then passed through 1D transposed convolutional layers to construct predicted freq. response.

Model for On-Interposer Solenoid Inductor with Magnetic Core

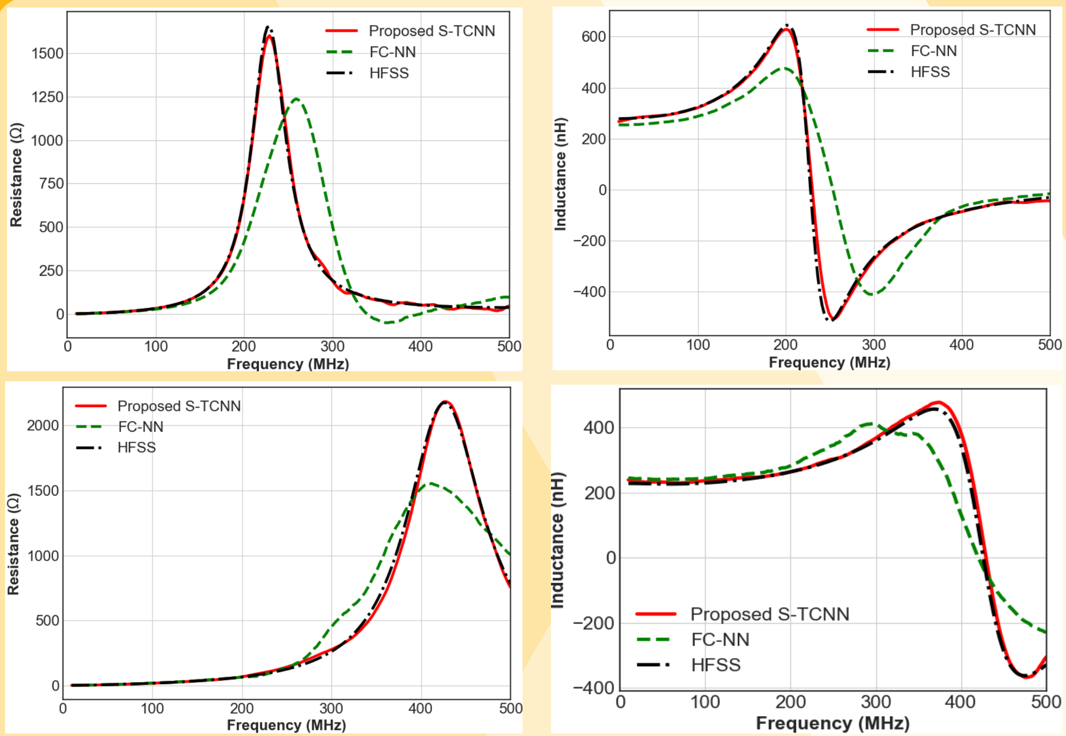


Parameter		Unit	Min	Max
Gap between windings	g	mil	2	20
Number of windings	N		3	13
Size of via	s_v	μm	50	103
Copper Trace Width	w_c	mil	2	20
Copper Thickness Bottom	$t_{c,b}$	μm	35	170
Copper Thickness Top	$t_{c,t}$	μm	35	170
Magnetic Core Thickness	t_d	μm	50	650
Magnetic Core Width	w_d	μm	50	350

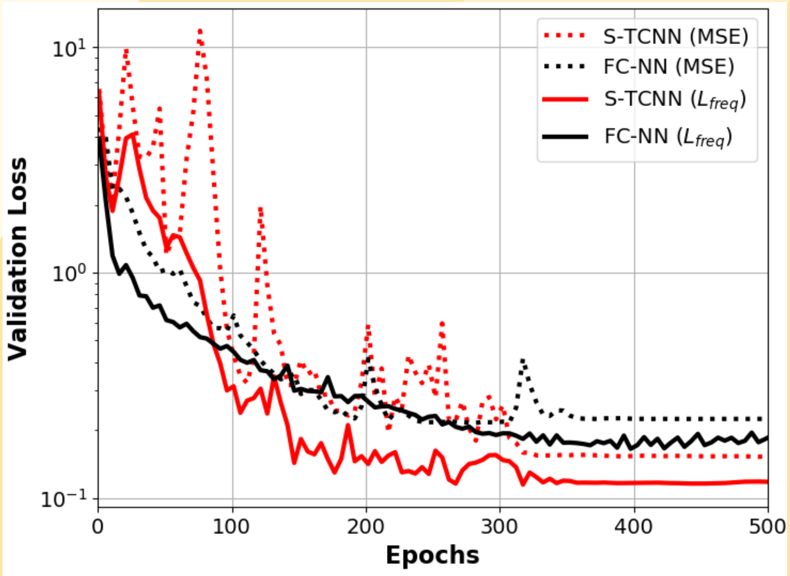
- Solenoid inductor with NiZn magnetic core is considered.
  - Integrated alongside the chiplets on interposer.
- 8 parameters define the geometry of the inductor.
- Inductance and resistance between 10 MHz and 500 MHz at 200 freq. points.
- 1000 data points based on Latin Hypercube Sampling (800 training, 200 test)

Results:

Performance of S-TCNN for Learning Inductance and Resistance

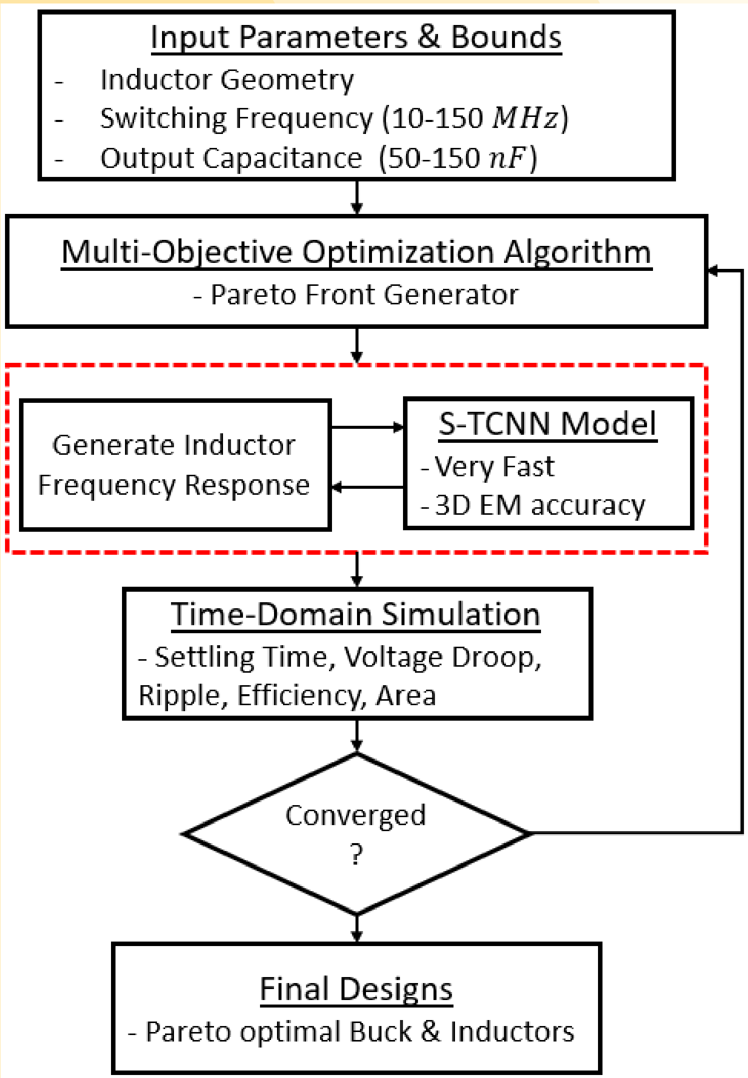


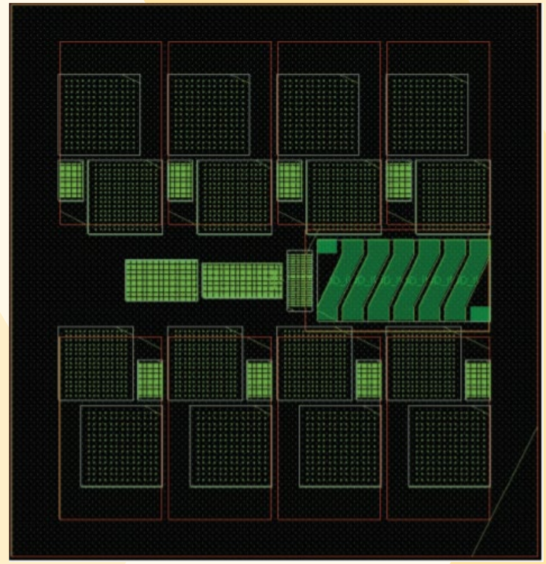
	FC-NN		Proposed S-TCNN	
	MSE	L_{freq}	MSE	L_{freq}
Validation NMSE	0.228	0.177	0.152	0.120
Run Time	0.001 sec		1.503 sec	
(for 1k freq. responses)	*(HFSS: ∼ 25 hours)*			

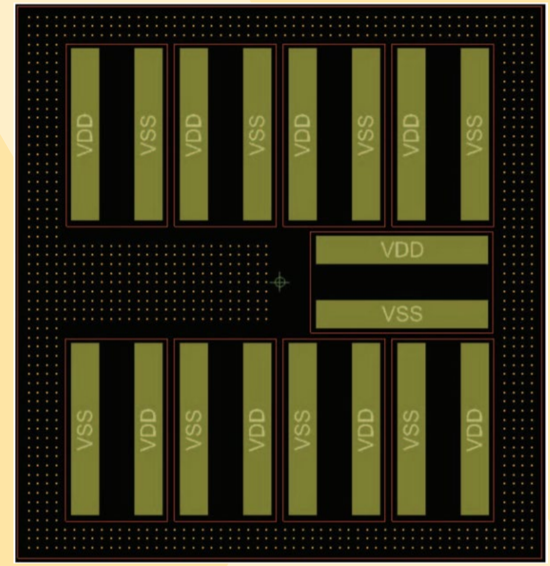


- S-TCNN is compared to regular FC-NN
- **10.8% improvement** in predictive accuracy as compared to FC-NN with MSE loss.
- Proposed loss function increased accuracy of FC-NN by 5.1% and S-TCNN by 3.2%.
  - Convergence of test error is also faster → better generalization.

Generating Pareto Front for IVR





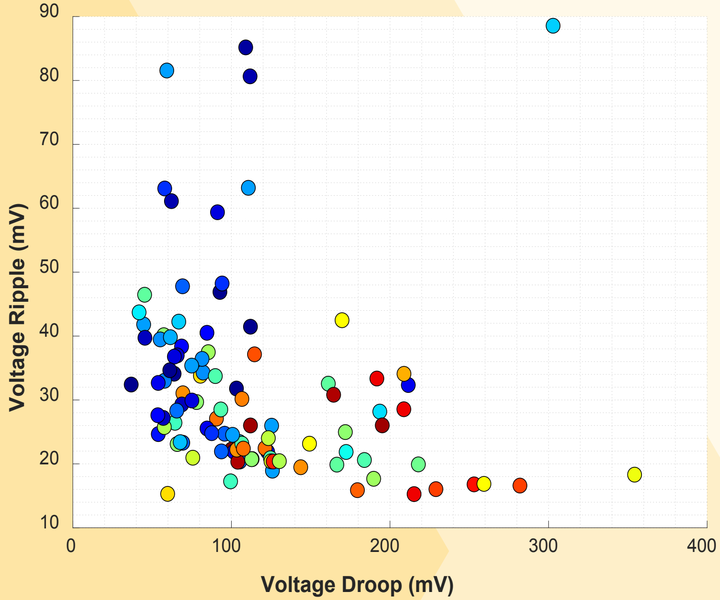


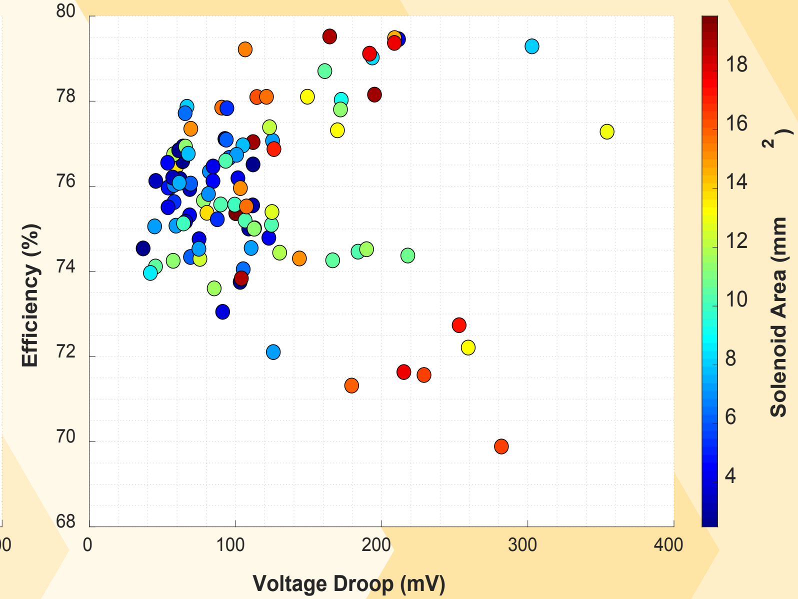
Kim et al. "Architecture, Chip, and Package Co-design Flow for 2.5D IC Design Enabling Heterogeneous IP Reuse", DAC'19.

- Switching frequency (10-150 MHz) and output capacitance (50-150 nF) included as parameters of IVR.
- Total of 10 input parameters and 5 objectives.
- The floorplan is fixed and corresponding PDN parasitics are included in time-domain simulations.
- NSGA-II is used to generate Pareto Front.

Results:

5-dimensional Pareto Front

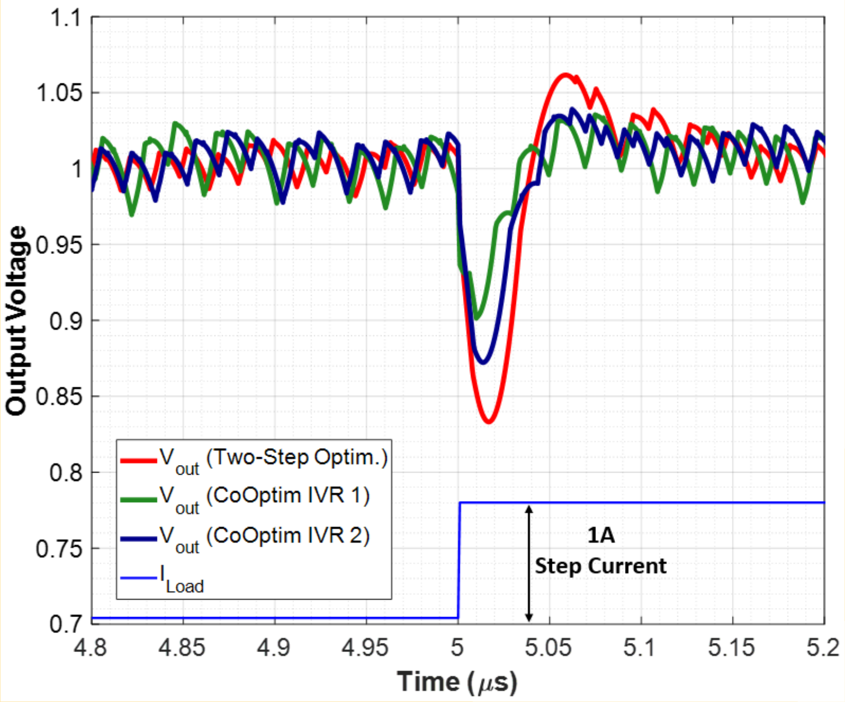




	T_{set}	V_{droop}	V_{ripple}	Eff.	Ind. Area
T_{set}	1.00	0.23	-0.19	0.38	0.09
V_{droop}	0.23	1.00	0.45	-0.30	0.07
V_{ripple}	-0.19	0.45	1.00	-0.47	-0.37
Eff.	0.38	-0.30	-0.47	1.00	0.19
Ind. Area	0.09	0.07	-0.37	0.19	1.00

- Each point in the Pareto front is optimal, but prioritize different objectives.
- 105 Pareto optimal designs are generated.
- Optimal trade-offs can be seen from pair-wise plots and correlation matrix.
  - Ex: Conversion efficiency and settling time; inductor area and efficiency & voltage ripple.

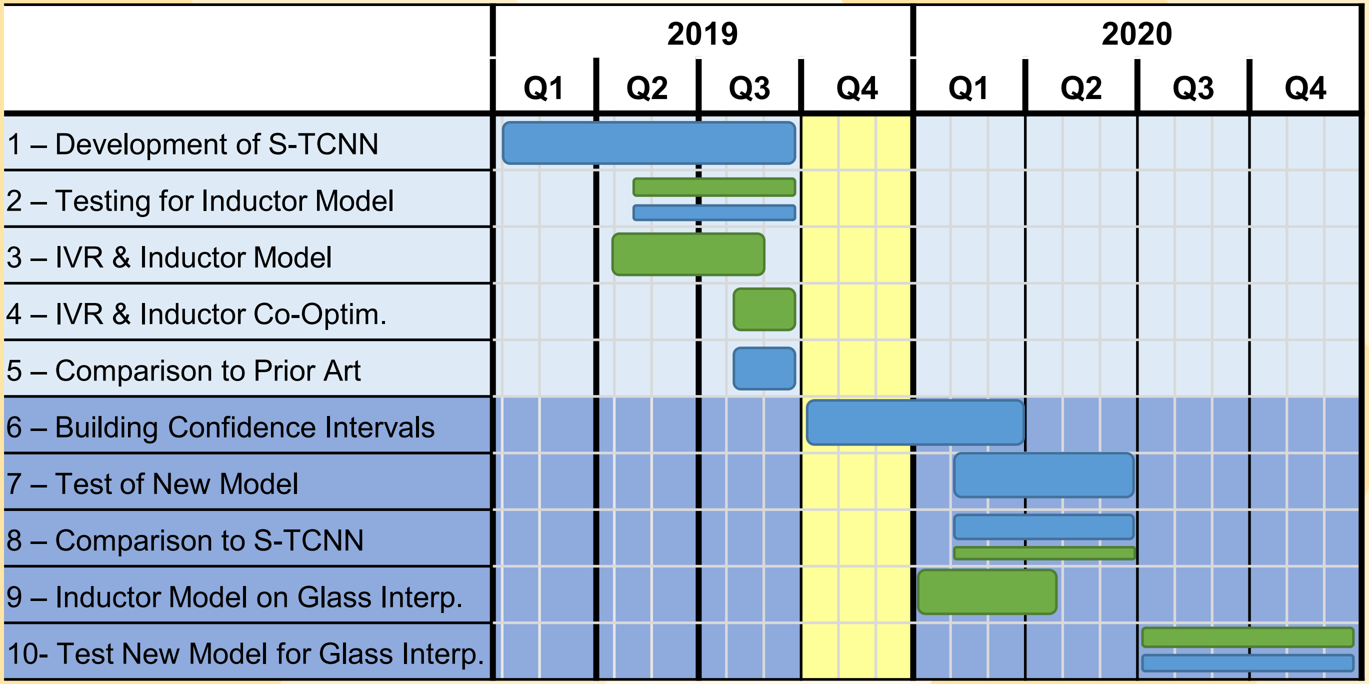
Results:
Comparison to Two-Step Optimization



	Two-Step Optimization	Co-Optimized IVR 1	Co-Optimized IVR 2
Switching Freq.	125 MHz	100 MHz	115 MHz
Capacitance	100 nF	115 nF	128 nF
Inductance	29.8 nH	20.7 nH	23.8 nH
ESR	3.63 Ω	1.01 Ω	1.12 Ω
DC Resistance	10.5 mΩ	15.7 mΩ	30.2 mΩ
Area	5.12 mm^2	4.64 mm^2	2.48 mm^2
Efficiency	76.6 %	77.8 %	76.3 %
Voltage Droop	167 mV	98.6 mV	127 mV
Voltage Ripple	38.8 mV	49.3 mV	40.2 mV
Settling Time	115 ns	80 ns	75 ns

- Co-Optimization is compared to a thorough Two-Step Optimization.
- Two designs are selected to prioritize performance (IVR1) and inductor area (IVR2).
- IVR2 have 51.6% reduced area with 40 ns faster settling time compared to Two-Step optimization.
- IVR1: 9.8% reduced area with 40.9% less voltage droop, 26.1% less settling time and 1.2% more efficiency.
- Other designs can also be selected from the generated Pareto front to prioritize other objectives.

Timeline



Light blue: ML Model development and application to power delivery
Dark blue: New model development and apply to glass-interposer
Light Yellow: Current time window

Application to IVR & Embedded Inductor
ML Model Development

Summary

- Introduced Spectral Transposed Convolutional Networks (S-TCNN) to predict frequency responses.
  - First use of convolutional networks to handle frequency responses in EDA.
- Transposed convolutional layers are shown to be effective to upsample design parameters to their corresponding freq. domain characteristics.
- Proposed a new loss function to increase generalization capability of neural networks.
  - Both for S-TCNN and regular fully-connected nets.
- Overall, S-TCNN showed 10.8% better predictive accuracy compared to conventional models in EDA.
- Used the derived model for IVR & inductor co-optimization, and achieved up to:
  - **51.5% reduced inductor area**
  - **40.9% reduced voltage droop**
  - **26.1% reduced settling time**

  compared to Two-Step Optimization.

H. M. Torun et al.,
"A Spectral Convolutional Net for Co-Optimization of Integrated Voltage Regulators and Embedded Inductors", ICCAD'19