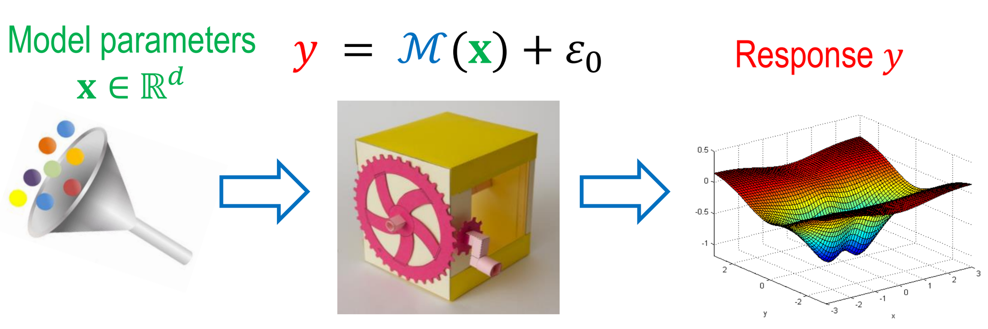
Bridging the Gap Between Artificial Neural Networks and Kernel Regressions in EM Applications

Nastaran Soleimani, Riccardo Trinchero , Flavio Canavero
Dept. Electronics and Telecommunications, Politecnico di Torino
Nastaran.soleimani@polito.it, http://emc.polito.it

Research context and motivation

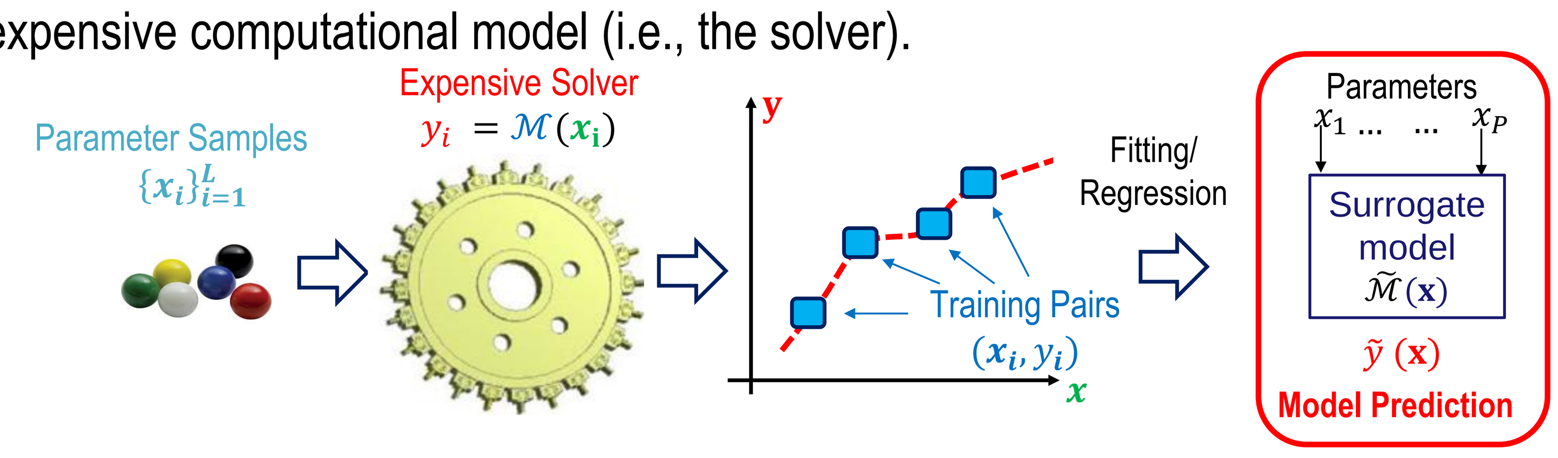
- **Optimization** and **uncertainty quantification** are key ingredients for the design of microwave structures and electronic devices.
- Such tasks are usually carried out synthetically via computer experiments (simulations), based on the **computational model**
- **Computational model** is a procedure (e.g., a solver) able to compute quantities of interest from the input parameters (e.g., geometrical/electrical parameters).



WARNING
The **computational cost** of the computational model can be **huge!!!**

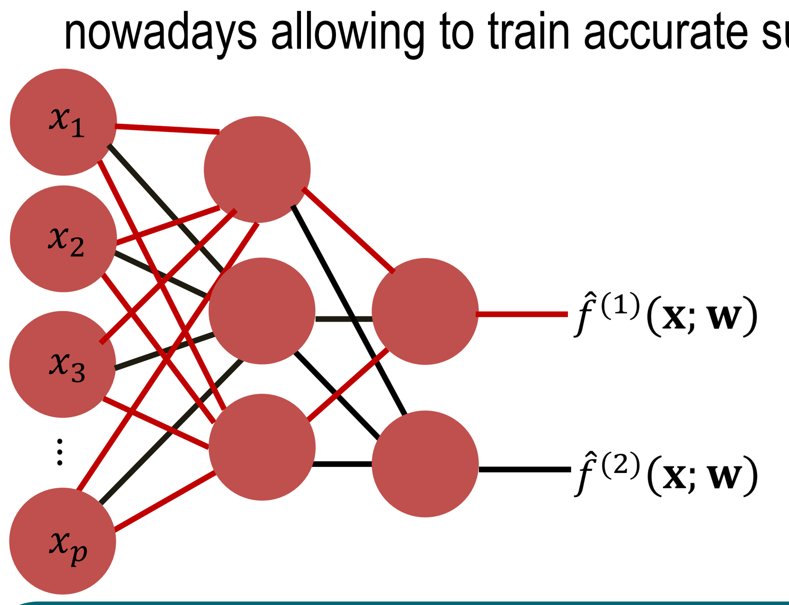
Addressed research questions/problems

- **Surrogate model** $\widetilde{\mathcal{M}}$ is "a model of a model", i.e., a model of the computationally expensive computational model (i.e., the solver).



- The surrogate model **accuracy** in complex non-linear problem with dozens input parameters depends on the fitting or regression techniques used to train it.

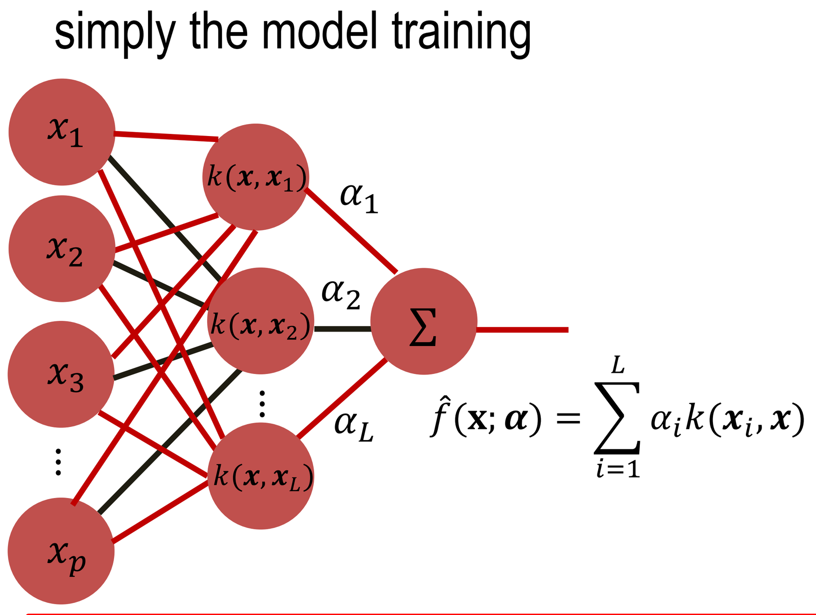
➢ **Artificial Neural Network (ANN)** is the most popular Machine Learning method nowadays allowing to train accurate surrogate model.



PROS:
- NON linear model structure
- Flexible topology
- Natural extension to multi-output

CONS:
- **NON-convex optimization problem**
  ⟶ Hard to train, data-hungry

➢ **Kernel regression** provides a clever alternative to ANN structure allowing to heavily simply the model training



PROS:
- Linear model structure
- Convex optimization problem
  - Fast to train
  - Fast convergence w.r.t. training samples

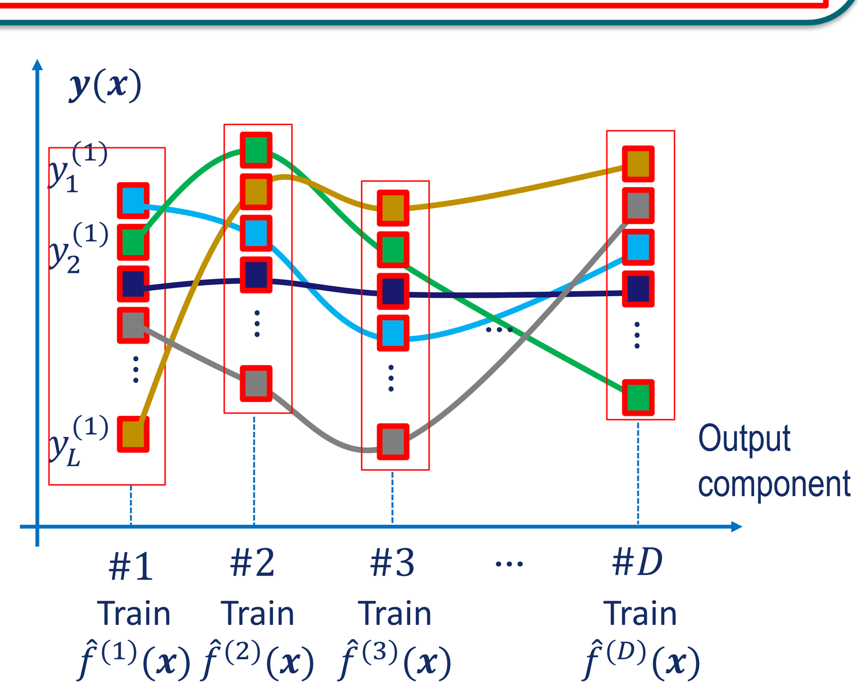
CONS:
- Fixed topology
- **Scalar-valued methods**

WARNING: Most of the EM applications require a MULTI-OUTPUT formulation!!!

- **Multi-Output Scenario & Scalar Regression**

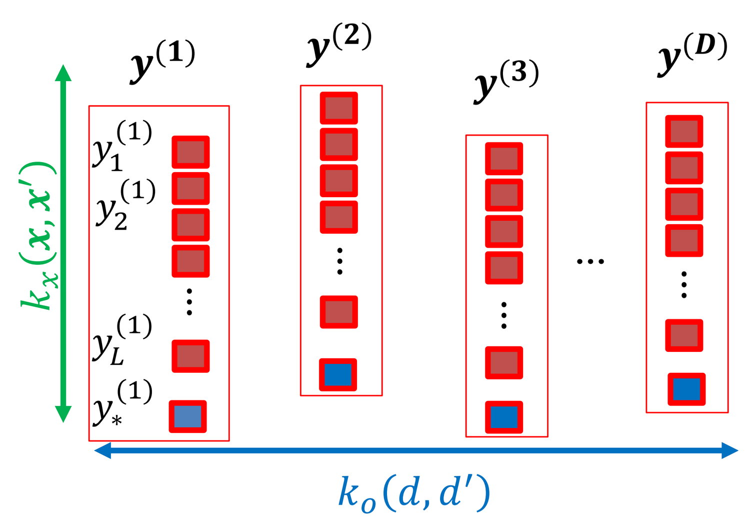
IDEA:
Use a scalar regression for each output components

Too many models and hyperparameters to tune!!!
No protection against noise!!!



Methodology

- **Multi-Output Kernel Ridge Regression:**



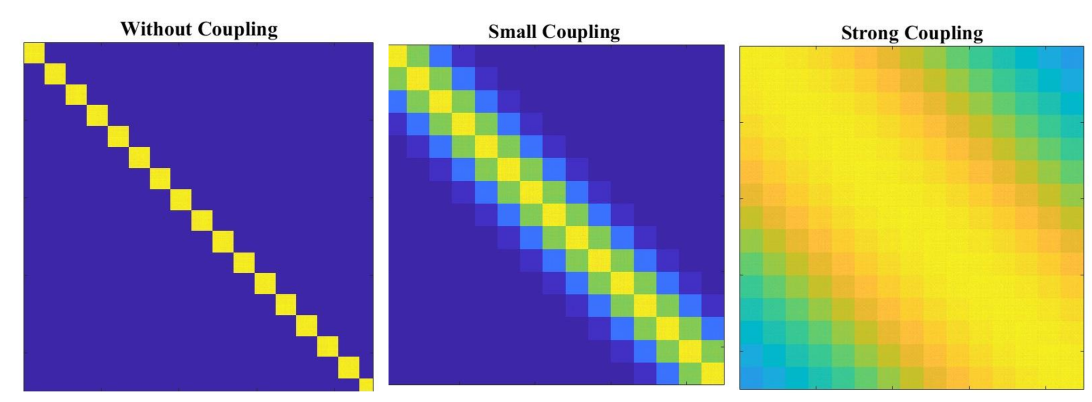
$$\hat{\boldsymbol{f}} = \underset{\bar{\boldsymbol{f}} \in \mathcal{H}}{\operatorname{argmin}} \sum_{d=1}^{D} \sum_{l=1}^{L} \left(y_l^{(d)} - \bar{f}^{(d)}(\boldsymbol{x}_l) \right)^2 + \lambda \|\bar{\boldsymbol{f}}\|^2$$

$$\hat{f}^{(d')}(\boldsymbol{x}_*) = \sum_{d=1}^{D} \sum_{i=1}^{L} k\big((\boldsymbol{x}_i, d), (\boldsymbol{x}_*, d')\big) c_{i,d}$$

where $k\big((\boldsymbol{x}_i, d), (\boldsymbol{x}_*, d')\big)$ is a **"new" kernel function** acting on both the input space and output components

E.g., for a **separable kernel**:

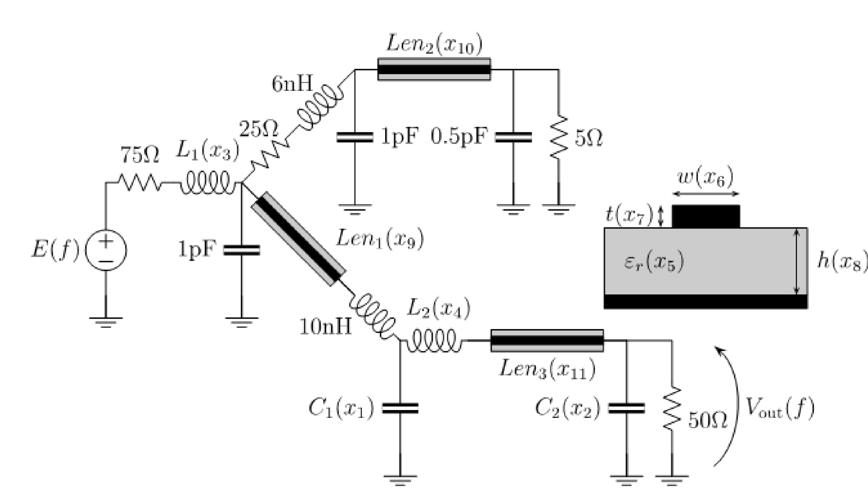
$$k((\boldsymbol{x}, d), (\boldsymbol{x}', d')) = k_x(\boldsymbol{x}, \boldsymbol{x}') \cdot k_o(d, d')$$



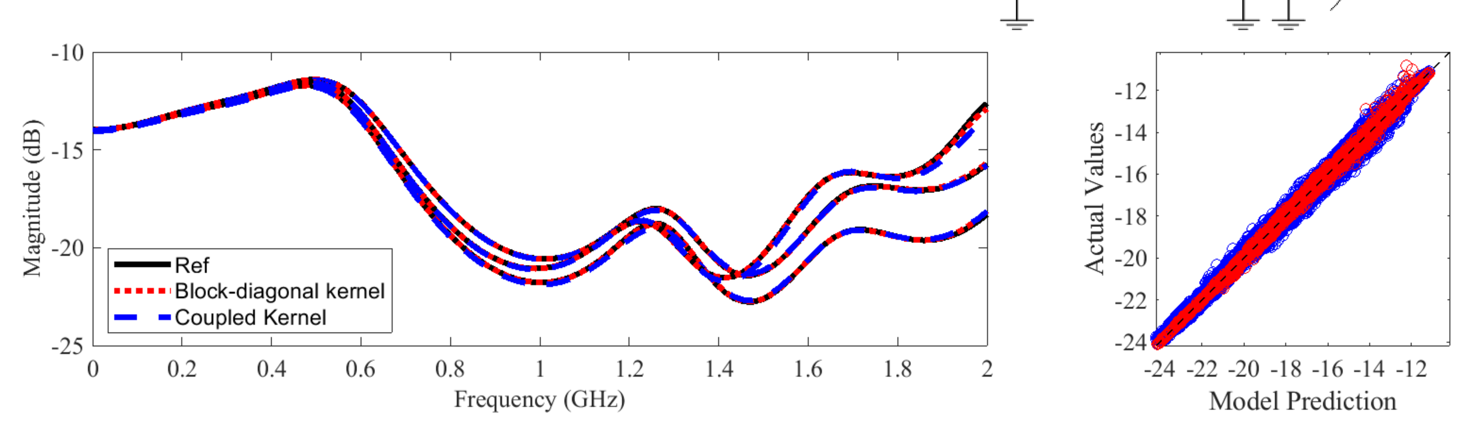
Novel contributions

- **Example I**: High Speed Link

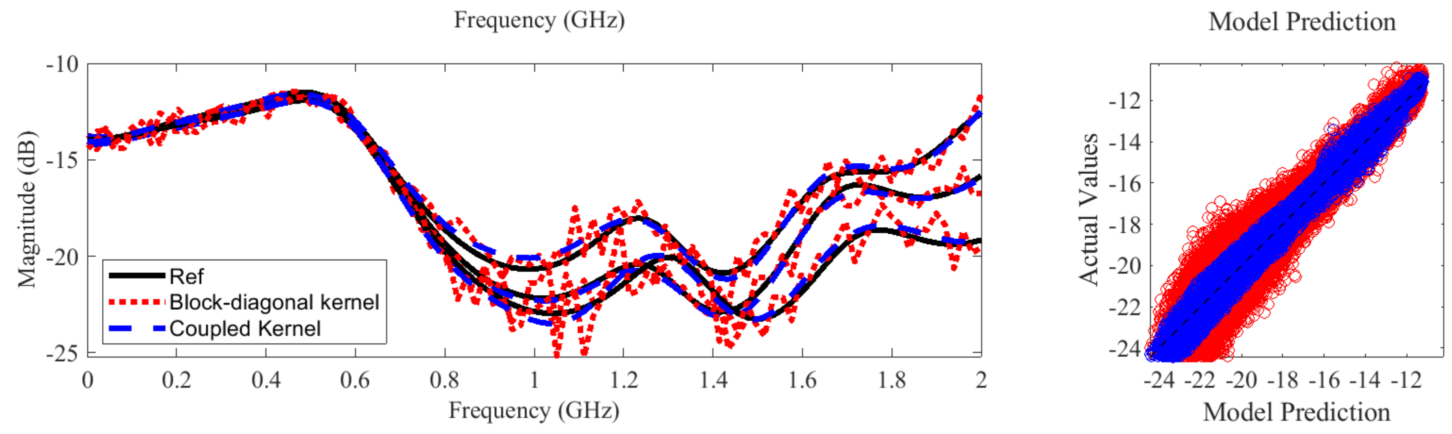
GOAL: Build a surrogate model for finding the magnitude of the frequency response of the function $y(f;x) = \frac{Vout(f;x)}{E(f)}$ as a function of 11 **circuit parameters with huge variability.**



- **Noise Free (training samples)**

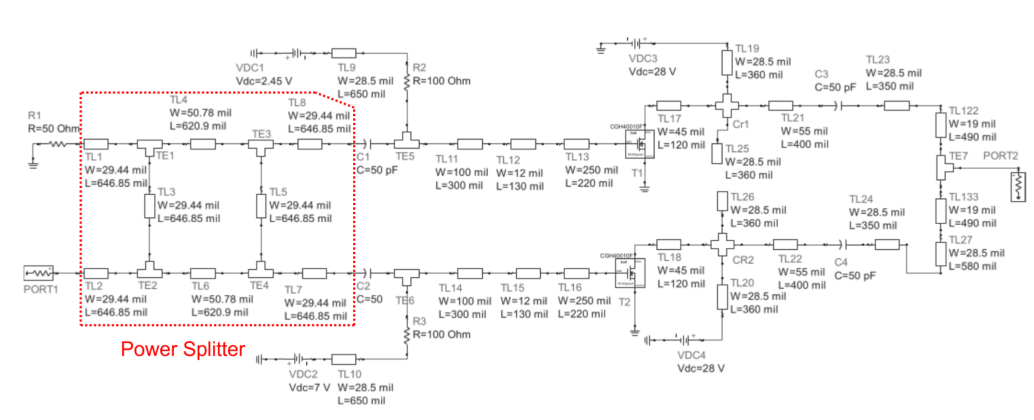


- **Noisy (training samples)**



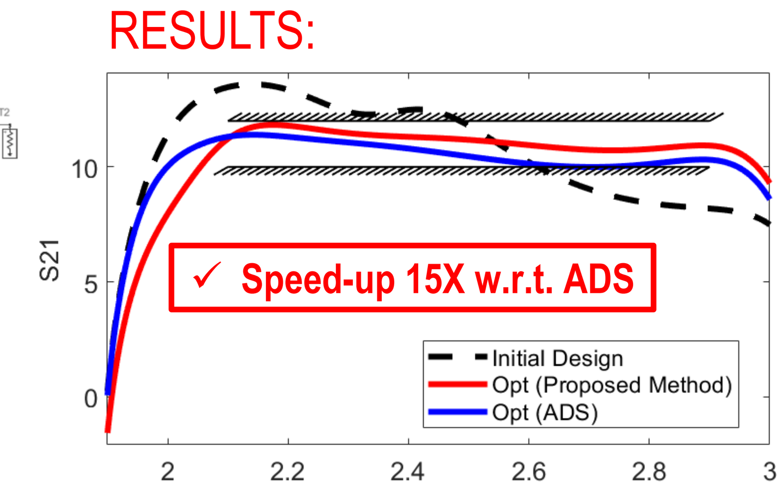
- **Example II** : Doherty Power Amplifier for Wireless Applications

GOAL: Optimize the power splitter such that: $10\text{dB} \leq S_{21}(f) \leq 12\text{dB}$ for $f \in [2.1, 2.9]$GHz



20 parameters

RESULTS:



Future work

- How can we reduce the computational cost of the training phase?
- How can we automatically obtain the optimal structure of the multi-output kernel?

POLITECNICO DI TORINO

EMC Group