Kernel-Based Machine Learning Surrogates for RF Circulator Design

Kernel-Based Machine Learning Surrogates for RF Circulator Design

Marco Atlante[1,2], Riccardo Trinchero[1], Igor S. Stievano[1],
Mihai Telescu[2], Norbert Parker[2], Vincent Laur[2], Noël Tanguy[2]
(1) Politecnico di Torino, 10129 Torino, Italy
(2) Univ Brest, Lab-STICC, CNRS, UMR 6285, F-29200 Brest, France
E-mail: {marco.atlante, riccardo.trinchero, igor.stievano}@polito.it,
{mihai.telescu, norbert.parker, vincent.laur, noel.tanguy}@univ-brest.fr

***Abstract* — Machine learning has attracted significant interest in the microwave community in recent years. It has the potential to greatly enhance design flows, but the choice of the algorithm and, more importantly, of the deployment strategy are nontrivial and may seem intimidating with respect to traditional approaches. This paper is intended as a proof of concept. It demonstrates a simple, effective, and elegant way to build and use feature-based surrogate models in order to improve the design process of RF circulators. The models are able to predict key metrics of the frequency response of the device, depending on its geometry. These models are then used to select configurations that meet the desired specifications.**

***Keywords* — Microwave Circulators, Surrogate Modeling, Kernel Ridge Regression, Design Automation, Machine Learning**

I. Introduction

The present paper proposes an original methodology in circulator design, employing surrogate models built using a machine learning algorithm. Traditionally, circulator design starts from a set of application-specific specifications. These usually include bandwidth, size, losses, etc. As an example Table 1 summarizes a representative set of requirements. Based on simplified models and heuristics, the designer selects a technology, defines an initial geometry, and simulates it with a full-wave solver. The design is gradually refined to meet the specifications through trial-and-error and the use of optimizers built in the full-wave solver. Once the target is reached, the process moves to the prototyping phase. When faced with a new application and new specifications, the process is repeated. The alternative flow proposed in this work aims to drastically reduce computational effort. First, one or more surrogate models are trained using a limited set of full-wave data. This is a one-time computation effort, since the resulting surrogate models can be stored. Whenever a set of specifications is provided, feasible solutions are identified by querying the surrogate models, which basically act as parametric predictors mapping geometry to key performance metrics.

II. Feature extraction for Circulator Design

Circulators are passive non-reciprocal devices, commonly used as duplexers in single-antenna full-duplex systems. Their non-reciprocity relies on a magnetically saturated ferrite, whose permeability becomes anisotropic. Traditional implementations use "soft" ferrites polarized by a permanent magnet, which ensures good electrical performance. However, they result in bulky and costly devices due to the hybrid assembly of ferrite pucks into stripline or microstrip structures. A more compact alternative employs pre-oriented hexaferrites ("hard" ferrites), which retain strong remanent magnetization and high anisotropy without an external field. This enables self-biased (magnetless) circulators, achieving volume reductions above 90% compared to conventional solutions [1], [2].

Table 1. Representative Design Scenario Specifications

Specification	**Scenario**
Relative bandwidth (100·bw/central frequency)	3%
Central frequency (i.e., reference frequency)	18 GHz
Isolation/Adaptation (minimum in-band value, i.e., reference level)	15 dB
Maximum in-band insertion losses	1.2 dB
Maximum size	$7.5\,\mathrm{mm}^2$
Minimum line width	20 µm

This work focuses on self-biased circulators, but all developments are readily adaptable to other technologies and topologies with minimal effort. To implement the flow in Section I, we first construct a surrogate model that predicts performance metrics (e.g., maximum in-band loss, bandwidth) from geometry alone. The training strategy relies on an a priori exploration of the design space via full-wave simulations of multiple geometries. This one-time effort avoids further full-wave calls during optimization or selection. A total of 400 geometries were generated using Latin Hypercube Sampling (LHS), which enables efficient coverage of the parameter space. Each geometry is defined by three variables: resonator radius, impedance matching line length, and width (see insert in Fig. 2), with ranges reported in Table 2. The

Table 2. Design parameters for the circulator

Design Parameter	**Radius** [µm]	**Length** [µm]	**Width** [µm]
Base Value	950	45	100
Range	±100	±25	±95

high-fidelity full-wave simulations are performed in Ansys HFSS™, producing S-parameter responses across frequency for all LHS-sampled geometries. As shown in Fig. 1, these curves illustrate the impact of geometric variation on device behavior. Rather than using the full S-parameter traces as targets, we adopt a feature-based approach.

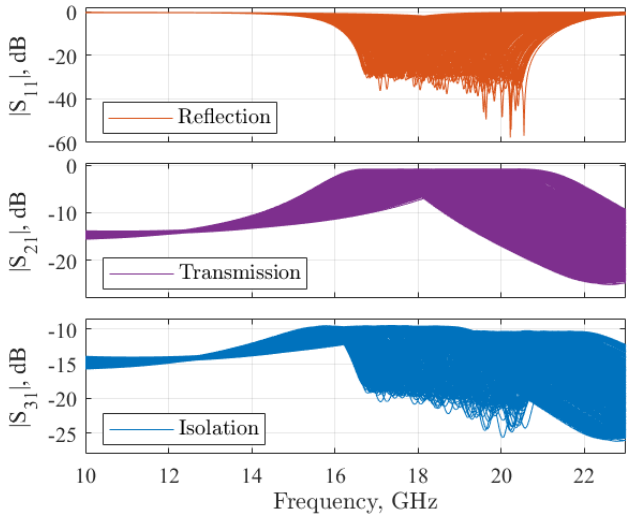

Fig. 1. S-parameter responses across the sampled design space. The three panels show reflection ($|S_{11}|$), transmission ($|S_{21}|$), and isolation ($|S_{31}|$), respectively, each in dB.

From each simulation, we extract a small set of scalar values that capture performance aspects of direct interest to the designer. These features reflect application-level specifications listed in Table 1. Among them, the effective bandwidth plays a central role in performance evaluation. To compute it, we apply a fixed isolation/adaptation threshold ($-15\,$dB, as defined in Table 1) to the S_{11} and S_{31} curves. The portion of interest, defined along the threshold is bounded by the two crossing points closest to the resonance peak, regardless of whether they occur on S_{11} or S_{31}. These two points, f_{BWlow} and f_{BWhigh} (magenta squares in Fig. 2), define the effective bandwidth: $BandWidth = f_{BWhigh} - f_{BWlow}$. Because the algorithm generates quasi-random geometries, in some cases the bandwidth may be 0 Ghz, i.e. either S_{11} or S_{31} is always above the selected threshold. Should this occur, the sample is discarded and not used in the training stage.

For each geometry, the reference frequency f_{res} is redefined as the midpoint: $f_{res} = (f_{BWlow} + f_{BWhigh})/2$. It is marked in Fig. 2 with a red cross. This threshold-based scheme translates application-level specs into a compact, meaningful representation of device behavior, enabling constraint filtering and performance-driven optimization.

A second constraint checks that the in-band insertion loss stays within the specified limit across the entire effective bandwidth. We locate the two frequencies where the S_{21} curve crosses the insertion loss threshold ($-1.2\,$dB in Table 1). These points, f_{ILlow} and f_{ILhigh} (blue circles in Fig. 2), define the interval: $MaxLossBand = f_{ILhigh} - f_{ILlow}$.

We then check whether $MaxLossBand \geq BandWidth$, ensuring the insertion loss remains below threshold across the entire span. Since both intervals are centered around the reference, comparing their lengths suffices.

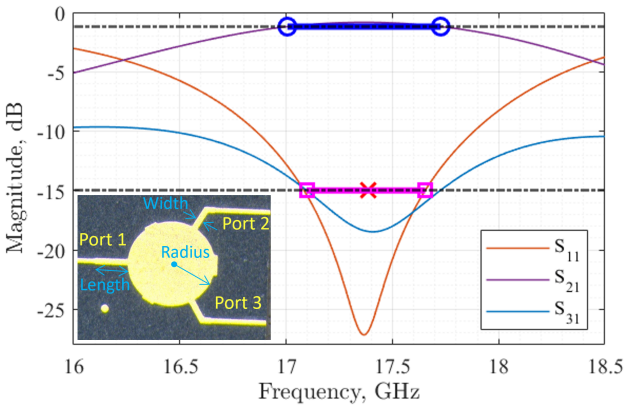

Fig. 2. Simulated S-parameter curves with highlighted performance segments. The magenta segment identifies the effective bandwidth, delimited by f_{BWlow} and f_{BWhigh}, and centered at f_{res} (red cross). The blue segment indicates the frequency range where the insertion loss constraint is satisfied, bounded by f_{ILlow} and f_{ILhigh}. Insert: example of circulator geometry featuring resonator radius, impedance matching line length and width.

From each valid simulation, we obtain an input vector $\boldsymbol{x}_n = [\text{Radius, Width, Length}]^T$ and an output vector $\boldsymbol{y}_n = [\text{BandWidth, } f_{\text{res}}\text{, MaxLossBand}]^T$, with $n = 1, \ldots, N$ and $N = 400$.

III. Modeling via Kernel Ridge Regression

The dataset defined in Section II is then split into disjoint training and test subsets. We use the first $L = 300$ samples for training, and reserve the remaining 100 samples for testing. The training subset $\mathcal{D}_{\text{train}} = \{(\boldsymbol{x}_l, y_l^{(i)})\}_{l=1}^{L}$ is used to build three independent models, one per output ($i = 1, 2, 3$).

All results presented in this paper were obtained using kernel ridge regression (KRR). KRR has a favorable bias-variance tradeoff, flexibility with ℓ_2 regularization and, above all, robustness in low-data regimes [3]–[5]. Each model predicts the i-th output as:

$$\hat{y}^{(i)}(\boldsymbol{x}) = \sum_{l=1}^{L} \alpha_l^{(i)}\, k(\boldsymbol{x}, \boldsymbol{x}_l), \tag{1}$$

where $k(\cdot,\cdot)$ is a positive-definite kernel and $\alpha_l^{(i)}$ are model weights. The weights are computed in closed form as:

$$\boldsymbol{\alpha}^{(i)} = (\mathbf{K} + \lambda \mathbf{I})^{-1}\, \boldsymbol{y}^{(i)}, \tag{2}$$

where $\mathbf{K} \in \mathbb{R}^{L\times L}$ is the kernel matrix with entries $K_{lj} = k(\boldsymbol{x}_l, \boldsymbol{x}_j)$, and $\boldsymbol{y}^{(i)} \in \mathbb{R}^L$ is the i-th output vector. The regularization parameter λ controls the trade-off between data fidelity and model smoothness, enhancing generalization on unseen inputs. We use a Matérn 5/2 kernel:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \left(1 + \sqrt{5}r + \frac{5}{3}r^2\right) \exp(-\sqrt{5}r), \tag{3}$$

with $r = \|\boldsymbol{x} - \boldsymbol{x}'\|/\sigma$, where σ is the kernel length scale. Model hyperparameters are tuned via Bayesian optimization

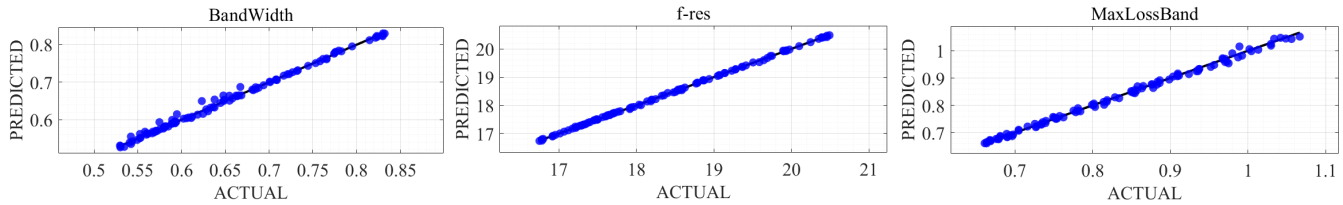

Fig. 3. Predicted versus actual values for the three scalar output features used in surrogate modeling: bandwidth (left), center frequency f_{res} (middle), and maximum loss (right). Each point represents one of the 100 samples in the test set.

(60 iterations, 3-fold CV) by minimizing the validation RMSE. Training of each model completes in about 10 s. The model's accuracy is visually assessed in Fig. 3, where scatter plots compare predicted and actual values on the 100 test samples. The alignment along the diagonal confirms the quality of the training.

IV. Candidate Selection via Surrogate Filtering

The trained surrogate model offers a fast, cost-effective alternative to full-wave simulation, enabling scalable design exploration, especially valuable in early stages, where evaluating thousands of candidates is impractical. To take advantage of this capability, we generate a large set of $M = 5000$ new geometries via LHS within the ranges of Table 2. These new input vectors, denoted as $\boldsymbol{x}_j$, are then evaluated using the surrogate model of Eq. (1), without the need to perform any additional full-wave simulations. This yields the virtual dataset $\mathcal{V} = \{(\boldsymbol{x}_j, \hat{y}_j^{(i)})\}_{j=1}^{M}$, where $\hat{y}_j^{(i)}$ denotes the i-th performance metric predicted for the j-th input samples (i.e., $\boldsymbol{x}_j$). The full evaluation takes less than one second and is followed by a two-stage filtering process. First, we discard all candidates that violate basic fabrication constraints (e.g., maximum footprint). Then, based on the predicted metrics, we apply the performance specifications from Table 1, selecting only the designs that satisfy all targets in terms of relative bandwidth, center frequency alignment, and maximum insertion loss. The remaining configurations form the reduced candidate set: $\mathcal{S} = \{(\boldsymbol{x}_k, \hat{y}_k^{(i)})\}_{k=1}^{m}$, with $m \leq M$. Out of the $M = 5000$ virtual candidates, only $m = 32$ satisfied all constraints and were retained in the set $\mathcal{S}$. For the purpose of illustrating the present paper and validating the results, this selected subset was simulated in HFSS. The corresponding responses are shown in Fig. 4 (top). The final stage of the design flow consists in selecting a single configuration, typically the best performing candidate in terms of bandwidth. In practice, this means that only one full-wave simulation is required to validate the final design. In our case, the response of the selected configuration is shown in Fig. 4 (bottom). For this circulator, the surrogate model predicts an effective bandwidth of 0.6267 GHz, while the HFSS simulation yields 0.6270 GHz, confirming the accuracy of the model and the effectiveness of the selection strategy.

V. Conclusions

This work presented a surrogate-assisted approach for circulator design, combining full-wave simulations, kernel-based regression, and constraint filtering. The surrogate model enables fast exploration of the design space, while the proposed screening process selects feasible candidates with minimal recourse to full-wave analysis. The overall strategy offers an efficient and scalable solution for the design of high-frequency devices.

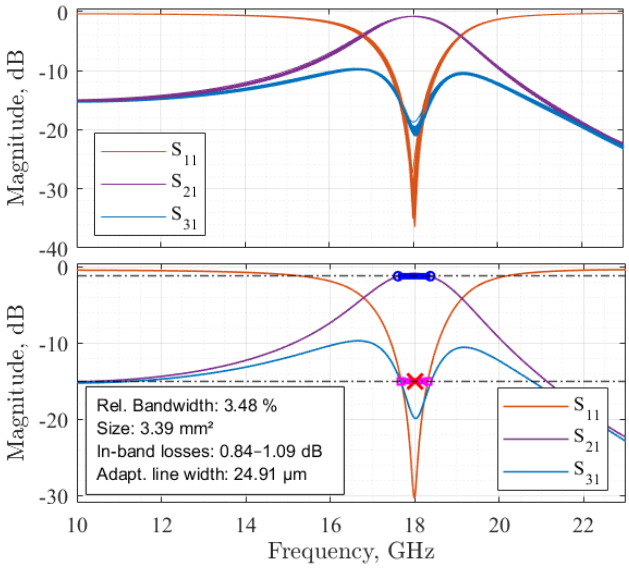

Fig. 4. Top: full-wave responses of the $m = 32$ selected configurations in $\mathcal{S}$ that satisfy all constraints. Bottom: S-parameter curves of the best-performing candidate in terms of relative bandwidth, along with its performance metrics summarized in the inset.

References

[1] N. Parker-Soues, et al., "Compact Self-Biased Q-Band Circulators", *IEEE Trans. Magn.*, vol. 57, no. 4, pp. 1–6, Apr. 2021.

[2] Y. Le Noane, et al., "Uncertainty Quantification for a Microstrip Self-Biased Ku-band Circulator", Asia-Pacific Microwave Conference (APMC2023), Dec. 2023, Taipei, Taiwan.

[3] N. Soleimani, R. Trinchero and F. Canavero, "Bridging the Gap Between Artificial Neural Networks and Kernel Regressions for Vector-Valued Problems in Microwave Applications," IEEE Transactions on Microwave Theory and Techniques, vol. 71, no. 6, pp. 2319-2332, June 2023

[4] A. Rudi, L. Carratino, and L. Rosasco, "Falkon: An optimal large scale kernel method," Advances in neural information processing systems, vol. 30, 2017.

[5] J. Shawe-Taylor and N. Cristianini, *Kernel methods: an overview.* Kernel Methods for Pattern Analysis, Cambridge: Cambridge University Press, 2004, pp. 25–46.