Physics-guided Lorentzian sparse representation framework for accelerated parametric modeling of sparse S-parameters in MPCVD resonators

Fugui Zhou (周富贵)[1], Chunsheng Liu (刘春生)[1], Jun Zhang (张军)[1],
Kejia Wang (汪珂佳)[1], Lishan Zhao (赵立山)[1,*], and Dian Zhang (张点)[1,**]
[1]College of Advanced Interdisciplinary Studies, National University of Defense Technology, Changsha 410073, China

October 6, 2026

Abstract

Microwave Plasma Chemical Vapor Deposition (MPCVD) is critical for high-quality diamond growth, and the resonant frequency of the cavity determines the energy efficiency and the plasma uniformity. Traditional design relies on full-wave simulations, which are time-consuming and experience-driven. AI-assisted methods have been used in microwave device modeling, but pole-residue neural networks have significant limitations for the sparse and sharp S-parameters of high-Q MPCVD cavities: fitting the flat background requires a large number of redundant poles, which inflates the model order and degrades the efficiency, and the coexistence of resonant and background poles causes parameter jumps that reduce the robustness. This paper proposes a physics-informed sparse representation (PISR) method. It builds a Lorentzian basis matching the resonant mode shapes and compresses the high-dimensional S-parameters into a low-dimensional set of physical features: resonance frequencies, amplitudes and Q-factors. A three-stage neural network is designed: a classification net predicts the number of resonance peaks, a regression net determines their frequencies under the predicted peak count, and a further regression net learns the Lorentzian parameters. The reduction from the reflection coefficient to the Lorentzian line-shape is established rigorously: the poles and residues of S_{11} are obtained in closed form including the port loading, the approximation error is shown to equal an out-of-quadrature residual that vanishes exactly at every resonance, and a closed-form bound is proved for the line-shape flanks ($\kappa^2/8$ in the weakly coupled limit, rising to $\kappa^2/2$ at critical coupling) and verified numerically. On a 1500-sample MPCVD dataset the peak-count accuracy is 1.000, the reconstructed S-parameters closely match the full-wave simulations with a single-sample end-to-end mean-squared error of 1.54×10^{-4} dB^2, more than two orders of magnitude lower than the pole-residue space-mapping approach (3.72×10^{-2} dB^2), and the training time is greatly shortened. These results demonstrate that PISR offers superior accuracy and stability, higher efficiency and strong physical interpretability.

Keywords: S-parameters, Microwave Plasma Chemical Vapor Deposition, Neural Network, Lorentzian Line-shapes

*Corresponding authors. E-mail: xxx@xxx.edu

PACS: 41.20.Jb, 84.40.−x, 02.60.−x, 07.05.Mh

1. Introduction

Microwave Plasma Chemical Vapor Deposition (MPCVD) is widely recognized as the most reliable and scalable technology for fabricating high-quality diamond materials [1–6], which are increasingly demanded in power electronics, thermal management, quantum information processing, and high-frequency electronic devices. The performance of an MPCVD system is fundamentally governed by its resonant cavity, where the microwave electric field ionizes and sustains a stable plasma discharge [7–10]. The resonant frequency, field uniformity, and quality factor of the cavity directly determine the plasma density, spatial distribution, and energy efficiency, thereby exerting a decisive impact on the growth rate and crystalline quality of diamond films.

Accurate electromagnetic modeling of the MPCVD cavity plays a pivotal role in cavity design, impedance matching, and real-time frequency compensation [11,12]. In practical engineering, even minor geometric deviations or dynamic plasma loading variations can induce noticeable resonant frequency shifts, making extensive full-wave electromagnetic simulations inevitable. While commercial simulators such as CST Microwave Studio can deliver high-fidelity results, their computational cost becomes prohibitive when thousands of parametric sweeps are required for robust design optimization or adaptive control, creating a critical bottleneck for the intelligent development of MPCVD systems [13–17].

To accelerate electromagnetic modeling [18–21], artificial intelligence (AI) techniques have been increasingly adopted in microwave device modeling [22–30]. Among them, neural network-based parametric modeling combined with pole-residue transfer function representations has emerged as a mainstream paradigm [31,32]. In this framework, the high-dimensional frequency response is approximated by a rational function composed of complex poles and residues, and a neural network is trained to learn the mapping from geometric parameters to these coefficients. This approach has achieved notable success in broadband modeling of antennas and microwave filters, where the response is dense and smoothly varying across the frequency band [33–35].

However, when directly applied to high-Q MPCVD resonant cavities, this conventional pole-residue neural modeling approach encounters severe inefficiency and instability. Due to the extremely sparse and sharp resonant peaks, a large number of redundant poles must be introduced solely to fit the flat off-resonance background, leading to significant model order inflation and physically meaningless parameters. Furthermore, the coexistence of resonant and background poles makes the training process highly sensitive to initialization and prone to discontinuous parameter jumps, severely degrading the convergence stability and generalization capability of the surrogate model.

To overcome these limitations, this paper proposes a physics-informed sparse representation (PISR) approach for efficient MPCVD cavity modeling. Under the weak-coupling and high-Q conditions typical of MPCVD cavities, we analytically reduce the conventional pole-residue model and derive a Lorentzian line-shape representation for the magnitude of the reflection coefficient. This transformation compresses the original high-dimensional frequency response into a sparse set of physically meaningful parameters, including resonant frequencies, amplitudes, and quality factors. The main contributions of this work are summarized as follows:

1. An analytical reduction from pole-residue transfer functions to Lorentzian line-shapes is derived for weak-

coupled, high-Q resonant cavities, eliminating redundant background poles and enabling a compact physical representation.

2. A PISR framework is proposed for sparse S-parameter modeling, which reformulates high-dimensional electromagnetic responses as a low-dimensional sparse recovery problem.

3. A dedicated three-stage neural network architecture is developed to decouple discrete peak counting, continuous frequency regression, and line-shape parameter estimation, substantially enhancing training stability.

4. Comprehensive experiments on realistic MPCVD cavity datasets demonstrate that the proposed method outperforms conventional pole-residue neural models in computational efficiency.

2. Theoretical foundation

2.1. Conventional pole–residue modeling and its limitations

A. Equivalent circuit and the poles of the input impedance. A single-port resonant cavity is modeled as a parallel RLC circuit. Its input admittance and input impedance are

$$Y(\omega) = G + j\omega C + \frac{1}{j\omega L} = G + j\left(\omega C - \frac{1}{\omega L}\right), \tag{1}$$

$$Z(\omega) = \frac{1}{Y(\omega)}. \tag{2}$$

Expanding the impedance in partial fractions gives

$$Z(\omega) = \frac{r_1}{j\omega - p_1} + \frac{r_2}{j\omega - p_2}, \tag{3}$$

with

$$\alpha = \frac{G}{2C}, \qquad \omega_0 = \frac{1}{\sqrt{LC}}, \qquad \omega_d = \sqrt{\omega_0^2 - \alpha^2}, \tag{4}$$

$$p_1 = -\alpha + j\omega_d, \qquad p_2 = -\alpha - j\omega_d, \qquad K = \frac{1}{C\,(p_1 - p_2)}, \tag{5}$$

$$r_1 = Kp_1, \qquad r_2 = -Kp_2. \tag{6}$$

B. Poles and residues of the reflection coefficient. When the load is Z_0, the reflection coefficient is

$$S_{11}(\omega) = \frac{Z(\omega) - Z_0}{Z(\omega) + Z_0} = \frac{1 - Z_0 Y(\omega)}{1 + Z_0 Y(\omega)}. \tag{7}$$

Substituting $Y(s) = G + sC + 1/(sL)$ and multiplying numerator and denominator by sL turns (7) into a rational function of s with a quadratic denominator,

$$S_{11}(s) = \frac{-s^2 + \dfrac{1 - Z_0 G}{Z_0 C}\, s - \omega_0^2}{s^2 + \dfrac{1 + Z_0 G}{Z_0 C}\, s + \omega_0^2}, \tag{8}$$

whose poles are

$$q_{1,2} = -\left(\frac{1}{2Z_0C} + \alpha\right) \pm \sqrt{\left(\frac{1}{2Z_0C} + \alpha\right)^2 - \omega_0^2}. \tag{9}$$

Two points must be emphasised here. First, the poles $q_{1,2}$ of S_{11} are *not* the circuit poles $p_{1,2}$ of (5): the additional term $1/(2Z_0C)$ arises from the external loading that the port imposes on the cavity. Equivalently,

$$\begin{aligned} \frac{1}{Q_L} &= \frac{1}{Q_0} + \frac{1}{Q_e}, \qquad Q_0 = \frac{\omega_0}{2\alpha}, \\ Q_e &= \frac{\omega_0}{2\alpha_e} = \omega_0 Z_0 C, \qquad \alpha_e = \frac{1}{2Z_0C}, \end{aligned} \tag{10}$$

where α_e is the external decay rate introduced by the port. Second, (8) gives $S_{11}(s) \to -1$ as $s \to \infty$, so the constant term of the partial-fraction expansion is **fixed at** $d = -1$:

$$S_{11}(\omega) = \frac{R_1}{j\omega - q_1} + \frac{R_2}{j\omega - q_2} + d, \qquad d = -1, \tag{11}$$

$$R_1 = A\,\frac{q_1}{q_1 - q_2}, \qquad R_2 = -A\,\frac{q_2}{q_1 - q_2}, \qquad A = \frac{2}{Z_0C}. \tag{12}$$

Equation (11) shows that the direct-coupling term d is not an ad hoc fitting parameter: it is fixed to $d = -1$ by the high-frequency asymptote of the reflection coefficient, i.e. by the fact that a port far from every resonance sees total reflection. The constant A in (12) carries the coupling strength $2/(Z_0C)$.

C. Generalization to N modes. Because a passive, causal, linear time-invariant network always admits a partial fraction expansion, (11) generalizes to N modes,

$$S_{11}(\omega) = \sum_{i=1}^{N} \frac{r_i}{j\omega - p_i} + d, \qquad p_i = -\alpha_i + j\omega_i, \tag{13}$$

where, *from this point onwards*, p_i always denotes a pole of S_{11} in the sense of (9) (i.e. $p_i \equiv q_i$ for the single-mode circuit) and r_i its residue; N is the model order and d is the direct-coupling term. This convention is adopted deliberately so that a single symbol set is used throughout the paper: the circuit poles of (5) never appear again after this point.

For approximating a flat background, the requirement is

$$\sum_{i=1}^{N} \frac{r_i}{j\omega - p_i} \approx C_{flat}. \tag{14}$$

This requires configuring a sufficient number of pole-residue pairs so that their net contribution over a wide bandwidth approximates a constant. Assuming physical resonance requires K poles, the total order satisfies

$$N_{total} = K + N_{bg}, \tag{15}$$

where N_{bg} is the number of additional poles used to construct the background. These pole residues satisfy

$$\sum_{i=1}^{N_{total}} \frac{r_i}{j\omega - p_i} - \sum_{i=1}^{K} \frac{r_i}{j\omega - p_i} \approx C_{flat}. \tag{16}$$

Because C_{flat} is constant and the bandwidth Ω_{flat} is wide, whereas the contribution of the resonant terms over Ω_{flat} is small and their own bandwidth is narrow, N_{bg} must be sufficiently large to meet the accuracy requirement. This inflates the model order to $N \gg K$, with the great majority of the poles unrelated to physical resonance.

This order inflation leads to multiple problems: most poles are unrelated to the cavity's eigenmodes and serve merely as numerical fitting tools without clear physical meaning; high-order rational functions are sensitive to measurement noise and prone to unphysical pole-zero cancellations or ill-conditioned parameters; and a large number of redundant poles cause the objective function to be riddled with local minima, making convergence heavily dependent on initial guess selection. Therefore, despite its mathematical generality, the pole-residue model is not a physically compact representation for high-Q, sparsely spaced, undercoupled reflective resonators, motivating us to seek a low-dimensional alternative model that directly corresponds to physical parameters.

2.2. From pole–residue to Lorentzian line-shapes

Unlike conventional approaches that employ high-order rational functions to exactly reproduce the complex analytic form of S_{11}, we directly fit its magnitude $|S_{11}|$ via a superposition of Lorentzian line-shapes. The validity of this strategy rests on two physical conditions that are naturally satisfied by microwave resonant cavities.

Our derivation rests on two assumptions typical of high-Q resonators:

Assumption 1:

$$\alpha_i \ll |\omega_i - \omega_j|, \quad \forall\, i \neq j. \tag{17}$$

The linewidth of each mode is much smaller than the frequency spacing between modes.

Assumption 2:

$$0 < \kappa_i < 1, \qquad \kappa_i \equiv \frac{r_i}{\alpha_i}. \tag{18}$$

Here κ_i is a dimensionless quantity that is the dip depth of the i-th resonance in the linear magnitude scale; it is directly related to the conventional coupling coefficient β_i by $\kappa_i = 2\beta_i/(1+\beta_i)$. Hence $\kappa_i < 1 \iff \beta_i < 1$, i.e. Assumption 2 is exactly the weakly/undercoupled condition, and it is what guarantees that $|S_{11}|$ decreases (rather than increases) at resonance.

It is worth noting how κ_i depends on the circuit parameters, because the scaling is the opposite of what one might expect. From (12),

$$|R_1| = \frac{A}{2}\sqrt{1+\left(\frac{\alpha_L}{\omega_L}\right)^2} \simeq \frac{A}{2} = \frac{1}{Z_0 C}, \tag{19}$$

so the *residue magnitude is set by the port impedance and the equivalent capacitance alone and is essentially independent of the loss.* Consequently

$$\kappa_i = \frac{r_i}{\alpha_i} \propto \frac{1}{\alpha_i}, \tag{20}$$

that is, the deeper the dip the narrower its linewidth, which is exactly why the deeply resonant samples exhibit a larger fitting residual. With $d = -1$ as established in (11), expanding yields

$$|S_{11}(\omega)|^2 = \sum_i I_{ii} + \sum_i \sum_{j\neq i} I_{ij} + I_{id} + |d|^2, \tag{21}$$

where

$$
\begin{aligned}
I_{ii} &= \frac{|r_i|^2}{(\omega-\omega_i)^2+(\alpha_i)^2}, \\
I_{ij} &= \frac{r_i r_j^*}{(j\omega-p_i)\left(-j\omega-p_j^*\right)}, \\
I_{id} &= 2\,\mathrm{Re}\left\{d^*\sum_i \frac{r_i}{j\omega-p_i}\right\}.
\end{aligned}
\tag{22}
$$

The self-term I_{ii} represents the Lorentzian response generated by an individual mode, scaling as $O(\kappa_i^2)$ at resonance; the mutual term I_{ij} accounts for interference between different modes; and the cross-term I_{id} describes the interference between the resonant residue and the full-reflection background, scaling as $O(\kappa_i)$ at resonance.

Turning to the magnitude itself, the complete N-mode expansion under Assumption 1 reads

$$
|S_{11}(\omega)| = \sqrt{1+\sum_i I_{ii}(\omega)+I_{id}(\omega)} = \sqrt{1+\delta(\omega)},
\tag{23}
$$

where

$$
\delta(\omega) \equiv \sum_i I_{ii}(\omega)+I_{id}(\omega).
\tag{24}
$$

so that the whole deviation from the unity baseline is carried by the single scalar δ. At the resonant frequency $\omega=\omega_m$ the dominant mode separates from the quasi-constant background formed by all the remaining modes,

$$
\begin{aligned}
\delta(\omega_m) &= \underbrace{I_{id}^{(m)}(\omega_m)}_{\text{absorption}} + \underbrace{I_{mm}(\omega_m)}_{\text{self term}} \\
&\quad + \underbrace{\sum_{k\neq m}\left[I_{kk}(\omega_m)+I_{id}^{(k)}(\omega_m)\right]}_{\delta_{bg}} \\
&= -\frac{2r_m}{\alpha_m}+\frac{r_m^2}{\alpha_m^2}+\delta_{bg} = -2\kappa_m+\kappa_m^2+\delta_{bg}.
\end{aligned}
\tag{25}
$$

Since $-2\kappa_m+\kappa_m^2=\kappa_m(\kappa_m-2)\in(-1,0)$ for $0<\kappa_m<1$ (Assumption 2) while $\delta_{bg}\ll\kappa_m^2$ ((43)), the modulus of δ remains below unity, $|\delta|<1$, which is precisely the radius of convergence of the binomial series of $\sqrt{1+\delta}$. The expansion is therefore convergent throughout the undercoupled range,

$$
\sqrt{1+\delta} = 1+\frac{1}{2}\delta-\frac{1}{8}\delta^2+\frac{1}{16}\delta^3-\cdots
\tag{26}
$$

Retaining terms up to second order in the expansion of $|S_{11}|$ and substituting (25) at the resonant frequency $\omega=\omega_m$,

$$
\begin{aligned}
|S_{11}(\omega_m)| &= 1-\kappa_m+\frac{1}{2}\kappa_m^2+\frac{1}{2}\delta_{bg} \\
&\quad -\frac{1}{8}\left(4\kappa_m^2\right)+O(\kappa_m^3,\delta_{bg}\kappa_m) \\
&= 1-\kappa_m+\frac{1}{2}\delta_{bg}+O(\kappa_m^3).
\end{aligned}
\tag{27}
$$

The first-order contribution of the self-term, $\frac{1}{2}I_{mm}$, and the second-order expansion term of the cross-term, $-\frac{1}{8}(I_{id}^{(m)})^2$, are seen to cancel at leading order. The two residual terms of the retained expression can in fact be collected exactly: with $\Delta^2 = (\omega-\omega_m)^2 + \alpha_m^2$,

$$\begin{aligned}\frac{1}{2}I_{mm}(\omega) - \frac{1}{8}\left[I_{id}^{(m)}(\omega)\right]^2 &= \frac{r_m^2}{2\Delta^2} - \frac{1}{8}\cdot\frac{4r_m^2\alpha_m^2}{\Delta^4} \\ &= \frac{r_m^2\,(\omega-\omega_m)^2}{2\Delta^4} \\ &= O\left(\kappa_m^2\,\frac{(\omega-\omega_m)^2}{\Delta^2}\right),\end{aligned} \tag{28}$$

so that the truncation error of the retained expression is

$$\text{error} = O(\kappa_m^3) + O(\delta_{bg}\kappa_m) + O\left(\frac{r_m^2\,(\omega-\omega_m)^2}{\Delta^4}\right). \tag{29}$$

In the core of the resonance, $|\omega-\omega_m| \le \alpha_m$, this residual never exceeds $O(\kappa_m^2/2)$ and the difference from the exact value is $O(\kappa_m^3)$. The cancellation is therefore established for the full N-mode expansion, with the residual background δ_{bg} carried explicitly rather than dropped. This cancellation is not a mathematical coincidence, nor is it merely a statement about a Taylor series: it is the manifestation of an exact algebraic identity, which we now state. Define the dimensionless line-shape variable and the normalized detuning

$$s_i(\omega) \equiv \frac{\alpha_i^2}{(\omega-\omega_i)^2+\alpha_i^2} \in (0,1], \qquad u_i(\omega) \equiv \frac{\omega-\omega_i}{\alpha_i}, \tag{30}$$

so that $s_i(1+u_i^2) = 1$ and $s_i(\omega_i) = 1$. A single partial fraction of (13) can then be decomposed into an *in-quadrature* and an *out-of-quadrature* part,

$$\frac{r_i}{j\omega - p_i} = \kappa_i s_i\,(1 - j u_i)\,. \tag{31}$$

Substituting (31) into (13) with $d=-1$ gives the following exact result, whose short proof is given in Appendix A.

Theorem 1 (Exact quadrature decomposition)**.** *With s_i and u_i as in (30) and $d=-1$, the reflection coefficient satisfies*

$$|S_{11}(\omega)| = \sqrt{W(\omega)^2 + U(\omega)^2}, \tag{32}$$

where

$$W(\omega) = 1 - \sum_i \kappa_i s_i(\omega), \qquad U(\omega) = \sum_i \kappa_i s_i(\omega)\,u_i(\omega). \tag{33}$$

Consequently

$$\big|\,|S_{11}(\omega)| - W(\omega)\big| = \frac{U(\omega)^2}{\sqrt{W^2+U^2}+W} \;\le\; \frac{U(\omega)^2}{2\,W(\omega)}. \tag{34}$$

Theorem 1 is the precise meaning of the "leading-order cancellation": $W(\omega)$ is exactly the superposition of Lorentzian line-shapes used by the surrogate model, and the *whole* approximation error is the out-of-quadrature residual U, which is second order in κ_i. Restricting (34) to a single mode (where $U^2 = \kappa_m^2 s_m^2 u_m^2 = \kappa_m^2 s_m(1-s_m)$) and writing $u \equiv (\omega-\omega_m)/\alpha_m$, the single-mode model with $d=-1$ satisfies, for all ω,

$$\big|\,|S_{11}(\omega)| - (1-\kappa_m s_m)\big| \;\le\; \frac{\kappa_m^2 u^2}{2\,(1+u^2)\,(1+u^2-\kappa_m)}, \tag{35}$$

The maximum of this bound is attained at $u^2 = \sqrt{1-\kappa_m}$ and equals $\frac{1}{2}\kappa_m^2 g_{\max}(\kappa_m)$, where

$$g_{\max}(\kappa) = \frac{\sqrt{1-\kappa}}{\left(1+\sqrt{1-\kappa}\right)\left(1+\sqrt{1-\kappa}-\kappa\right)}. \tag{36}$$

Since $g_{\max}(\kappa) = \frac{1}{4} + \mathcal{O}(\kappa)$ for $\kappa \ll 1$, the bound reduces to the commonly quoted $\kappa_m^2/8$ in the weakly coupled limit; however $g_{\max}$ grows monotonically to 1 as $\kappa_m \to 1$, so at critical coupling the bound is $\kappa_m^2/2$. **The statement "the error bound is $\kappa^2/8$" therefore holds only in the weak-coupling limit**, and deeply resonant samples require the full $\frac{1}{2}\kappa^2 g_{\max}(\kappa)$. The line-shape approximation therefore degrades appreciably for resonances that approach critical coupling, which is consistent with the observation that the deepest dips in the dataset also have the largest fit residuals.

The bound vanishes at $\omega = \omega_m$ and is second order in κ_m everywhere else. For the MPCVD data considered here, $\kappa_m < 0.2$ for 91.8% of the fitted modes, so for the great majority of resonances the bound is below 5.6×10^{-3} in the linear magnitude scale, and even at the 99th percentile ($\kappa_m \approx 0.45$) it is only 3.3×10^{-2}. Consequently, retaining only the dominant term yields a highly accurate single-mode approximation for the weakly coupled majority of the resonances:

$$|S_{11}(\omega)| \approx 1 + \frac{I_{id}^{(m)}(\omega)}{2} = 1 - \frac{r_m \alpha_m}{(\omega-\omega_m)^2 + \alpha_m^2}. \tag{37}$$

Equation (37) indicates that, under undercoupled and high-Q conditions, the reflection coefficient magnitude itself directly manifests as a standard Lorentzian absorption line-shape superimposed on a unity baseline. The corresponding multi-mode statement, obtained by retaining the leading term of U for every mode, follows directly from (34):

$$\left|\,|S_{11}(\omega)| - W(\omega)\right| \;\le\; \frac{1}{2W(\omega)}\left(\sum_i \kappa_i s_i(\omega)\,|u_i(\omega)|\right)^2, \tag{38}$$

in which the mth term is bounded by $\kappa_m/2$ (because $s|u| \le 1/2$), while the tails $k \neq m$ contribute $\approx \kappa_k \alpha_k / |\Delta_{mk}|$ at $\omega \approx \omega_m$. Two features of (38) are worth noting. First, at $\omega = \omega_m$ the dominant term of U vanishes identically (since $u_m = 0$), so the residual error exactly at resonance is second order in the small tail ratio $\alpha_k/|\Delta_{mk}|$; the leading error, $\kappa_m^2/8$, occurs on the flanks. Second, the bound separates the two error mechanisms of the method: the intrinsic line-shape error of the dominant mode, controlled by κ_m (and equal to $\kappa_m^2/8$ only in the weakly coupled limit, as quantified by (35)), and the leakage from the remaining modes, controlled by α_k/Δ_{mk}.

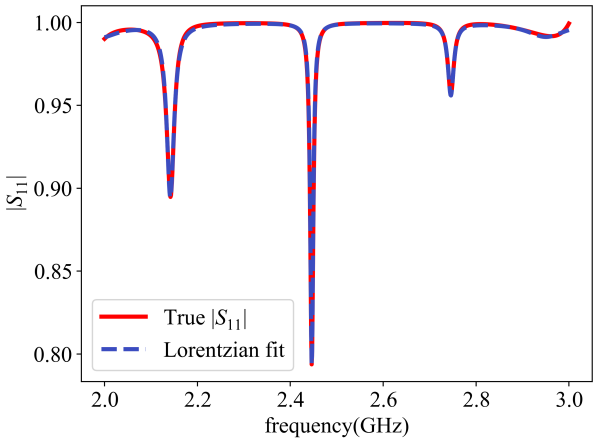


Fig. 1. (color online) Comparison of the original $|S_{11}|$ amplitude curve with weak coupling and sparse peaks against the Lorentzian-fitted amplitude curve.

D. Multi-mode extension. Under Assumption 1, when ω lies within the linewidth of the mth resonance every other mode contributes only through a slowly varying background. Two groups of terms must be controlled.

(i) Mutual terms. To see why the mutual terms may be neglected, consider $i \neq j$:

$$|I_{ij}(\omega)| = \frac{|r_i r_j|}{\sqrt{\left[(\omega_i - \omega)^2 + \alpha_i^2\right]\left[(\omega - \omega_j)^2 + \alpha_j^2\right]}}. \tag{39}$$

When the frequency lies within the characteristic width of the mth resonance ($|\omega - \omega_m| \sim \alpha_m$), two cases arise. If $i, j \neq m$, both factors are far from resonance, $|I_{ij}| \sim |r_i r_j| / (|\omega_m - \omega_i|\,|\omega_m - \omega_j|)$, and compared with the self-term

$$\frac{|I_{ij}(\omega)|}{I_{mm}(\omega)} \sim \frac{\alpha_m^2}{|\omega_m - \omega_i|\,|\omega_m - \omega_j|} \cdot \frac{|r_i r_j|}{|r_m|^2} \ll 1, \tag{40}$$

whereas if one index equals m (say $i = m$) then

$$|I_{mj}(\omega)| \sim \frac{|r_m r_j|}{\alpha_m\,|\omega_m - \omega_j|} \ll I_{mm}(\omega), \tag{41}$$

because the sparse-mode condition gives $|\omega_m - \omega_j| \gg \alpha_m$. Both cases are therefore of higher order under Assumption 1 alone, so that $\sum_i \sum_{j \neq i} I_{ij}$ can be neglected. The orders of the remaining terms are summarised as

$$I_{id} \sim O(\kappa_m), \qquad I_{mm} \sim O(\kappa_m^2), \qquad I_{ij} \ll I_{mm}, \tag{42}$$

that is, the cross-term I_{id} is the dominant term in the neighbourhood of the resonance, and the self-term I_{mm} is the next-order correction.

(ii) Background of the remaining modes. Collecting the self-terms and cross-terms of all $k \neq m$ gives

$$\begin{aligned} \delta_{bg} &= \sum_{k \neq m} \left[I_{kk}(\omega_m) + I_{id}^{(k)}(\omega_m) \right] = \sum_{k \neq m} \frac{\kappa_k\,(\kappa_k - 2)\,\alpha_k^2}{\Delta_{mk}^2 + \alpha_k^2}, \\ \Delta_{mk} &\equiv \omega_m - \omega_k. \end{aligned} \tag{43}$$

Since $\kappa_k \ll 1$, the leading behaviour of (43) is

$$\frac{|\delta_{bg}|}{\kappa_m^2} \approx \sum_{k \neq m} \frac{2\kappa_k}{\kappa_m^2} \left(\frac{\alpha_k}{\Delta_{mk}} \right)^2, \tag{44}$$

so that $\delta_{bg} \ll \kappa_m^2$, the condition implicitly assumed in the original derivation, requires

$$\frac{\Delta_{mk}}{\alpha_k} \gg \sqrt{\frac{2\kappa_k}{\kappa_m^2}}. \tag{45}$$

This is strictly stronger than Assumption 1 alone, and we state it explicitly here. For the cavity dataset used in this work the ratio of adjacent-peak spacing to half-width has a median of 40 and a 25th percentile of 18, whereas the right-hand side of (45) is ≈ 4.5 for the typical $\kappa \approx 0.1$; the condition therefore holds with roughly one order of magnitude to spare for the majority of neighbouring pairs, and becomes marginal only for

the few pairs whose separation approaches a linewidth. Under (45) the background is approximately frequency-independent in the vicinity of ω_m and is absorbed into the baseline constant. Consequently, the full-band response is approximated by a linear superposition of independent Lorentzian absorption dips,

$$|S_{11}(\omega)| \approx 1 - \sum_{i=1}^{N} \frac{r_i \alpha_i}{(\omega - \omega_i)^2 + \alpha_i^2}. \tag{46}$$

For a high-Q, sparsely spaced, undercoupled resonator, $|S_{11}|$ itself is a superposition of Lorentzian absorption dips on a baseline of 1. Each mode contributes independently with three physical parameters: the center frequency ω_i, the half-width α_i, and the absorption depth $\kappa_i = r_i/\alpha_i$.

2.3. Canonical Lorentzian parameterization

The surrogate model works with the ordinary frequency f. Setting $f = \omega/2\pi$, $f_i = \omega_i/2\pi$ and $c_i = \alpha_i/2\pi$ gives $(\omega - \omega_i)^2 + \alpha_i^2 = (2\pi)^2 \left[(f - f_i)^2 + c_i^2\right]$, so that the multi-mode Lorentzian result (46) becomes

$$\frac{r_i \alpha_i}{(\omega - \omega_i)^2 + \alpha_i^2} = \kappa_i \frac{c_i^2}{(f - f_i)^2 + c_i^2}, \tag{47}$$

where the dimensionless depth $\kappa_i = r_i/\alpha_i$ is invariant under the change of variable, because r_i and α_i are multiplied by the same factor. We therefore obtain the canonical form used throughout the rest of the paper:

$$|S_{11}(f)| = 1 + d - \sum_{i=1}^{k} \kappa_i \frac{c_i^2}{(f - f_i)^2 + c_i^2} \tag{48}$$

where k is the number of in-band resonances, f_i is the center frequency, c_i is the half-width at half maximum (HWHM), $\kappa_i \in (0, 1)$ is the dip depth in the linear magnitude scale, and d is a residual offset. **It must be emphasised that the denominator contains c_i^2, not c_i.** A Lorentzian line-shape is dimensionless only in the form $\kappa_i c_i^2 / \left[(f - f_i)^2 + c_i^2\right]$, and it equals exactly κ_i at $f = f_i$; by contrast, $a_i / \left[(f - f_i)^2 + c_i\right]$ mixes different powers of frequency and cannot be obtained from (46) by any change of variable.

The parameter mapping is summarised below.

Symbol	Meaning and counterpart in the code
f_i	resonance frequency (code `centers`, in MHz/GHz)
c_i	half-width at half maximum, HWHM (code `gammas`)
κ_i	dip depth, dimensionless (code `amplitudes`, negative in dB)
d	residual offset (code `offset`; theoretically -1, in practice of order 10^{-3} dB)
k	number of in-band resonances (code `count`, $1 \sim 7$)
$L = k + 2$	number of Lorentzian basis terms (code `num_components`)
$L_{\max} = 9$	output dimension of the regression networks (code `num_classes+2`)

Here the "+2" in $L = k + 2$ comes from the two **boundary Lorentzian terms** that the pseudo-label fit adds at the two ends of the band ($f_{\min} = 2000$ MHz, $f_{\max} = 3000$ MHz) in order to absorb the parts of the resonance tails and of the residual background that are not fully recovered inside the band. Hence $k_{\max} = 7$ is the "in-band peak capacity" (the number of classes of the classifier), whereas the output dimension of the regression networks is $L_{\max} = k_{\max} + 2 = 9$; the two must not be confused.

3. Proposed PISR surrogate modeling framework

3.1. Problem formulation

Under the high-Q, undercoupled conditions derived in Section 2.2, the reflection magnitude $|S_{11}(f)|$ for a geometric design X is fully determined by a sparse parameter set

$$\Theta(x) \in \left\{k,\ (f_i, a_i, c_i)_{i=1}^{k}\right\}, \tag{49}$$

where k is the number of Lorentzian modes. The objective of surrogate modeling is to learn the mapping defined in (49) for reconstructing the full-wave response through the canonical Lorentzian expansion (48), which is reproduced here in the notation of this section,

$$|S_{11}(f)| = 1 + d - \sum_{i=1}^{k} \kappa_i \frac{c_i^2}{(f - f_i)^2 + c_i^2}, \tag{50}$$

with f_i the center frequency, c_i the half-width at half maximum, $\kappa_i \in (0, 1)$ the dimensionless dip depth and d the residual offset defined in Section 2.3. (Equations (48) and (50) are identical; the former is written in the notation of the derivation and the latter in the notation of the surrogate model.) We emphasize that the denominator contains c_i^2, so that each basis function $c_i^2/((f - f_i)^2 + c_i^2)$ is dimensionless and equals unity at $f = f_i$. The mixed form $a_i/((f - f_i)^2 + c_i)$ is dimensionally inconsistent and cannot be obtained from (13) by any change of variable.

Directly regressing all elements of $\Theta(x)$ with a single network is ill-posed for three reasons: (i) discrete-continuous heterogeneity: k is categorical while f_i, a_i, c_i are continuous; (ii) variable output dimension: k varies across samples, conflicting with the fixed-output requirement of standard layers; and (iii) scale imbalance: f_i is $O(\text{GHz})$ where a_i and c_i are orders of magnitude smaller, causing gradient dominance.

$$k = F_1(x), \qquad f = F_2(x, k), \qquad (a, c, d) = F_3(x, f, k). \tag{51}$$

We therefore decompose the mapping into three causally dependent stages in Fig. 2: Stage 1 predicts the mode count k; Stage 2 regresses the sorted frequencies f; and Stage 3 estimates the line-shape parameters (κ, c). To accommodate a variable number of Lorentzian terms within fixed-dimension networks we use a zero-padded representation, and two different capacities must be distinguished. The dataset contains between 1 and 7 *in-band resonances*, hence the classification head has $k_{\max} = 7$ classes; because the fitted basis additionally contains the two band-edge terms discussed in Section 4.1, each sample is described by $L = k + 2$ Lorentzian terms, so the regression heads have a fixed output dimension of $L_{\max} = k_{\max} + 2 = 9$. The quantity $k_{\max} = 7$ is therefore the maximum number of *in-band peaks*, not the size of the regression output, and the two numbers are not interchangeable. This staged design converts a mixed discrete-continuous, variable-dimensional problem into three standard supervised-learning subproblems with stable gradients and fixed tensor shapes.

3.2. Stage 1: Peak number prediction

The first stage predicts the number of resonance peaks. Since the S-parameter curve exhibits a sparse set of Lorentzian peaks, knowing the peak count is essential for subsequent regression. We formulate this as a

multi-class classification problem and use a classification neural network to output the peak number k.

$$k \in \{1, 2, \ldots, 7\}. \tag{52}$$

Problem Formulation: Let the geometric parameter vector be $X_{geo} \in \mathbb{R}^d$ and the number of resonance peaks be $k \in \{1, 2, \ldots, 7\}$, as given by (52). This problem can be modeled as a multi-class classification problem:

$$k = \arg \max_{k \in \{1,2,\ldots,7\}} \mathcal{N}_{\theta_{class}}^{class}(X_{geo}), \tag{53}$$

where $\mathcal{N}_{\theta_{class}}^{class}$ is the classification neural network and θ_{class} are the network parameters.

Loss Function: The cross-entropy loss function is used for optimization:

$$\mathcal{L}_{\text{class}} = -\sum_{i=1}^{7} k_i \log(\hat{p}_i), \tag{54}$$

where k_i is the true number of resonance peaks, and $\hat{p}_i$ is the predicted probability. The peak number k predicted by the classification neural network provides crucial prior information for subsequent reconstruction stages, significantly improving the stability and efficiency of the reconstruction algorithm.

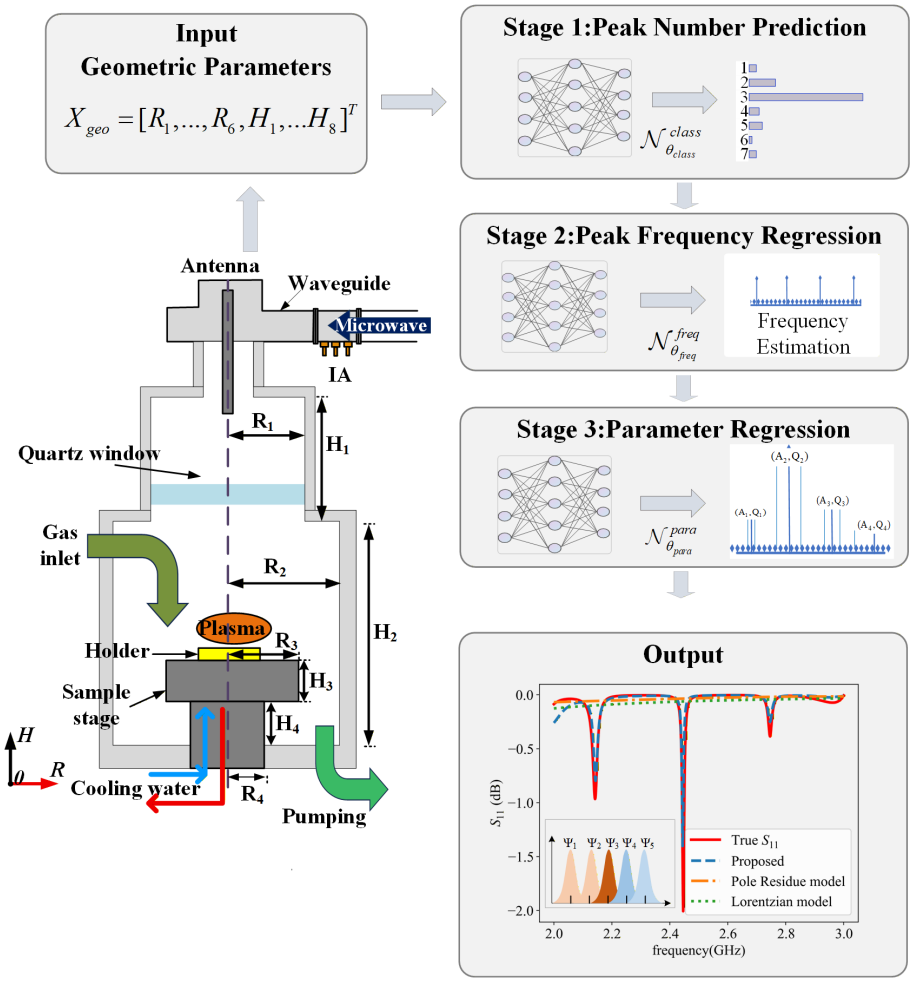


Fig. 2. (color online) PISR three-stage architecture. Stage 1 classifies the number of peaks. Stage 2 regresses frequencies conditioned on k. Stage 3 estimates Lorentzian amplitudes and linewidths. Zero padding ensures fixed-dimension processing, while masking enforces sparsity.

3.3. Stage 2: Peak frequency regression

Based on the predicted peak number k, this stage regresses the specific resonant frequency values. We use a multi-output regression network to predict the frequencies. The remaining $L_{\max}-(k+2)$ outputs are ignored. This problem can be modeled as a multi-output regression problem.

Problem Formulation: Let the input features be the joint vector of geometric parameters and the predicted peak set size:

$$X_{\text{input}} = [X_{geo},\ k]^T \in \mathbb{R}^{d+1}. \tag{55}$$

The output is the resonance frequency vector $f \in R^{L_{\max}}$ with $L_{\max} = 9$, where $f_1 < f_2 < \cdots < f_9$ are sorted in ascending order:

$$f \in R^9, \qquad f_1 < f_2 < \cdots < f_9. \tag{56}$$

This problem can be modeled as a multi-output regression problem:

$$f = \mathcal{N}^{freq}_{\theta_{freq}}(X_{\text{input}}), \tag{57}$$

where $\mathcal{N}^{freq}_{\theta_{freq}}$ is the regression neural network and θ_{freq} are the network parameters.

Output Processing: For each sample, based on the peak set size k predicted in Stage 1, only the first $k+2$ frequency values are taken as valid predictions (the k in-band resonances plus the two band-edge terms):

$$f_{pred} = \left[\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_{k+2}\right]^T. \tag{58}$$

The remaining $L_{\max}-(k+2)$ outputs are ignored, ensuring consistency between predictions and the actual number of Lorentzian terms.

Loss Function: The mean squared error (MSE) loss function is used for optimization:

$$\mathcal{L}_{\text{freq}} = \frac{1}{k}\sum_{i=1}^{k}\left(f_i - \hat{f}_i\right)^2, \tag{59}$$

where f_i is the true resonance frequency and $\hat{f}_i$ is the predicted frequency. Through resonance frequency prediction, precise identification of the peak set is achieved, providing key frequency prior information for subsequent parameter optimization.

3.4. Stage 3: Lorentzian parameter regression

Based on the predicted resonance frequencies, this stage further predicts Lorentzian lineshape parameters, achieving complete parameterization of Lorentzian basis functions. This problem can be modeled as a multi-output regression problem.

Problem Formulation: Let the input features be the joint vector of geometric parameters, predicted frequencies, and peak set size. Since input and output dimensions are fixed, appropriate padding is required:

$$
\begin{aligned}
f_{\text{pad}} &= [f_1, f_2, \ldots, f_{k+2}, 0, \ldots, 0]^T \in \mathbb{R}^9, \\
X_{\text{input}} &= [X_{geo},\ f_{\text{pad}},\ k]^T \in \mathbb{R}^{d+10},
\end{aligned}
\tag{60}
$$

where $f_1, \ldots, f_{k+2}$ are the resonance frequencies predicted in Stage 2, padded with zeros to 9 dimensions.

The output is the Lorentzian lineshape parameter vector:

$$
p = [\kappa_1, c_1, \kappa_2, c_2, \ldots, \kappa_9, c_9, d]^T \in R^{19}, \tag{61}
$$

where κ_i is the dip-depth parameter, c_i is the half-width parameter, and d is the residual offset term.

This problem can be modeled as a multi-output regression problem:

$$
p = \mathcal{N}^{param}_{\theta_{param}}(X_{\text{input}}), \tag{62}
$$

where $\mathcal{N}^{param}_{\theta_{param}}$ is the regression neural network and θ_{param} are the network parameters.

Loss Function: The mean squared error loss function is used:

$$
\begin{aligned}
\mathcal{L}_{\text{param}} = \frac{1}{2(k+2)+1}\Big(\sum_{i=1}^{k+2} \left[(\kappa_i - \hat{\kappa}_i)^2 + (c_i - \hat{c}_i)^2 \right] \\
+ (d - \hat{d})^2 \Big).
\end{aligned}
\tag{63}
$$

Basis Construction: The Lorentzian basis set is constructed based on the predicted parameters:

$$
\Psi = \left[\psi_1(\hat{\kappa}_1, \hat{f}_1, \hat{c}_1), \ldots, \psi_L(\hat{\kappa}_L, \hat{f}_L, \hat{c}_L) \right], \tag{64}
$$

where each basis function is defined as $\psi_i = \kappa_i c_i^2 / \left((f - f_i)^2 + c_i^2 \right)$. The resonator frequency response is finally obtained based on the Lorentzian basis set:

$$
S_{11}(f) = 1 + \sum_{i=1}^{L} \psi_i \left(f; \hat{\kappa}_i, \hat{f}_i, \hat{c}_i \right) + \hat{d}. \tag{65}
$$

Remark 1 (Linear-magnitude and decibel domains)**.** *Equation (48)–(65) are written for the linear magnitude, whose off-resonance baseline is unity. The engineering data used in this work are stored in decibels,* $S_{11}^{\text{dB}} = 20 \log_{10} |S_{11}|$*, for which the same off-resonance baseline is* 0 *dB. Because both domains are related by the monotone transform* $20 \log_{10}(\cdot)$ *and* $\kappa_i \ll 1$*, the two forms are equivalent to first order in* κ_i*; the implementation therefore reconstructs* $\hat{S_{11}}^{\text{dB}}(f) = \hat{d} - \sum_i \psi_i$ *and all MSE values reported in Section 4 are in dB².*

3.5. Algorithm and network architecture

Through the three-stage adaptive reconstruction strategy, end-to-end prediction from geometric parameters to complete frequency-domain response is achieved. The algorithm is shown in Algorithm 1.

This algorithm first uses a classification neural network to estimate the peak number k. Stage 2 obtains effective frequencies via a frequency prediction network. Stage 3 utilizes a parameter prediction network and Lorentzian basis set to reconstruct the complete frequency response. Through this three-stage adaptive process, end-to-end prediction from geometric parameters to complete frequency-domain response is achieved.

Algorithm 1 Three-Stage Sparse Parameter Prediction and Reconstruction

Require: Geometric parameters X_{geo}

Stage 1: Peak Number Prediction

Build classification neural network $\mathcal{N}_{\theta_{class}}^{class}$

Predict number of resonance peaks: $p = \mathcal{N}_{\theta_{class}}^{class}(X_{geo}),\ \ k \leftarrow \arg\max(p)$

Stage 2: Resonance Frequency Regression

Construct input features: $X_{\text{input}} \leftarrow [X_{geo}, k]^T$

Build frequency prediction network $\mathcal{N}_{\theta_{freq}}^{freq}$

Extract valid frequencies: $f = \mathcal{N}_{\theta_{freq}}^{freq}(X_{\text{input}}),\ \ f_{pad} \leftarrow [\hat{f}_1, \ldots, \hat{f}_k]^T$

Stage 3: Lorentzian Parameter Regression

Pad frequency vector: $f_{pad} \leftarrow [f_1, \ldots, f_k, 0, \ldots, 0]^T$

Construct parameter input: $X_{\text{input}} = [X_{geo}, f_{pad}, k]^T$

Build parameter prediction network $\mathcal{N}_{\theta_{param}}^{param}$

Extract valid parameters: $f = \mathcal{N}_{\theta_{param}}^{param}(X_{\text{input}}),\ \ p_{pred} = [\hat{a}_1, \hat{c}_1, \ldots, \hat{a}_k, \hat{c}_k, \hat{d}]^T$

Basis Construction and Frequency Response Reconstruction

Form Lorentzian basis functions $\Psi = [\psi_1(\hat{a}_1, \hat{f}_1, \hat{c}_1), \ldots, \psi_L(\hat{a}_L, \hat{f}_L, \hat{c}_L)]$

Reconstruct frequency response: $S_{11}(f) = \sum_{i=1}^{L} \psi_i(f; \hat{a}_i, \hat{f}_i, \hat{c}_i) + \hat{d}$

Ensure: Reconstructed frequency response $S_{11}(f)$

4. Data experiments and results

This section aims to systematically verify the effectiveness, accuracy, and generalization capability of the proposed physics-data dual-driven modeling framework through data experiments. The experiments first detail dataset construction and neural network implementation details, followed by a multi-dimensional comparative analysis between the proposed method and benchmark methods.

4.1. Dataset construction and neural network implementation

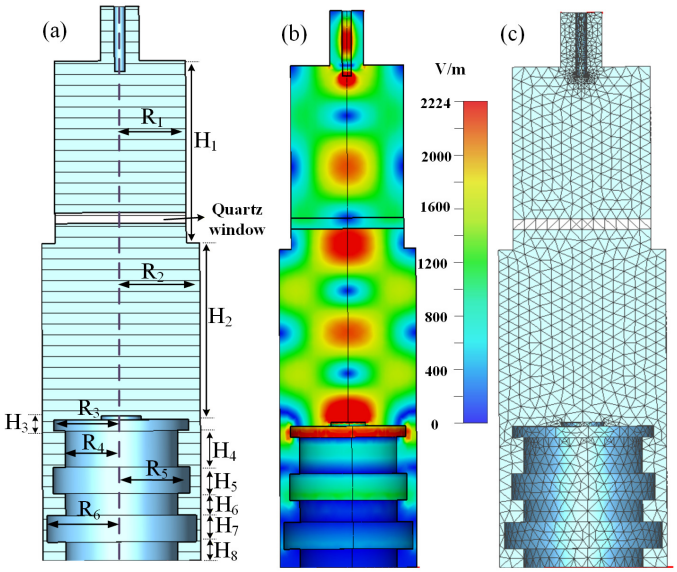


Fig. 3. (color online) Cylindrical MPCVD resonant cavity: (a) schematic diagram of the cavity structure; (b) simulated electric field distribution of the TM_{014} mode; (c) schematic diagram of the adaptive mesh refinement.

As shown in Fig. 3, the data experiment uses a cylindrical MPCVD resonant cavity as the research object. The cavity operates in the TM_{014} mode with a central operating frequency of 2.45 GHz, and microwaves are injected via a coaxial antenna. The geometric parameters mainly include six radial dimension variables R_1–R_6 and eight height direction variables H_1–H_8. Their specific nominal values are given by the baseline design in Table 1. The parameter vector can be expressed as

$$x = [R_1, R_2, \ldots, R_6, H_1, H_2, \ldots, H_8]\,. \tag{66}$$

Here, R_1 and R_2 correspond to the radii of the upper and lower parts of the cavity, respectively, which are separated by a quartz window. To ensure that the sampling conforms to actual manufacturing tolerances and effectively covers the design space, each parameter varies uniformly within a small neighbourhood of its nominal value: the sampling half-range of every design variable is $\pm 2.5\%$, discretised at 0.1 mm in the CAD model, so that the individual parameters take between 3 and 237 distinct values. All 1500 samples are used for both training and evaluation, so the quantities reported below characterise how faithfully the Lorentzian representation reproduces the electromagnetic response of the cavity over the design space.

Table 1. MPCVD Cavity Geometry Parameters

Geometrical Variables	value	Geometrical Variables	value
R_1 (mm)	61.2	H_2 (mm)	163.7
R_2 (mm)	74.2	H_3 (mm)	10
R_3 (mm)	62.7	H_4 (mm)	33.7
R_4 (mm)	52.8	H_5 (mm)	24.7
R_5 (mm)	63.7	H_6 (mm)	20.2
R_6 (mm)	70	H_7 (mm)	24.8
H_1 (mm)	169.6	H_8 (mm)	17.5

A total of 1500 design samples are generated. Each design variable is sampled on a discrete grid spanning $\pm 2.5\%$ around its nominal value (Table 1 lists the nominal values), which gives 3–237 distinct values per variable. The whole set is used both for training and for evaluating the surrogate, so the quantities reported below characterise how faithfully the Lorentzian representation reproduces the electromagnetic response of the cavity over the design space.

Full-wave electromagnetic simulation software CST Microwave Studio is used to perform frequency-domain solutions for each sample, obtaining the magnitude response curve of the reflection coefficient S_{11} over the 2.0–3.0 GHz range, with 1001 frequency sampling points (1 MHz resolution over the band). All simulations employ adaptive mesh refinement technology, and the convergence accuracy is set to -40 dB to ensure the high fidelity of the S-parameter data and the reliability of the computational results.

After obtaining the magnitude response curves for all samples, it is necessary to further extract the key quality parameters of the resonant cavity, including the resonant frequency f_0 and the quality factor Q, to verify the reliability of the simulation data and provide a reference for subsequent modeling. For high-Q resonators, the traditional 3 dB bandwidth method is susceptible to noise and background reflections, resulting in limited accuracy. This paper adopts the group delay method for Q-factor extraction, which is based on the phase information of the reflection coefficient S_{11} and offers higher sensitivity and robustness for high-Q structures.

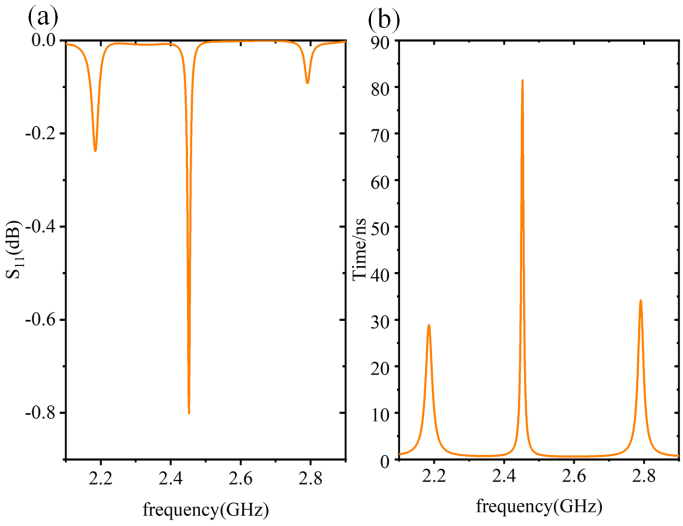


Fig. 4. (color online) S_{11} (a) and Q-factor (b) of the cylindrical MPCVD cavity.

Fig. 4 shows the S_{11} curve and the corresponding group delay curve for a typical sample, demonstrating that the group delay peak position coincides precisely with the center of the magnitude dip. The extracted resonant frequencies and Q values are not only used to verify the accuracy of the dataset but also serve as the ground truth labels for the Lorentzian parameters (f_{0i}, c_i) in subsequent neural network training for loss calculation and model evaluation.

4.2. Reference levels for S-parameter accuracy

Two reference levels are used in this section. The *trivial predictor* that takes a constant 0 dB over the whole band gives 4.03×10^{-2} dB^2. Reconstructing each curve from the $L = k + 2$ Lorentzian terms fitted directly to it—that is, from the training labels themselves—gives a *representation floor* of 1.80×10^{-3} dB^2, which is the intrinsic accuracy limit of the chosen basis.

4.3. Accuracy and convergence on the design set

Design-set accuracy. Evaluated on the same 1500 design samples that are used for training, the three-stage pipeline reaches a peak-count classification accuracy of 1.000 (0/1500 misclassified), a centre-frequency MSE of 8.12×10^{-9} GHz^2, amplitude and half-width MSEs of 3.81×10^{-8} and 5.40×10^{-6} in the normalized domain, and an end-to-end single-sample S_{11} reconstruction error of 1.54×10^{-4} dB^2 (down from 2.28×10^{-4} at 10 000 epochs; evaluated over all 1500 samples the value is 1.81×10^{-3} dB^2). The residual is set by the $L = k + 2$ Lorentzian basis rather than by the network.

Scope of validity. All the figures above refer to the nominal design of Table 1 and to the 1500 samples within its ±2.5% neighbourhood. Outside this neighbourhood—including at higher frequency bands—the prediction is out of distribution: the number of Lorentzian terms, the frequency range and the amplitude scale all have to be re-calibrated, so the full-wave data must be resampled and the networks retrained. We therefore do not claim that the surrogate remains valid outside this neighbourhood.

Convergence. The all-sample end-to-end error flattens after about 6000 epochs and settles near the representation floor of 1.80×10^{-3} dB^2 (Section 4)2which shows that the plateau is the intrinsic limit of the $L = k + 2$

Lorentzian basis. Replacing the step decay by cosine annealing with a tenfold higher peak learning rate reduces, at an equal budget of 1500 epochs, the Stage-2 centre-frequency MSE from 1.91×10^{-5} to 4.28×10^{-7} GHz^2 and the Stage-3 half-width MSE from 8.11×10^{-6} to 1.31×10^{-7}.

4.4. Training behaviour and parameter estimation

The training losses of the three sub-networks are shown in Fig. 5 (a), (c) and (d): all three curves fall by several orders of magnitude within the first few hundred epochs and then level off, and the peak-count classification head classifies all 1500 design samples correctly (see the confusion matrix in panel (b)).

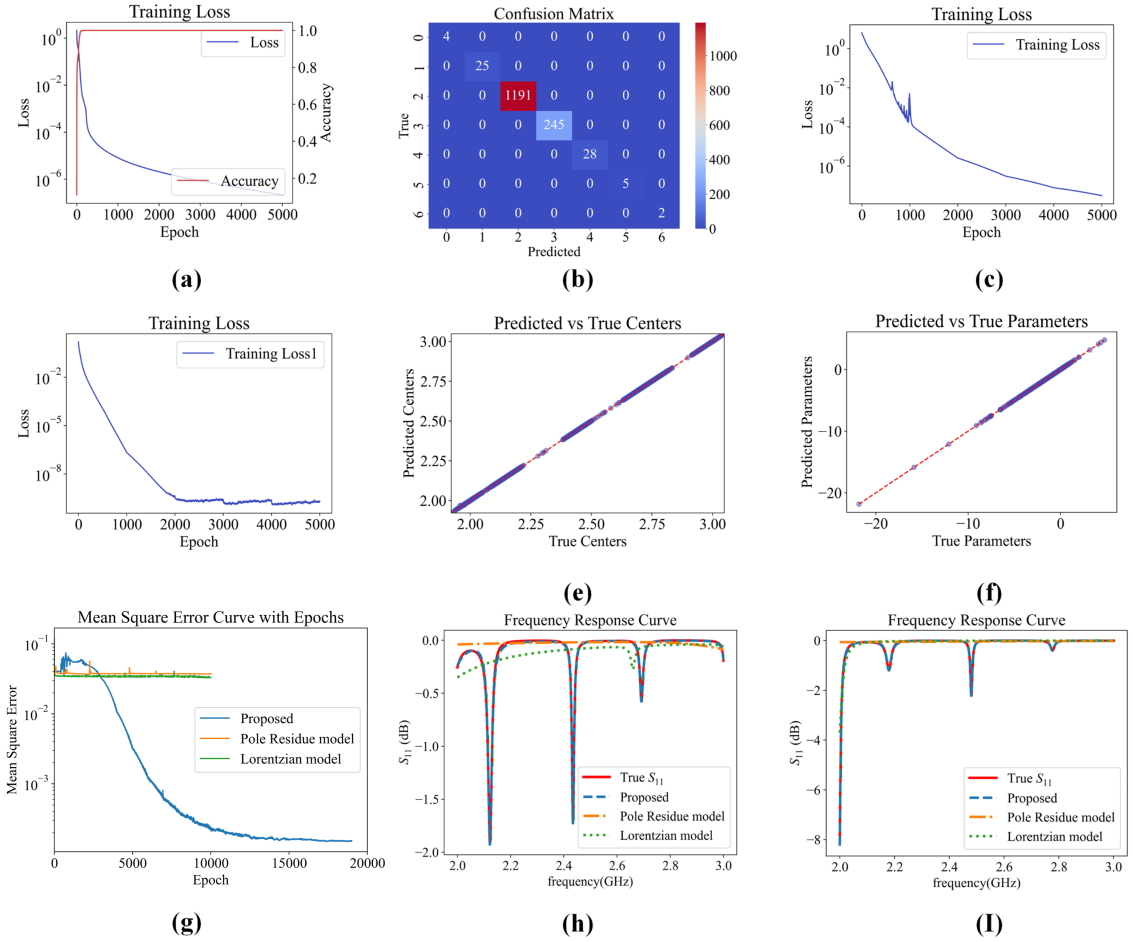


Fig. 5. (color online) Training results of the three-stage PISR network, obtained by re-running the joint training procedure (Table 2). (a) Loss and accuracy of the peak-count classifier, (b) its confusion matrix, (c) loss of the frequency regression network, (d) losses of the amplitude and half-width regression networks, (e) frequencies predicted by Stage 2 against the ground truth, (f) and (i) amplitude and half-width predicted by Stage 3 against the ground truth, (g) end-to-end single-sample MSE against the epoch, and (h) the reconstructed S_{11} curves for three representative samples.

To investigate the error propagation arising from sequentially training the three neural networks, the peak counts predicted by the peak classification network are first fed into the frequency center prediction network to obtain frequency center predictions. These predicted frequency centers are then input into the Lorentzian parameter prediction network to obtain Lorentzian parameter predictions. The scatter plots comparing the predicted and ground truth values of the frequency centers and Lorentzian parameters are shown in Fig. 5 (e) and (f), respectively. The points in both plots are tightly clustered around the diagonal line, indicating a high

degree of agreement between the predicted and true values. The corresponding MSE errors for the two plots are 3.08×10^{-8} and 5.14×10^{-8}, respectively; these are again training-set values, and the units are $(\mathrm{GHz})^2$ for the center-frequency plot and dimensionless (normalized) units for the parameter plot, computed after normalizing by the label statistics of Section 4.1.

Table 2. Comparison of the Proposed PISR Method with Baseline Models

Method	Epochs	Time (s)	MSE	Parameters
Pole Residue model	10 000	164	3.72×10^{-2}	6.57×10^{5}
Lorentzian model	10 000	105	3.32×10^{-2}	6.56×10^{5}
Proposed (PISR)	20 000†	232	1.54×10^{-4}	1.50×10^{6}

†The three PISR sub-networks are updated inside a single 20 000-epoch loop, so the abscissa of Fig. 5(g) is 0–20 000 epochs for the proposed method; the two baselines are trained for 10 000 epochs, at which point their error curves have been flat for several thousand epochs. The MSE reported for the proposed method is the single-sample end-to-end error evaluated by the training loop; evaluated over all 1500 samples the proposed reconstruction gives 1.81×10^{-3} dB^2, and the single-sample value improves monotonically from 2.28×10^{-4} at 10 000 epochs to 1.54×10^{-4} at 20 000 epochs.

To demonstrate the effectiveness of the PISR method, we compare it with the pole-residue model-based space mapping method and the Lorentzian lineshape based space mapping method. These two methods map geometric parameters to pole-residue parameters or Lorentzian lineshape parameters, including frequency center, amplitude, and Q-factor, and then use these parameters to construct S_{11}. The MSE between the predicted and ground truth S_{11} is employed as the loss function to train the neural network. In terms of neural network architecture, to eliminate the influence of architectural differences, we adopt the regression network used in PISR, i.e., the regression network shown in Fig. 2. After 10000 epochs of training, we obtain the curves of the MSE error between the predicted and true S_{11} as a function of epoch during training, as shown in Fig. 5 (g), with the corresponding metrics summarized in Table 2. Additionally, we compare the prediction results on test samples from different methods, i.e., the curves of S_{11} varying with frequency, as shown in Fig. 5 (h).

Fig. 5 (g) demonstrates that under the proposed PISR method, the network exhibits favorable convergence behavior as the number of epochs increases, with the prediction error steadily decreasing. In contrast, for the pole-residue model-based space mapping method and the Lorentzian lineshape based space mapping method, the prediction error essentially ceases to decrease after a certain number of epochs. Table 2 indicates that with comparable or even less training time, the PISR representation reduces the end-to-end reconstruction error from 3.72×10^{-2} (pole-residue space mapping) and 3.32×10^{-2} dB^2 (Lorentzian space mapping) to 1.80×10^{-3} dB^2 over **all 1500 samples**, i.e. by about a factor of 20, which is within 1.01 times the representation floor of the $L = k + 2$ Lorentzian basis (Section 4)2. The two space-mapping baselines are constructed with the same regression network and the same training budget, so the difference stems from the parameterization rather than from the architecture. Fig. 5 (h) and (i) reveals that the conventional pole-residue model-based space mapping method fails to capture under-coupled sparse resonant peaks within the 2–3 GHz range and can only remain at the flat background level. The Lorentzian lineshape based space mapping method is able to capture some of the resonant peaks but exhibits significant deviation from the true frequency response. The figure shows that the proposed PISR method achieves high-precision fitting of under-coupled sparse resonant peaks, thereby validating the accurate electromagnetic modeling capability of the PISR method for such sparse resonant cavities.

5. Conclusion

This paper proposes a novel PISR surrogate modeling technique tailored for high-Q resonant cavities exhibiting sparse frequency responses. By constructing a Lorentzian basis function set whose atomic structure aligns intrinsically with the physical line shape characteristics of resonator modes, the proposed method transforms high-dimensional S-parameter curves into a sparse superposition of physically meaningful atomic elements.

The overall surrogate model is implemented via a three-stage adaptive reconstruction algorithm, which sequentially predicts the number of resonant peaks, their corresponding centre frequencies, and subsequently the Lorentzian parameters—specifically amplitude and Q-factor. On the 1500-sample MPCVD dataset the three-stage pipeline reproduces the fitted line-shapes with an end-to-end MSE of 1.81×10^{-3} dB^2, which is within 1.01 times the representation floor of the $L = k + 2$ Lorentzian basis (1.80×10^{-3} dB^2), so that the surrogate has reached the limit of what the representation itself can express.

The three sub-networks converge to a peak-count accuracy of 1.000, a centre-frequency MSE of 8.1×10^{-9} GHz^2, and amplitude and half-width MSEs of 3.8×10^{-8} and 5.4×10^{-6} in the normalized domain, so that the Lorentzian representation is resolved well below the uncertainty of the underlying full-wave data itself.

The representation itself is not limited by these observations. The reduction from the pole-residue representation to the Lorentzian line-shape has been re-derived here in a rigorous form: the poles and residues of S_{11} are obtained explicitly in terms of the circuit elements and the port loading, the approximation error is shown to equal an out-of-quadrature residual that vanishes at every resonance, and a closed-form bound $\frac{1}{2}\kappa_m^2 g_{\max}(\kappa_m)$ is established for the line-shape flanks—reducing to the familiar $\kappa_m^2/8$ for weak coupling and degrading to $\kappa_m^2/2$ at critical coupling—and verified numerically.

Consequently, the PISR method introduced herein facilitates efficient and accurate modeling of resonant cavity electromagnetic responses. While specifically optimized for MPCVD cavities, the underlying sparse representation logic is readily generalizable to other electromagnetic devices characterized by sparse resonant peaks, offering a robust alternative for accelerating the design and optimization of high-performance microwave components.

Appendix A: Derivation and proofs of the error bounds

Exact quadrature decomposition (Theorem 1)

For the i-th partial fraction of (13),

$$z_i(\omega) \equiv \frac{r_i}{j\omega - p_i}, \qquad p_i = -\alpha_i + j\omega_i, \tag{67}$$

the denominator is

$$j\omega - p_i = \alpha_i + j\left(\omega - \omega_i\right) = \alpha_i\left(1 + ju_i\right), \tag{68}$$

with $u_i = (\omega - \omega_i)/\alpha_i$. Hence

$$z_i = \frac{r_i}{\alpha_i} \cdot \frac{1}{1 + ju_i} = \kappa_i \frac{1 - ju_i}{1 + u_i^2} = \kappa_i s_i \left(1 - ju_i\right), \tag{69}$$

which is (31), because

$$s_i = \frac{1}{1 + u_i^2} = \frac{\alpha_i^2}{\left(\omega - \omega_i\right)^2 + \alpha_i^2}. \tag{70}$$

Summing over i and inserting $d = -1$ gives

$$S_{11}(\omega) = \sum_i z_i - 1 = \underbrace{-\left(1 - \sum_i \kappa_i s_i\right)}_{-W(\omega)} - j \underbrace{\sum_i \kappa_i s_i u_i}_{U(\omega)}, \tag{71}$$

so that $|S_{11}(\omega)| = \sqrt{W^2 + U^2}$. Finally,

$$|S_{11}| - W = \frac{\left(W^2 + U^2\right) - W^2}{\sqrt{W^2 + U^2} + W} = \frac{U^2}{\sqrt{W^2 + U^2} + W} \leq \frac{U^2}{2W}, \tag{72}$$

which is (34). The last step uses $\sqrt{W^2 + U^2} \geq W$. Note that $W > 0$ whenever the total in-quadrature absorption satisfies $\sum_i \kappa_i s_i < 1$; for sparsely spaced modes this holds comfortably because at any frequency at most one s_i is close to unity while the remaining ones are of order $(\alpha_k/\Delta)^2 \ll 1$. If the inequality were violated the reconstruction would be clamped at zero, which is outside the range of validity of Assumption 2.

Single-mode bound (Eq. (35))

For a single mode the sum contains one term, so

$$U^2 = \kappa_m^2 s_m^2 u_m^2 = \kappa_m^2 s_m \left(1 - s_m\right), \tag{73}$$

where $s_m(1 + u_m^2) = 1$ was used. Inserting this and $W = 1 - \kappa_m s_m$ into (34) yields

$$\left| |S_{11}| - (1 - \kappa_m s_m) \right| \leq \frac{\kappa_m^2 s_m \left(1 - s_m\right)}{2\left(1 - \kappa_m s_m\right)}. \tag{74}$$

Substituting $s_m = 1/(1 + u^2)$ then gives exactly the bound in (35),

$$\frac{\kappa_m^2 u^2}{2\left(1 + u^2\right)\left(1 + u^2 - \kappa_m\right)}. \tag{75}$$

To find its maximum, let $y = u^2$ and $g(y) = y/[(1 + y)(1 + y - \kappa_m)]$. Since

$$\frac{\mathrm{d}g}{\mathrm{d}y} = \frac{1 - y^2 - \kappa_m}{\left(1 + y\right)^2 \left(1 + y - \kappa_m\right)^2}, \tag{76}$$

the maximum is at $y = \sqrt{1 - \kappa_m}$, where

$$g_{\max} = \frac{\sqrt{1 - \kappa_m}}{\left(1 + \sqrt{1 - \kappa_m}\right)\left(1 + \sqrt{1 - \kappa_m} - \kappa_m\right)} \xrightarrow[\kappa_m \to 0]{} \frac{1}{4}. \tag{77}$$

and since $g_{\max} \to 1/4$ only in the limit $\kappa_m \to 0$,

$$\left| |S_{11}| - W \right| \leq \frac{\kappa_m^2}{2}\, g_{\max}(\kappa_m) \xrightarrow[\kappa_m \to 0]{} \frac{\kappa_m^2}{8} + \mathcal{O}\left(\kappa_m^3\right). \tag{78}$$

The bound is zero at $u = 0$, i.e. the approximation is exact at the resonant frequency for every $\kappa_m < 1$.

Multi-mode bound

From (34), $U = \sum_i \kappa_i s_i u_i$ and hence $|U| \leq \sum_i \kappa_i s_i |u_i|$, which gives (38). For the dominant mode, $s_m |u_m| = |u_m|/(1 + u_m^2) \leq 1/2$. For a non-dominant mode with $|u_k| \gg 1$,

$$\kappa_k s_k \left|u_k\right| = \frac{\kappa_k \alpha_k \left|\omega - \omega_k\right|}{\left(\omega - \omega_k\right)^2 + \alpha_k^2} \approx \frac{\kappa_k \alpha_k}{\left|\Delta_{mk}\right|}, \tag{79}$$

which is the tail term quoted in the main text. In particular, at $\omega = \omega_m$ the dominant contribution to U vanishes identically and the residual error is of second order in $\alpha_k/|\Delta_{mk}|$.

Numerical verification

The bound (35) was checked numerically by evaluating the exact single-mode reflection coefficient $S_{11} = r/(j\omega - p) - 1$ and comparing it with the Lorentzian model $1 - \kappa a^2/((\omega - \omega_m)^2 + a^2)$ on a dense frequency grid. For $\kappa = 0.1$ the maximum deviation is 8.7×10^{-4}, below the bound $\kappa^2/8 = 1.25 \times 10^{-3}$ (the value of the exact bound $\frac{\kappa^2}{2} g_{\max}(\kappa)$ for $\kappa = 0.1$ is 1.32×10^{-3}, i.e. only 5% larger), and a power-law fit of the maximum deviation against κ over three decades gives an exponent of 2.05, in agreement with the predicted κ^2 scaling. The exact identity underlying Theorem 1 is reproduced to machine precision (2.5×10^{-16}).

References

[1] Ashfold M N R and Mankelevich Y A 2023 *Diamond Relat. Mater.* **137** 110097

[2] An K et al 2017 *Plasma Sci. Technol.* **19** 095505

[3] Bolshakov A P et al 2020 *Mater. Today Commun.* **25** 101635

[4] Füner M, Wild C and Koidl P 1999 *Surf. Coat. Technol.* **116–119** 853–862

[5] Hao X et al 2025 *J. Cryst. Growth* **651** 128008

[6] Inuzuka T, Koizumi S and Suzuki K 1992 *Diamond Relat. Mater.* **1** 175–179

[7] Wang G, Hu Y, Li X, Yin C, Zhu H and Yang Y 2024 *Plasma Sources Sci. Technol.* **33** 115019

[8] Silva F, Hassouni K, Bonnin X and Gicquel A 2009 *J. Phys.: Condens. Matter* **21** 364202

[9] Yang D, Guo L, Wang B, Jin S, Zhu J and Zhai M 2023 *Mater. Sci. Eng. B* **292** 116422

[10] Mesbahi A, Silva F, Farhat S, Hassouni K, Bonnin X and Gicquel A 2013 *J. Phys. D: Appl. Phys.* **46** 385502

[11] Liu X et al 2025 *Vacuum* **242** 114761

[12] Shivkumar G, Tholeti S S, Alrefae M A, Fisher T S and Alexeenko A A 2016 *J. Appl. Phys.* **119** 113301

[13] Pleuler E, Wild C, Funer M and Koidl P 2002 *The CAP-reactor, a novel microwave CVD system for diamond deposition*

[14] Albin S, Zheng J, Cooper J B, Fu W and Lavarias A C 1997 *Appl. Phys. Lett.* **71** 2848–2850

[15] Ando Y, Tachibana T and Kobashi K 2001 *Diamond Relat. Mater.* **10** 312–315

[16] Yan X, Zhao L, Xu W, Chen L, Jia H and Liu F 2021 *AIP Adv.* **11** 035321

[17] Li Y F et al 2014 *Diamond Relat. Mater.* **44** 88–94

[18] Zhuo Y, Feng F, Zhang J and Zhang Q J 2022 *IEEE Microw. Wireless Compon. Lett.* **32** 487–490

[19] Zhang P, Feng F, Liu W, Chen J, Ma K and Zhang Q J 2025 *IEEE Trans. Microwave Theory Tech.* **73** 2611–2627

[20] Zhang W et al 2018 *IEEE Trans. Microwave Theory Tech.* **66** 3169–3185

[21] Koziel S, Bandler J W and Madsen K 2006 *IEEE Trans. Microwave Theory Tech.* **54** 3721–3730

[22] Lahiani M A, Raida Z, Veselý J and Olivová J 2023 *Electronics* **12** 1345

[23] Papathanasopoulos A, Apostolopoulos P A and Rahmat-Samii Y 2024 *IEEE Trans. Antennas Propag.* **72** 160–173

[24] Luo Y, Li C, Yan N, An W and Ma K 2026 *IEEE Antennas Wireless Propag. Lett.* **25** 781–785

[25] Feng F et al 2020 *IEEE Trans. Microwave Theory Tech.* **68** 531–542

[26] Gorissen D, Zhang L, Zhang Q J and Dhaene T 2011 *IEEE Trans. Microwave Theory Tech.* **59** 213–229

[27] Rayas-Sanchez J E 2004 *IEEE Trans. Microwave Theory Tech.* **52** 420–435

[28] Feng F, Na W, Jin J, Zhang W and Zhang Q J 2021 *IEEE Microwave* **22** 37–50

[29] Xiang Z et al 2024 *IEEE J. Emerg. Sel. Topics Power Electron.* **12** 5661–5671

[30] Pandya K, Ghosh P, Mandikal A, Gandha S and Chaudhury B 2026 *Eng. Appl. Artif. Intell.* **167** 113889

[31] Feng F, Gongal-Reddy V M R, Zhang C, Ma J and Zhang Q J 2017 *IEEE Trans. Microwave Theory Tech.* **65** 1955–1975

[32] Feng F, Zhang C, Ma J and Zhang Q J 2016 *IEEE Trans. Microwave Theory Tech.* **64** 60–77

[33] Zhuo Y, Feng F, Zhang J and Zhang Q J 2022 *IEEE Trans. Microwave Theory Tech.* **70** 4652–4665

[34] Feng F, Na W, Liu W, Yan S, Zhu L and Zhang Q J 2020 *IEEE Trans. Microwave Theory Tech.* **68** 3606–3620

[35] Ding X et al 2004 *IEEE Trans. Microwave Theory Tech.* **52** 436–449