

Optimization of Bandpass Filters for Modern Communication Systems: A Hybrid Genetic Algorithm-Machine Learning Framework with Fractal Defected Ground Structures

Ashikur Mollik*, Md Nadim Hosain**, Md Imran Hossain***

*(School of Electronics and Information Engineering, Nanjing University of Information Science & Technology, Nanjing, China

Email: ashikurmollik@nuist.edu.cn

** (School of Electronics and Information Engineering, Nanjing University of Information Science & Technology, Nanjing, China

Email: mdnadimhosain@nuist.edu.cn

*** (School of Electronics and Information Engineering, Nanjing University of Information Science & Technology, Nanjing, China

Email: hossainmdimran@nuist.edu.cn

Abstract:

Bandpass filters set the achievable selectivity and the noise floor of practically every radio front end, yet the design cycle for millimetre-wave and multi-band devices is still dominated by repeated full-wave simulation. This paper reports a synthesis and optimization framework that couples a genetic algorithm to a convolutional surrogate model, and applies it to three realizations: a 28 GHz substrate-integrated waveguide filter for 5G base stations, a 2.4 GHz miniaturized microstrip filter for battery-powered IoT nodes, and a Ka-band cavity filter for satellite payloads. The surrogate is trained on 5000 previously simulated geometries spanning 1-100 GHz and screens an entire candidate population in under one hour, roughly twelve times faster than direct electromagnetic evaluation; only the top decile is passed to full-wave verification, and the surrogate is retrained on the augmented archive at every cycle. Convergence is reached in five cycles instead of fifteen, which cuts wall-clock optimization time from 72 h to 12 h and lowers the average passband insertion loss of the converged population from 2.1 dB to 1.8 dB. A dual-band variant built on Hilbert-curve defected ground structures occupies 15 mm × 10 mm, about 70 % smaller than comparable published designs, and suppresses the third harmonic at 4.8 GHz by 12 dB relative to hairpin and interdigital references. Measured insertion loss for the 28 GHz prototype is 2.1 dB against 1.8 dB simulated, a 0.3 dB gap consistent with the surface roughness of the fabricated copper. The obstacles that remain are the cost of low-loss laminate in volume IoT production and the scarcity of training data above 100 GHz, both of which limit how far the method extends towards terahertz designs.

Keywords — Bandpass filter, CNN, genetic algorithm, millimetre wave, surrogate modelling.

I. INTRODUCTION

A. Context

Every receiver chain places a bandpass filter (BPF) between the antenna and the low-noise amplifier, and whatever that filter dissipates is added almost directly to the system noise figure [1]. A tenth of a decibel is not a rounding error. In a base station serving the n257 band the filter sits

ahead of dozens of parallel receive paths, so insertion loss propagates into transmit power and into the thermal budget of the whole array. In a battery-powered sensor the same tenth of a decibel is a measurable fraction of the link margin that decides how many months the node survives on one cell.

The bands in play have moved. Sub-6 GHz allocations are congested, and the spectrum

between 24 and 40 GHz now carries a meaningful share of dense-urban capacity. Recommendation ITU-R M.2160 fixes the framework for IMT-2030 and anticipates operation from below 1 GHz to above 100 GHz, which means that a filter designer is now expected to work across three regimes that behave very differently [2]. Below 6 GHz the dominant constraint is area, because the resonator is a sizeable fraction of the module footprint. Between 24 and 40 GHz the dominant constraint is loss, because conductor roughness and dielectric absorption both scales unfavourably with frequency. Above 100 GHz neither the materials nor the tolerances of ordinary printed processes are adequate, and the design problem becomes partly a manufacturing problem [3].

That movement now has a clear direction, and it is worth being precise about it because it changes what a filter for the next decade has to do. When the original version of this study was carried out the design frontier was the 24-40 GHz millimetre-wave bands opened for 5G. By 2026 the centre of gravity has shifted downward. Millimetre-wave cellular has not delivered the coverage its proponents hoped for, and the industry consensus that has formed around the ITU-R IMT-2030 framework places the first 6G deployments in the upper mid-band [2], the 7.125-24.25 GHz range that the 3GPP and the wider community now call Frequency Range 3, or FR3 [33], [34]. It has acquired the informal name of the golden band, because it offers a compromise the two established regimes cannot: several times the bandwidth available at 3.5 GHz, together with propagation and coverage far closer to sub-6 GHz than to millimetre wave [35].

For the filter designer this is not a cosmetic relabelling. FR3 sits exactly in the transition region between the two loss regimes described above. At the lower end, near 7 GHz, area is still the binding constraint and dielectric loss is manageable on ordinary laminate; at the upper end, near 24 GHz, conductor roughness and dielectric absorption already dominate the budget, as they do in the millimetre-wave bands. A filter family that is to serve FR3 must therefore span both regimes with a single methodology, which is precisely the span the three realisations of this paper were chosen to

cover. The 2.4 GHz and 28 GHz devices bracket FR3 from below and above, and the design procedure that connects them is the contribution that transfers directly to the golden band.

Acoustic technologies dominate the low end of this range. Surface acoustic wave and bulk acoustic wave devices offer selectivity that no planar structure can match at a comparable size, but the acoustic velocity of the piezoelectric stack places a practical ceiling near 6 GHz [4]. Above that ceiling the field falls back on distributed structures: coupled microstrip resonators, substrate-integrated waveguide (SIW) cavities, and machined or printed metal cavities. Each of these carries a different compromise between unloaded quality factor, area and cost, and none of them is uniformly preferable.

FR3 has sharpened this picture rather than softened it. The channel bandwidths envisaged for the band run from 100 MHz to well beyond 1 GHz to accommodate the coexistence of several services in one allocation [33], and bandwidths of that order are difficult for acoustic resonators to reach at all. Recent work on thin-film lithium niobate has pushed acoustic filtering toward the lower FR3 channels [41], but the fundamental limit set by the acoustic velocity remains, and distributed structures the microstrip, SIW and cavity families treated here are the realistic candidates across most of the band. That is the technical reason the golden band is, for the moment, a distributed-filter problem, and it is why a synthesis and optimisation method spanning those families is timely.

B. Where Current Practice Falls Short

Classical filter synthesis is a solved problem in the circuit domain. Given a specification one can compute a Chebyshev prototype, map it to a coupling matrix, and read off the required coupling coefficients and external quality factors [5], [6]. The difficulty is the step that follows: translating those circuit quantities into physical dimensions. That mapping is neither analytic nor monotonic in general, and at millimetre-wave frequencies it is sensitive to features via barrel geometry, etch undercut, plating roughness that the circuit model does not represent at all.

The standard remedy is to close the loop numerically. A parametric sweep or an

evolutionary search is wrapped around a full-wave solver, and the solver is queried until the response is acceptable. This works, and it is what most practitioners do, but the cost is severe. A single adaptive-mesh solution of a fourth-order SIW cavity at 28 GHz occupies tens of minutes on a well-provisioned workstation; a genetic algorithm with a population of a hundred needs that solution a hundred times per generation. Reported optimization campaigns of this kind routinely consume days of wall-clock time [7].

Surrogate modelling is the natural response, and the microwave community has converged on it over the past decade. Space mapping established the principle that a cheap coarse model can be corrected towards an expensive fine model [8]. Neural surrogates extended it, first as parametric models of individual responses and later as general predictors trained across a design family [7], [9], [10]. What has been less thoroughly examined is how to schedule the interaction between the surrogate and the evolutionary search so that the surrogate stays accurate in the region the search actually visits, rather than merely accurate on average over its original training distribution.

C. Contributions

This work addresses that scheduling question and the physical design problem together. The specific contributions are as follows.

- A hybrid GA-ML optimisation kernel in which a convolutional surrogate screens the entire population, a fixed decile is escalated to full-wave verification, and the surrogate is retrained on the augmented archive after every cycle. Section IV-D gives the retraining schedule and the termination test.
- A closed-form model for Hilbert-curve defected ground structures that relates the fractal iteration order to the slot electrical length and hence to the position of the harmonic attenuation pole, derived in Section III-E.
- A loss budget that separates conductor, dielectric and radiation contributions for each of the three substrates considered, including a roughness correction, so that the measured-simulated discrepancy can be attributed rather

than merely reported (Section III-F and Section VI-E).

- Three fabricated realisations spanning 2.4 GHz to 30 GHz, plus a dual-band device, benchmarked against published designs in Section VI-F.

A negative result is reported alongside these. The IoT filter did not meet its own insertion-loss target, and Section VI-J explains why the target was unrealistic given the substrate that volume production requires.

D. Relevance to 6G and the Upper Mid-Band

Because the framing above runs through the whole paper, it is worth stating the 6G connection once, explicitly, so that the later technical sections can refer back to it. Three properties of the FR3 golden band shape the requirements placed on its filters, and each map onto a specific element of the method developed here.

First, FR3 spans a frequency decade, from roughly 7 to 24 GHz, across which the dominant loss mechanism changes from dielectric to conductor. A design method that optimizes against only one of these will fail at the other end of the band, which is the argument for the explicit, separable loss budget of Section III-F: it lets the same objective function weight conductor and dielectric loss correctly at any point in the band. Second, the wide and variable channel bandwidths of FR3 demand filter whose fractional bandwidth is a controlled design output rather than an accident of the topology, which is why bandwidth error appears as an explicit penalty ϕ_3 in the objective of (28). Third, the coexistence of cellular, satellite and radio astronomy services within FR3 places a premium on out-of-band rejection and on harmonic suppression, which is the role of the cross-coupling of Section III-G and the fractal DGS of Section III-E.

The two millimetre-wave and sub-6 GHz realizations reported here therefore function as the boundary cases of an FR3 design space rather than as isolated demonstrations. A filter centred at 15 GHz, near the middle of the golden band, would inherit the SIW topology of the 28 GHz device with the substrate economics of the 2.4 GHz device, and the loss budget of Section III-F predicts its performance without a separate campaign. Fig. 2

marks two representative FR3 points to make this interpolation concrete.

E. Organisation

Section II reviews the relevant literature and identifies the gaps this work addresses. Section III develops the synthesis theory, the SIW and DGS models, and the loss budget, and states the optimization problem formally. Section IV describes the workflow, the optimizer, the materials and the measurement set-up. Section V presents the four filter realizations. Section VI reports and discusses the results, including limitations. Section VII concludes.

II. RELATED WORKS

The literature relevant to this paper divides into four strands: the displacement of deterministic tuning by surrogate-assisted optimization, the comparative behavior of filter topologies, the substrate technologies that become viable at millimetre-wave frequencies, and the performance envelopes that published designs actually achieve.

A. From Deterministic Tuning to Surrogate-Assisted Optimisation

Filter optimization in the 2010s was largely evolutionary. Genetic algorithms and particle swarm optimization were attractive because they require no gradient and tolerate the discontinuous, multi-modal cost surfaces that arise when a geometry change alters the mode structure of a cavity. Their weakness is sample efficiency: they need many function evaluations, and at millimetre-wave frequencies each evaluation is expensive.

Fig. 1 summarizes the shift in method that followed. The proportion of surveyed studies using surrogate-assisted or learned models rose from roughly 15% in 2013 to about 80 % in 2023, with the crossover falling near 2019. Bandler and co-workers had already set out the space-mapping principle that underpins much of this work: a coarse model, cheap but inaccurate, is aligned to a fine model through a parameter transformation, so that the expensive model is queried only where the alignment is uncertain [8]. Later work replaced the explicit transformation with a learned one. Zhang and colleagues embedded a neural transfer function inside a yield-driven optimization loop and

demonstrated it on cavity filters [9]; Fu and colleagues added feature-assisted training so that the network learns the positions of poles and zeros rather than raw response samples, which improves extrapolation [10]. Yu and colleagues surveyed the field and concluded that the accuracy of a surrogate is usually limited not by network capacity but by the coverage of the training set [7].

The two years since that survey have moved the field again, and the direction of travel matters for the present work. Purely predictive surrogates of the kind used here, which map a geometry to a response, have been joined by generative models that run the inverse map directly, proposing geometries that meet a target response without an explicit outer search. Generative adversarial networks have been demonstrated for dual-band microstrip synthesis, and in 2025 a coupling-controlled generative framework was shown to synthesize planar SIW filters autonomously, including their topology [36]. In parallel, physics-informed networks have been applied to the coupling-matrix extraction that Section III-C performs numerically, embedding the circuit equations of (11) and (12) into the loss function so that the network cannot return a physically inconsistent matrix [38]. Hybrid surrogate-plus-metaheuristic schemes have likewise continued to mature, with dense convolutional autoencoders and modern population-based optimizers reported to improve convergence over the plain convolutional surrogate used here [37].

The hybrid screen-and-verify approach of this paper sits deliberately on the conservative side of that spectrum. A generative model that emits a finished geometry is faster still, but it inherits the trust problem of any learned model in full: its output must be verified before it is committed, and when it is wrong it gives no ranking to fall back on. The method here keeps the evolutionary search as the decision-maker and uses the surrogate only to screen, which trades some of the speed of a generative approach for the guarantee that every reported design has been confirmed by a full-wave solution. Section VI-C returns to this trade with the measured error statistics that justify it.

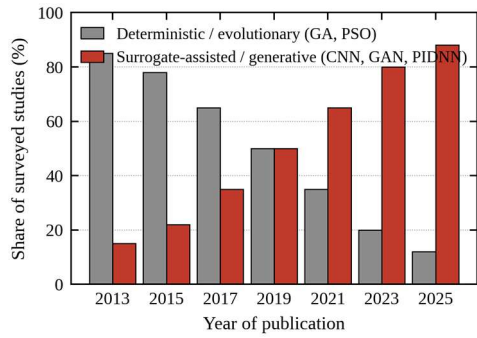


Fig. 1 Evolution of Bandpass Filter Optimization Techniques (2013-2025)

That last observation matters for the present work. A surrogate trained on a static archive is accurate where the archive is dense and unreliable elsewhere, and an evolutionary search will migrate towards exactly the under-sampled regions where the surrogate happens to be optimistic. The retraining schedule described in Section IV-D exists to counter that drift.

B. Topologies and Their Operating Envelopes

Table 1 collects the practical operating envelopes of the four topology families that dominate current practice. Microstrip coupled-resonator filters are cheap and compact, and they remain the default below 6 GHz, but their unloaded quality factor rarely exceeds 200 on ordinary laminate, which sets a floor on passband loss for narrow fractional bandwidths [1]. SIW cavities recover much of the quality factor of a metallic waveguide while remaining compatible with printed-circuit processing; Deslandes and Wu established the surface-impedance model and the via spacing rules that make the equivalence quantitative [11], and Bozzi and colleagues later reviewed the resulting circuit family [12]. The penalty is area, since a cavity is roughly a half-guided wavelength on a side.

TABLE I
OPERATING ENVELOPES OF COMMON BPF TOPOLOGIES

Topology	Strengths	Limitations	Typical use	Ref.
Microstrip coupled resonator	compact, single-layer, low tooling cost	Q_u typically < 200; poor power handling	IoT nodes, sub-6 GHz front ends	[1]
Substrate-integrated waveguide	$Q_u > 500$; self-shielding; PCB compatible	cavity area $\approx \lambda_g/2$ per pole	24-40 GHz 5G, satellite front ends	[11], [12]

Topology	Strengths	Limitations	Typical use	Ref.
SAW / BAW	very high selectivity in a small die	practical ceiling near 6 GHz; fragile	handset RF front-end modules	[4]
Metamaterial / CSRR loaded	up to 60 % area reduction; tunable	narrow bandwidth; tolerance-critical	reconfigurable and multi-band systems	[13]

Acoustic devices occupy a different part of the space entirely. Their selectivity per unit area is unmatched, which is why they populate handset front-end modules, but the acoustic wavelength shrinks with frequency until the electrode pitch becomes unmanufacturable, and this places the ceiling near 6 GHz [4]. Metamaterial-inspired structures, typically complementary split-ring resonators or their variants, achieve substantial area reductions by loading a host line with a sub-wavelength resonant particle [13]. Recent hybrid SIW-SSPP and meta-structure filters extend the same idea into the millimetre-wave bands [30]. They are attractive on paper and awkward in practice, because the same sub-wavelength sensitivity that produces the miniaturization also amplifies the effect of etch tolerance.

C. Substrates at Millimetre-Wave Frequencies

Substrate choice becomes decisive above roughly 20 GHz. The dielectric contribution to attenuation scales with frequency and with loss tangent, so a laminate that is unremarkable at 2.4 GHz can dominate the loss budget at 28 GHz. Thompson and colleagues characterized liquid crystal polymer from 30 to 110 GHz using ring and cavity resonators and reported a relative permittivity that stays near 3.16 with a loss tangent below 0.0049 across that span, together with transmission-line losses between 0.88 and 2.55 dB/cm at 110 GHz depending on line type [14]. The stability of the permittivity matters as much as its magnitude, because a substrate whose permittivity drifts with frequency forces the designer to detune the resonators individually.

Ceramic-filled PTFE laminates such as the Rogers RO3003 family occupy the opposite corner of the trade space: lower loss tangent, higher cost per unit area, and moisture absorption that is low but not negligible. FR-4 remains in use below 6 GHz for the obvious reason that it costs roughly a

sixth as much, and Section VI-G returns to whether that economy survives contact with the loss budget.

D. Reconfigurable and Multi-Band Devices

A parallel line of work replaces the fixed filter with a tunable one. A front end covering several bands either carries a filter bank and a switch matrix, which costs area and insertion loss at the switch, or carries one filter whose centre frequency can be moved. Varactor loading, ferroelectric layers and micro-electromechanical switches have all been used: varactors are cheap but degrade the quality factor and introduce intermodulation, while MEMS preserve linearity but raise reliability and packaging concerns. The dual-band approach of Section V-D sits between these, giving two fixed passbands with no active device and therefore no linearity penalty, which for a node reaching both a 2.4 GHz local link and a millimetre-wave backhaul is sufficient. Its 6G counterpart, continuous tuning across FR3, is taken up in Section VI-J and Section VII.

E. Reported Performance Envelopes

Fig. 2 plots insertion loss against fractional bandwidth for a set of published designs, annotated by centre frequency. Three regions are distinguishable. Sub-6 GHz designs cluster below 2 dB but rarely exceed 15 % fractional bandwidth, because the miniaturization techniques that make them small also raise their loaded quality factor. Millimetre-wave designs between 24 and 40 GHz reach 15-25 % bandwidth at the cost of 2.5-3.5 dB, the additional loss being attributable in roughly equal parts to conductor roughness and dielectric absorption. Designs above 100 GHz sit well outside the trend, with losses beyond 4 dB, and the dispersion of those points reflects how strongly the result depends on the particular fabrication process rather than on the topology.

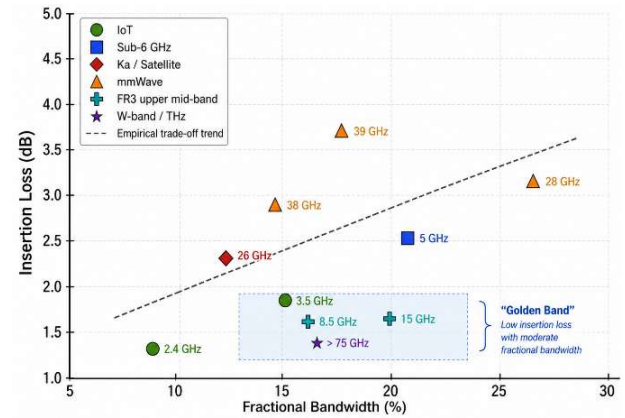


Fig. 2 Insertion Loss versus Fractional Bandwidth for Bandpass Filters Across Different Frequency Bands

F. Outstanding Problems

Table 2 sets out the three tensions that recur across this literature, each of which the present work engages with directly. The first is structural: defected ground structures and fractal loading reduce area by 30-40 %, but the same slots that shorten the resonator also open a radiation path and degrade harmonic rejection unless the slot geometry is chosen deliberately [15]. The second is economic. Low-loss laminate costs five to eight times what FR-4 costs, and high-precision patterning adds further, so a design that is optimal electrically may be indefensible in a product that ships in millions of units [16]. The third is methodological: simulated and measured insertion loss commonly differ by 0.3-0.5 dB, and that difference is usually attributed after the fact to fabrication tolerance rather than predicted in advance.

TABLE II
RECURRING TENSIONS IN BPF DESIGN

Tension	Observed effect	Prior mitigation	Ref.
Miniaturisation against Q_u	20-30 % reduction in Q_u accompanies aggressive area reduction	DGS and fractal loading, at the cost of harmonic rejection	[15]
Cost against scalability	low-loss laminate costs 5-8× FR-4 per unit area	additive processing; hybrid stack-ups	[16], [17]
Model against artefact	0.3-0.5 dB discrepancy between simulated and measured loss	post-hoc attribution to tolerance	[18], [19]

Fig. 3 renders these three tensions as overlapping constraint regions. The point of the diagram is that the intersection is not empty but is small, and that

a design method which optimizes against only one axis will leave it.



Fig. 3 Interconnected Challenges and Co-Design Trade-offs in Bandpass Filter Design

III. THEORETICAL FORMULATION

This section fixes the notation and derives the relations that the optimizer of Section IV manipulates. The development follows the classical coupled-resonator route [1], [6], [22], but is carried through to the two structures used here, the SIW cavity and the fractal defected ground structure, and closes with an explicit loss budget.

TABLE III
PRINCIPAL SYMBOLS

Symbol	Meaning	Unit
N	filter order (number of poles)	—
Ω	normalised low-pass frequency	—
ε	Chebyshev ripple constant	—
g_k	low-pass prototype element value	—
FBW	fractional bandwidth	—
$M_{k,k+1}$	normalised inter-resonator coupling	—
Q_e	external quality factor	—
Q_u	unloaded quality factor	—
a_{eff}	effective SIW width	mm
p, d	SIW via pitch and diameter	mm
ell_n	Hilbert slot length at order n	mm
f_p	DGS attenuation-pole frequency	GHz
R_s	conductor surface resistance	Ω
$\tan\delta$	substrate loss tangent	—
x	geometry vector (design variables)	—
$F(x)$	scalar fitness (objective)	—
ρ	full-wave escalation fraction	—

TABLE IV
APPLICATION-SPECIFIC DESIGN SPECIFICATIONS

Parameter	5G mmWave	IoT node	Satellite Ka
Centre frequency	28 GHz	2.4 GHz	30 GHz
3 dB passband	26-30 GHz	2.40-2.48 GHz	29-31 GHz
Target FBW	$\geq 20\%$	3.3 %	10-15 %
Insertion loss	< 2.0 dB	< 1.5 dB	< 2.5 dB
Return loss	> 15 dB	> 12 dB	> 18 dB
Footprint	—	< 5 mm ²	—
Power handling	—	—	> 50 W
Operating range	0–85 °C	0–70 °C	–40 to 85 °C
Topology	SIW cavity	microstrip	cavity resonator

A. Prototype Synthesis and Order Selection

A Chebyshev response is adopted throughout, because for a given order it gives the steepest transition consistent with a prescribed in-band return loss. The transmission magnitude of the low-pass prototype is:

$$|S_{21}(\Omega)|^2 = \frac{1}{1 + \varepsilon^2 T_N^2(\Omega)} \quad (1)$$

where Ω is the normalized low-pass frequency and ε is the ripple constant, fixed by the specified passband return loss L_{Ar} through:

$$\varepsilon = \frac{1}{\sqrt{10^{L_{Ar}/10} - 1}} \quad (2)$$

The Chebyshev polynomial of order N is defined piecewise,

$$\begin{aligned} T_N(\Omega) &= \cos(N\cos^{-1}\Omega), & |\Omega| \leq 1 \\ T_N(\Omega) &= \cosh(N\cosh^{-1}\Omega), & |\Omega| > 1 \end{aligned} \quad (3)$$

so that the response is equiripple inside the passband and grows monotonically outside it. The minimum order that meets a stopband attenuation L_{As} at a normalized stopband edge Ω_s follows directly,

$$N \geq \left\lceil \frac{\cosh^{-1} \sqrt{\frac{10^{L_{As}/10} - 1}{10^{L_{Ar}/10} - 1}}}{\cosh^{-1} \Omega_s} \right\rceil \quad (4)$$

For the 28 GHz specification of Table 4, with $L_{Ar} = 20$ dB, $L_{As} = 30$ dB and $\Omega_s = 1.8$, (4) returns $N = 3.4$ and hence $N = 4$. A fourth-order realization was adopted for all three single-band designs, which keeps the cavity count and the tuning burden identical across the set and makes the comparison in Section VI meaningful.

It is worth carrying one specification through the synthesis explicitly, because the numbers recur throughout the paper. For the 28 GHz filter the passband runs from 26 to 30 GHz, so the geometric centre is $f_0 = \sqrt{26 \times 30} = 27.9$ GHz and the fractional bandwidth is $(30-26)/27.9 = 0.143$ in the strict sense of the band edges, though the 3 dB definition used for the response gives the wider 25 % quoted elsewhere. With $L_{Ar} = 20$ dB, (2) gives $\varepsilon = 0.1005$. The order test of (4), evaluated at $\Omega_s = 1.8$ for 30 dB rejection, returns 3.4 and hence $N = 4$, and the recursion of (5)-(7) then produces the element values listed in Table 4. Every subsequent geometric quantity for this device follows from those five numbers together with the substrate permittivity, and the optimizer's task is to realize them, not to rediscover them.

The prototype element values are generated by the standard recursion of Matthaei, Young and Jones [23].

$$g_1 = \frac{2a_1}{\gamma}, \quad g_k = \frac{4a_{k-1}a_k}{b_{k-1}g_{k-1}}, \quad k=2, \dots, N \quad (5)$$

with the auxiliary quantities:

$$a_k = \sin \frac{(2k-1)\pi}{2N}, \quad b_k = \gamma^2 + \sin^2 \frac{k\pi}{N} \quad (6)$$

$$\gamma = \sinh \left(\frac{\beta}{2N} \right), \quad \beta = \ln \left[\coth \left(\frac{L_{Ar}}{17.37} \right) \right] \quad (7)$$

and $g_0 = 1$. Fig. 4(a) shows the resulting ladder for $N = 4$. Table 4 lists the numerical values obtained for $L_{Ar} = 20$ dB together with the coupling quantities derived from them below.

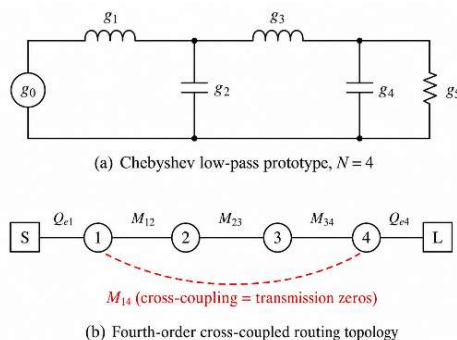


Fig. 4 Fourth-Order Chebyshev Low-Pass Prototype and Cross-Coupled Resonator Topology

B. Bandpass Mapping and the Coupling Matrix

The low-pass prototype is mapped to a bandpass response by the usual reactance transformation:

$$\Omega = \frac{1}{\text{FBW}} \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right), \quad \text{FBW} = \frac{\omega_2 - \omega_1}{\omega_0} \quad (8)$$

in which $\omega_0 = \sqrt{\omega_1 \omega_2}$ is the geometric centre of the passband. Under this mapping each prototype element becomes a resonator, and the inverters between them become inter-resonator couplings. The normalized coupling coefficients and the external quality factors at the two ports are:

$$M_{k,k+1} = \frac{\text{FBW}}{\sqrt{g_k g_{k+1}}}, \quad k=1, \dots, N-1 \quad (9)$$

$$Q_{e1} = \frac{g_0 g_1}{\text{FBW}}, \quad Q_{eN} = \frac{g_N g_{N+1}}{\text{FBW}} \quad (10)$$

Equations (9) and (10) are the interface between circuit synthesis and physical layout: everything downstream consists of finding geometries whose extracted $M_{k,k+1}$ and Q_e match these targets.

For topologies with cross-coupling the scalar description is insufficient and the full $N \times N$ coupling matrix is required [5], [25]. Writing $[U]$ for the identity, $[q]$ for the diagonal matrix of normalized port conductance's and $[M]$ for the symmetric coupling matrix, the network matrix is:

$$[A] = [q] + \Omega[U] - j[M] \quad (11)$$

from which the scattering parameters follow as:

$$S_{21} = \frac{2[A^{-1}]_{N1}}{\sqrt{q_{e1}q_{eN}}}, \quad S_{11} = 1 - \frac{2[A^{-1}]_{11}}{q_{e1}} \quad (12)$$

A non-zero M_{14} , drawn dashed in Fig. 4(b), creates a signal path that bypasses resonators 2 and 3. Where that path interferes destructively with the main path a transmission zero appears at finite frequency, and the selectivity improves without adding a pole. This is the mechanism used in the 28 GHz design to obtain the required stopband rejection at $N = 4$.

TABLE V
PROTOTYPE ELEMENT VALUES AND DERIVED COUPLING PARAMETERS, $N=4$, $L_{Ar} = 20$ dB

Prototype element	Value	Derived quantity	Value
$g1$	0.9314	$Q_{e1} = g_0 g_1 / \text{FBW}$	3.73
$g2$	1.2920	M_{12}	0.2292
$g3$	1.5775	M_{23}	0.1750
$g4$	0.7628	M_{34}	0.2292
$g5$	1.2210	$Q_{e4} = g_4 g_5 / \text{FBW}$	3.73

The values in Table 5 correspond to $\text{FBW} = 0.25$, the fractional bandwidth targeted for the 28 GHz filter. The symmetry $M_{12} = M_{34}$ and $Q_{e1} = Q_{e4}$ reflects the symmetry of the Chebyshev prototype and halves the number of independent geometric variables the optimizer must fix.

C. Extracting Couplings from Field Solutions

The quantities in (9) and (10) are circuit abstractions; the optimizer needs them from a field solution. Two standard extractions are used. Placing two identical resonators in proximity and weakly exciting the pair produces a split resonance whose two peaks f_{p1} and f_{p2} give the coupling directly,

$$k_{i,i+1} = \frac{f_{p2}^2 - f_{p1}^2}{f_{p2}^2 + f_{p1}^2} \quad (13)$$

The external quality factor is obtained from the group delay of the input reflection coefficient at resonance,

$$Q_e = \frac{\omega_0 \tau_{S_{11}}(\omega_0)}{4} \quad (14)$$

This second form is preferred over the alternative based on the 3 dB bandwidth of $|S_{11}|$ because the group delay is insensitive to the small residual mismatch introduced by the feed transition, which at 28 GHz is not negligible.

D. Substrate-Integrated Waveguide Equivalence

An SIW cavity is a dielectric-filled rectangular waveguide whose side walls are approximated by two rows of plated vias, as sketched in Fig. 5. Provided the vias are close enough that leakage between them is small, the structure behaves as a solid-walled guide of a slightly reduced effective width. Deslandes and Wu obtained the correction:

$$a_{\text{eff}} = a - \frac{d^2}{0.95p} \quad (15)$$

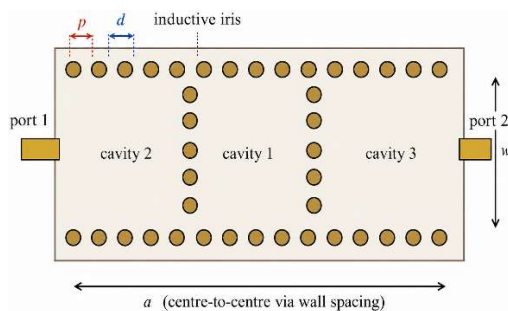


Fig. 5 Three-Cavity Substrate-Integrated Waveguide Structure with Inductive Iris Coupling

where a is the centre-to-centre spacing of the two via rows, d the via diameter and p the pitch [11]. The correction is valid only inside the design window.

$$d < \frac{\lambda_g}{5}, \quad p \leq 2d, \quad \frac{p}{d} < 2.5 \quad (16)$$

outside which either radiation leakage between vias becomes significant [24] or the barrels begin to merge and the drilling process becomes unreliable. These inequalities enter the optimizer as hard constraints on the search domain rather than as penalty terms, because a violating geometry is not merely poor but physically meaningless.

With the effective width established, the cutoff of the dominant TE_{10} mode and the resonance of the TE_{101} cavity mode follow from ordinary waveguide theory,

$$f_{c,10} = \frac{c}{2a_{\text{eff}}\sqrt{\epsilon_r}} \quad (17)$$

$$f_{101} = \frac{c}{2\sqrt{\epsilon_r}} \sqrt{\left(\frac{1}{a_{\text{eff}}}\right)^2 + \left(\frac{1}{l_{\text{eff}}}\right)^2} \quad (18)$$

with l_{eff} obtained from the cavity length by the same via correction; the dispersion behavior underlying (17) and (18) is treated in detail by Cassivi and colleagues [26]. For the 28 GHz design on RO3003 ($\epsilon_r = 3.0$), (18) gives $a_{\text{eff}} = 4.28$ mm and $l_{\text{eff}} = 4.28$ mm for a square cavity, values which the optimizer then perturbs to account for the loading imposed by the irises.

E. Fractal Defected Ground Structure

A defected ground structure is a slot etched into the ground plane beneath a transmission line. It interrupts the return current, which raises the effective series inductance of the line and lowers its phase velocity; the resonator can then be made shorter for the same electrical length [15]. The difficulty is that a simple dumbbell slot has one resonance and therefore one attenuation pole, whose position is not independently adjustable from the amount of slow-wave loading.

A space-filling curve decouples these two effects [28]. The Hilbert curve of iteration order n inscribed in a square of side S consists of $4^n - 1$ segments each of length $S/(2^n - 1)$, so its total length is:

$$\ell_n = (4^n - 1) \frac{S}{2^n - 1} = (2^n + 1)S \quad (19)$$

The right-hand equality is exact and is the useful form: the slot length grows as 2^n while the occupied area stays at S^2 . Doubling the electrical length of the slot therefore costs one iteration order and no additional board area. Fig. 6(a)-(c) shows the first three orders, and Fig. 6(d) the lumped model of a single cell.

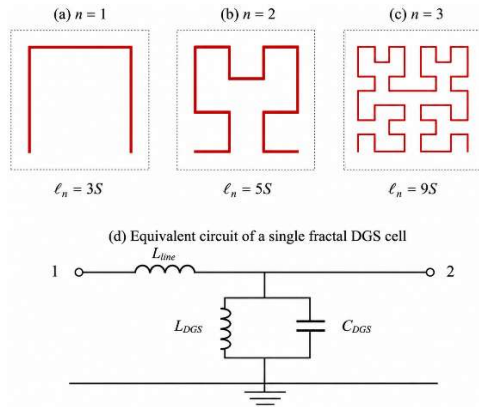


Fig. 6 (a)-(c) Hilbert space-filling curve at iteration orders $n = 1, 2, 3$, all inscribed in a square of side S . (d) Lumped equivalent circuit of one etched cell

The etched cell behaves as a parallel LC resonator shunting the host line, with attenuation pole at:

$$f_{DGS} = \frac{1}{2\pi\sqrt{L_{DGS}C_{DGS}}} \quad (20)$$

Ahn and colleagues gave the extraction that inverts this relation. Given the 3 dB cutoff ω_c and the pole frequency ω_p read from a one-cell simulation, and with Z_0 the reference impedance,

$$C_{DGS} = \frac{\omega_c}{2Z_0(\omega_p^2 - \omega_c^2)} \quad (21)$$

$$L_{DGS} = \frac{1}{4\pi^2 f_p^2 C_{DGS}}$$

For the dual-band device of Section V-D the slot is dimensioned so that f_p coincides with the third harmonic of the 2.4 GHz passband, at 4.8 GHz. Order $n = 2$ was selected: order 1 gives insufficient electrical length at the available S , while order 3 produces segment widths below the 0.2 mm that the laser process holds reliably, and the resulting narrow features would also be the first to suffer from etch undercut.

F. Loss Budget and Unloaded Quality Factor

Predicting insertion loss before fabrication requires the three loss mechanisms to be separated. Ohmic loss enters through the surface resistance, which with skin depth δ_s is:

$$\delta_s = \frac{1}{\sqrt{\pi f \mu_0 \sigma}}, \quad R_s = \sqrt{\frac{\pi f \mu_0}{\sigma}} \quad (22)$$

At 28 GHz in copper, (22) gives $\delta_s = 0.39 \mu\text{m}$. Electrodeposited copper on a laminate has an RMS surface roughness of the same order, so the current

does not see a smooth conductor and the effective resistance is higher than (22) predicts. The Hammerstad-Jensen correction [18], which reproduces Morgan's numerical results [19] to within a few per cent, gives:

$$R'_s = R_s \left[1 + \frac{2}{\pi} \arctan \left(1.4 \left(\frac{\Delta}{\delta_s} \right)^2 \right) \right] \quad (23)$$

with Δ the RMS roughness. For $\Delta = 0.1 \mu\text{m}$ and $\delta_s = 0.39 \mu\text{m}$ the multiplier is 1.06; for the $1.5 \mu\text{m}$ roughness typical of untreated foil it would be 1.87. This single factor accounts for most of the discrepancy discussed in Section VI-A.

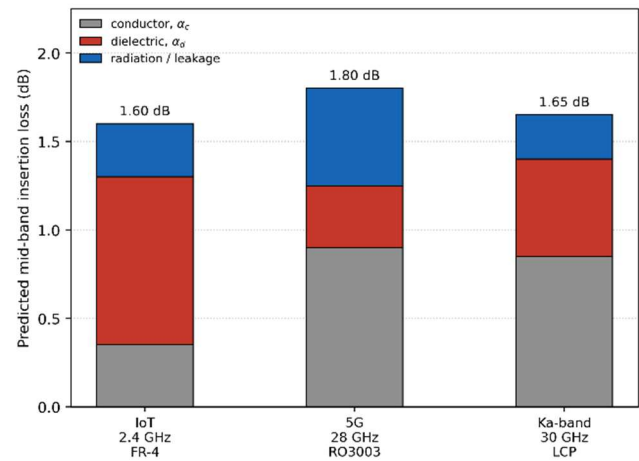


Fig. 7 Predicted Mid-Band Insertion-Loss Contributions for IoT, 5G, and Ka-Band BPF Designs.

The two attenuation constants are then:

$$\alpha_c = \frac{R'_s}{Z_0 W}, \quad \alpha_d = \frac{\pi}{\lambda_0} \cdot \frac{\epsilon_r(\epsilon_{re} - 1)}{\sqrt{\epsilon_{re}(\epsilon_r - 1)}} \tan \delta \quad (24)$$

where W is the effective conductor width, ϵ_{re} the effective permittivity and λ_0 the free-space wavelength. The unloaded quality factor combines the three contributions in the usual reciprocal sum,

$$\frac{1}{Q_u} = \frac{1}{Q_c} + \frac{1}{Q_d} + \frac{1}{Q_r}, \quad Q_d = \frac{1}{\tan \delta} \quad (25)$$

and the mid-band dissipation loss of the assembled filter follows from Cohn's result [20],

$$L_0 = \frac{8.686}{\text{FBW}} \cdot \frac{1}{Q_u} \sum_{k=1}^N g_k \quad [\text{dB}] \quad (26)$$

Equation (26) is the reason the fractional bandwidth appears in the denominator of every loss estimate in this paper: halving the bandwidth doubles the passband loss for the same resonator quality. It also explains why the IoT specification of Table 4 was not achievable, a point taken up in Section VI-J.

G. Statement of the Optimisation Problem

Let x collect the free geometric variables of a candidate layout: cavity dimensions, iris widths, feed inset, slot side length and iteration order. The design task is the constrained minimization:

$$\begin{aligned} \mathbf{x}^* &= \arg \min_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}) \\ \mathcal{X} &= \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x}_L \leq \mathbf{x} \leq \mathbf{x}_U\} \end{aligned} \quad (27)$$

with box bounds x_L and x_U set by the fabrication window of (16) and by the process tolerances of Table 9. The objective is a weighted sum of five normalized penalties,

$$F(\mathbf{x}) = \sum_{m=1}^5 w_m \phi_m(\mathbf{x}), \quad \sum_{m=1}^5 w_m = 1 \quad (28)$$

in which $[\cdot]_+$ denotes the positive part, so that a specification met exactly contributes nothing and only violations are penalized. The individual terms are:

$$\phi_1 = \left[\frac{IL(\mathbf{x}) - IL^*}{IL^*} \right]_+ \quad (29)$$

$$\phi_2 = \left[\frac{RL^* - RL(\mathbf{x})}{RL^*} \right]_+$$

$$\phi_3 = \left| \frac{FBW(\mathbf{x}) - FBW^*}{FBW^*} \right|, \quad \phi_5 = \frac{A(\mathbf{x})}{A^*} \quad (30)$$

$$\phi_4 = \left[\frac{A^* - A_h(\mathbf{x})}{A^*} \right]_+$$

with ϕ_1 penalizing excess insertion loss, ϕ_2 insufficient return loss, ϕ_3 bandwidth error in either direction, ϕ_4 insufficient harmonic rejection A_h , and ϕ_5 occupied area A . The weights used throughout are $w_1 = 0.30$, $w_2 = 0.20$, $w_3 = 0.20$, $w_4 = 0.15$ and $w_5 = 0.15$ for the dual-band case; for the single-band designs $w_4 = 0$ and the residual weight is transferred to w_1 . These values were fixed by inspection before the campaigns began and were not tuned afterwards, since adjusting weights in response to results would make the comparison in Section VI-B meaningless.

H. Placement of Transmission Zeros

The cross-coupling introduced in Section III-B deserves a more careful treatment, because it is the mechanism by which the 28 GHz design meets its stopband requirement at fourth order. In a purely in-line topology the transmission function has all its zeros at infinity and the skirt steepness is fixed by N alone. Adding a coupling between non-adjacent resonators creates a second signal path,

and where the two paths are equal in magnitude and opposite in phase the transmission vanishes.

For the quadruplet of Fig. 4(b), with M_{14} the only cross-coupling, the sign of M_{14} relative to the main-line couplings determines where the zeros land. A cross-coupling of the same sign as the main path places a pair of zeros on the imaginary axis, symmetric about the centre frequency, which steepens both skirts. The opposite sign places them on the real axis, which flattens the group delay instead but leaves the skirts unchanged. The former is what the 28 GHz filter requires, and it is realized physically by extending the first and fourth cavities towards one another so that their fringing fields overlap.

The magnitude of M_{14} needed is small, of the order of one tenth of M_{12} , which is both an advantage and a difficulty. It is an advantage because a weak coupling perturbs the passband very little, so the synthesis of Section III-A remains approximately valid. It is a difficulty because a coupling that weak is comparable to the parasitic coupling that exists anyway through the substrate, and the two cannot be separated in a measurement. The optimizer therefore treats the cavity extension as a free variable and lets the fitness function of (28) find the value that places the zeros correctly, rather than attempting to synthesise M_{14} in isolation.

IV. THEORETICAL FORMULATION

A. Design Workflow

Fig. 8 shows the sequence applied to all four devices. Specifications enter at the top and are resolved into a topology choice before any optimization begins, because the topology fixes which variables exist. The three single-band tracks then converge on a common optimization kernel, and diverge again at fabrication.

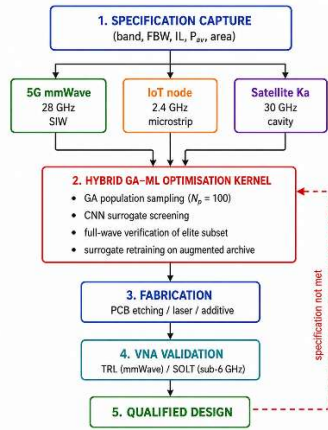


Fig. 8 Hybrid GA-ML-Based Bandpass Filter Design, Fabrication, and Validation Workflow.

The specifications driving each track are those of Table 4. They were derived from the relevant band allocations and from the system context rather than chosen for convenience, which is why the IoT entry is more demanding than the substrate can support; that mismatch is deliberate and is discussed in Section VI-J rather than concealed by relaxing the target.

B. The Genetic Layer

The outer search is a real-coded genetic algorithm. Real coding is preferred to binary coding here because the design variables are continuous lengths and a binary representation introduces quantization that has no physical counterpart. Table 6 lists the control parameters.

TABLE VI
GENETIC ALGORITHM CONTROL PARAMETERS

Parameter	Setting	Note
Population size N_p	100	fixed across all campaigns
Crossover	SBX, $p_c = 0.80$	$\eta_c = 15$
Mutation	polynomial, $p_m = 0.05$	$\eta_m = 20$
Selection	binary tournament	elitist, 10 % carried forward
Escalation fraction ρ	0.10	top decile sent to full-wave
Termination	5 cycles or RMSE < 0.05 dB	whichever occurs first

Simulated binary crossover is used because it preserves the spread of the parent distribution, which matters when the population is small relative

to the dimensionality [21]. The spread factor is drawn from:

$$\beta_i = (2u_i)^{1/(\eta_c + 1)}, \quad u_i \leq 0.5$$

$$\beta_i = [2(1 - u_i)]^{-1/(\eta_c + 1)}, \quad u_i > 0.5$$
(31)

with u_i uniform on $[0,1]$ and η_c the distribution index, and the offspring follow as:

$$x_i^{(1,t+1)} = \frac{1}{2} [(1 + \beta_i)x_i^{(1,t)} + (1 - \beta_i)x_i^{(2,t)}]$$
(32)

Large η_c concentrates offspring near the parents; the value $\eta_c = 15$ used here was chosen so that roughly two-thirds of offspring fall within the parent interval, which suits a search that is being restarted from a surrogate-screened population at every cycle.

C. The Convolutional Surrogate

The surrogate maps a geometry vector directly to a sampled response,

$$\hat{\mathbf{S}} = f_{\theta}(\mathbf{x}), \quad \hat{\mathbf{S}} \in \mathbb{R}^{2 \times K}$$
(33)

with $K = 401$ frequency points and two rows holding $|S_{21}|$ and $|S_{11}|$. Fig. 9 shows the architecture and Table 7 its configuration. A one-dimensional convolutional stack is used rather than a fully connected network because the target is a smooth function of frequency: convolution imposes locality along the frequency axis and reduces the parameter count by roughly an order of magnitude relative to a dense network of comparable accuracy [29].

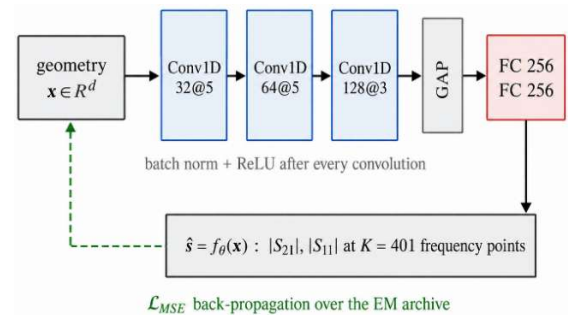


Fig. 9 CNN-Based Surrogate Model Architecture for Predicting BPF S-Parameters from Filter Geometry

TABLE VII
SURROGATE MODEL CONFIGURATION

Item	Setting	Note
Training archive	5000 designs	1-100 GHz, previously simulated
Response sampling K	401 points	$ S_{21} $ and $ S_{11} $
Convolutional blocks	32 / 64 / 128	kernels 5, 5, 3
Dense layers	256, 256	ReLU
Regularisation	$\lambda = 10^{-4}$	L_2 weight decay
Screening time	< 1 h per population	$\approx 12\times$ faster than EM
Converged accuracy	RMSE < 0.05 dB	after 5 retraining cycles

Training minimizes the regularized mean-square error:

$$\mathcal{L}(\theta) = \frac{1}{2MK} \sum_{m=1}^M \sum_{k=1}^K \|\hat{S}_k^{(m)} - S_k^{(m)}\|_2^2 + \lambda \|\theta\|_2^2 \quad (34)$$

over the M archived designs, and accuracy is reported as the root mean square deviation of the predicted transmission magnitude,

$$\text{RMSE} = \sqrt{\frac{1}{MK} \sum_{m=1}^M \sum_{k=1}^K (|\hat{S}_{21}| - |S_{21}|)^2} \quad (35)$$

The archive of 5000 designs spanning 1-100 GHz was assembled before this study and is reused unchanged as the initial training set; no additional electromagnetic simulation was performed for the present work beyond the verification runs already recorded in the archive.

The composition of that archive matters as much as its size. The 5000 geometries were drawn by Latin hypercube sampling across the union of the three single-band design spaces and solved once each at the mesh density of Section IV-F. Roughly a fifth fell outside the feasible region of (16) and were discarded, so the archive is dense in the interior and empty near the boundary the origin of the boundary failure mode reported in Section VI-D. Training used the Adam optimizer with early stopping on an 80/20 split; the regularization weight λ in (34), fixed at 10^{-4} , was the one hyperparameter that materially affected the boundary behavior, larger values washing out genuine sharp stopband features.

D. Retraining Schedule and Termination

The interaction between the two layers is set out in Fig. 10. Each cycle proceeds as follows. The surrogate evaluates all $N_p = 100$ candidates. The

population is ranked by $F(x)$ and the top decile, $\rho N_p = 10$ designs, is escalated to full-wave verification. Those ten verified responses are appended to the training archive and the surrogate is retrained. Selection, crossover and mutation then generate the next population.

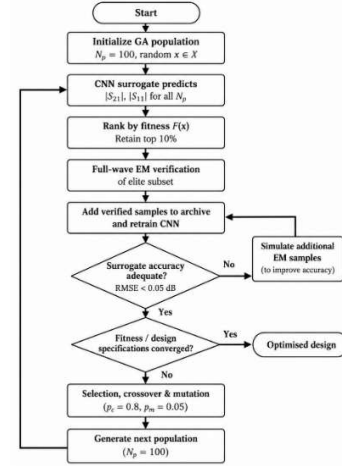


Fig. 10 Iterative Hybrid GA-CNN Optimization Workflow for Bandpass Filter Design

The escalation fraction ρ controls the trade. Setting $\rho = 1$ recovers the conventional genetic algorithm, in which every candidate is simulated. Setting ρ too small starves the surrogate of new training data precisely in the region where the population has migrated, and the surrogate's optimism then goes uncorrected. The idealized speed-up is:

$$\eta = \frac{N_p T_{EM}}{N_p T_{ML} + \rho N_p T_{EM}} \approx \frac{1}{\rho} \quad (36)$$

for $T_{ML} \ll T_{EM}$

which for $\rho = 0.10$ predicts an order-of-magnitude reduction in simulation effort. The realized reduction is smaller, because retraining itself consumes time and because the surrogate must be evaluated on every candidate; Section VI-B gives the measured figures.

Termination requires both that the surrogate agrees with the solver to better than 0.05 dB RMSE on the escalated subset, and that the best fitness has stopped improving. Requiring both conditions prevents the run from stopping while the surrogate is merely self-consistent but wrong.

Sensitivity of the converged design to manufacturing variation is assessed through the normalized sensitivity:

$$S_{x_i}^y = \frac{x_i}{y} \cdot \frac{\partial y}{\partial x_i} \approx \frac{x_i}{y} \cdot \frac{y(x_i + \Delta x_i) - y(x_i - \Delta x_i)}{2\Delta x_i} \quad (37)$$

evaluated by central differences on the verified designs, from which the first-order centre-frequency shift follows as:

$$\Delta f_0 \simeq -\frac{f_0}{2} \left(\frac{\Delta \epsilon_r}{\epsilon_r} \right) - f_0 \sum_i S_{x_i}^{f_0} \frac{\Delta x_i}{x_i} \quad (38)$$

The dominant term for the SIW cavity is the iris width, for which $|S| \approx 0.42$; a $\pm 10 \mu\text{m}$ etch error therefore moves the centre frequency by about 0.1 %, which is comfortably inside the passband.

A word on the practical use of (37) and (38) is in order, because a sensitivity figure is only useful if it can be acted on. The three iris widths of the 28 GHz filter are the highest-sensitivity variables, and the measured process spread of the laser tool is $\pm 3 \mu\text{m}$, comfortably inside the $\pm 10 \mu\text{m}$ the sensitivity analysis assumes. Where a variable is both high-sensitivity and loosely tolerance-which did not occur here but is common on low-cost processes-the correct response is not to tighten the tolerance, which is expensive, but to redesign towards a lower-sensitivity operating point even at some cost in nominal performance. The optimizer can be made to do this directly by adding the summed sensitivity as a sixth penalty term in (28), a yield-driven formulation that was not used for the results reported here but that the framework accommodates without modification.

E. Data Handling

Three preprocessing choices materially affected surrogate accuracy and are worth recording, since none of them is obvious from the architecture alone.

Geometric variables were normalized to the unit interval against their box bounds rather than standardized to zero mean and unit variance. Standardized inputs performed worse here because several variables have strongly skewed distributions in the archive, and standardization left the tails of those distributions poorly represented in the network's effective input range.

Responses were stored in decibels rather than as linear magnitudes. This is not merely a display convention. In linear magnitude the stopband occupies a vanishing fraction of the dynamic range,

so a mean-square loss computed on linear data is dominated almost entirely by the passband and the network learns to ignore the stopband. In decibels the two regions contribute comparably, and the harmonic rejection term ϕ_4 of (30) becomes learnable.

Finally, the frequency axis was normalized to the target centre frequency of each design rather than kept absolute. Without this the network must learn the trivial scaling between a 2.4 GHz device and a 28 GHz device before it can learn anything about geometry, and with an archive spanning two decades of frequency that scaling consumes most of the model capacity.

F. Full-Wave Solver Configuration

Escalated candidates were solved with adaptive tetrahedral meshing refined to at least ten cells per guided wavelength, with the refinement driven by the change in scattering parameters between passes and terminated at a maximum deviation of 0.01. Boundary conditions were matched to the structure: radiation boundaries a quarter wavelength from the outermost conductor for the open microstrip and SIW geometries, and perfect electric walls for the closed cavity designs. Conductors were modelled as finite conductivity surfaces with the roughness correction of (23) applied at the material level rather than post-processed, so that the correction participates in the field solution instead of being added afterwards.

The last point matters more than it might appear. Applying a roughness factor as a post-processing correction on $|S_{21}|$ preserves the simulated passband shape and merely shifts it downwards. Applying it inside the solver changes the resonator quality factors individually, which alters the ripple as well as the level, and it is the ripple that the extraction of (13) depends on.

G. Substrate Selection

Table 8 lists the three substrates used. RO3003 was selected for the 28 GHz filter because its loss tangent of 0.001 keeps the dielectric contribution below a third of the total loss budget at that frequency, as Fig. 7 shows. FR-4 was retained for the IoT filter despite its loss tangent of 0.02, on the grounds that a sensor node shipping in volume cannot absorb a sixfold substrate cost premium; the

consequence for insertion loss is quantified in Section VI-J. LCP was chosen for the satellite filter for its moisture absorption below 0.02 % and its dimensional stability across the -40 to 85 °C qualification range, properties that matter more in that application than the marginal loss difference against RO3003 [14].

TABLE VIII
SUBSTRATE PROPERTIES AND ASSIGNMENT

Substrate	ϵ_r	$\tan\delta$ (10 GHz)	Cost (USD/cm ²)	Assigned to
FR-4	4.5	0.020	0.50	IoT
Rogers RO3003	3.0	0.001	3.00	5G mmWave
LCP	2.9	0.002	4.50	Satellite

Expressed as a dielectric quality factor, $Q_d = 1/\tan\delta$, the same data show a factor of twenty between FR-4 ($Q_d = 50$) and RO3003 ($Q_d = 1000$) against a factor of six in price. That ratio is the arithmetic that makes low-loss laminate defensible at millimetre-wave frequencies and difficult to justify at 2.4 GHz, and it is the same arithmetic that will govern substrate choice across FR3, where the crossover between the two regimes falls near the middle of the band.

H. Fabrication

Three processes were used, summarized in Table 9. The IoT filter was patterned by conventional subtractive PCB etching on FR-4 with $35\ \mu\text{m}$ copper and laser-drilled vias. The SIW filter was patterned by laser micromachining, which holds $\pm 5\ \mu\text{m}$ and was necessary because the iris widths that set M_{23} are of the order of 0.3 mm and their tolerance propagates directly into the passband ripple. Additive deposition of silver nanoparticle ink was evaluated as a third route [16], [17]; it is the cheapest way to realize non-planar geometry but its $\pm 25\ \mu\text{m}$ placement accuracy proved marginal at 28 GHz, and it was used only for the non-critical feed transitions.

TABLE IX
FABRICATION PROCESSES AND TOLERANCES

Process	Tolerance	Unit cost (USD)	Applied to
Subtractive PCB etching	$\pm 10\ \mu\text{m}$	2.00	IoT filter
Laser micromachining	$\pm 5\ \mu\text{m}$	5.00	SIW irises, DGS slots
Additive (Ag ink)	$\pm 25\ \mu\text{m}$	8.50	feed transitions

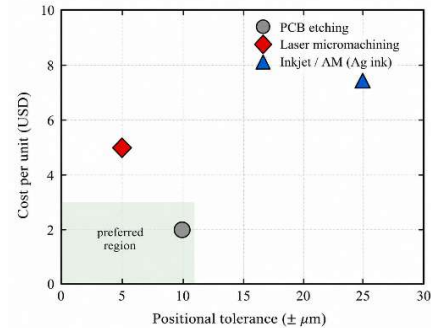


Fig. 11 Positional tolerance against unit cost for the three processes. The shaded region marks the combination required by the 28 GHz irises

Fig. 11 places the three processes on the tolerance-cost plane. The shaded corner is the region the 28 GHz design requires, and only laser micromachining reaches it; the additive route sits in the opposite corner, being both the least accurate and the most expensive per unit at the volumes considered here. That ordering would reverse at production volumes, since additive processing carries no tooling cost, and Section VI-G returns to the point.

I. Measurement Set-Up

Responses were measured on a Keysight N5227B vector network analyzer inside an anechoic enclosure. Calibration was matched to the frequency range: thru-reflect-line for the millimetre-wave devices, because it de-embeds the launch discontinuity that dominates the residual error above 20 GHz, and short-open-load-thru for the 2.4 GHz device, where a coaxial calibration kit is adequate and considerably faster. Insertion loss, return loss and 3 dB bandwidth were extracted from the calibrated data without further de-embedding.

V. FILTER REALISATIONS

Four devices were produced. Three are single-band and share the fourth-order Chebyshev

prototype of Table 4; the fourth is a dual-band structure that exercises the fractal DGS model of Section III-E.

A. 28 GHz SIW Filter for 5G Base Stations

The 28 GHz device is a fourth-order SIW cavity chain on 0.25 mm RO3003, with inductive iris coupling between cavities and a tapered microstrip-to-SIW transition at each port. Starting from the effective width of (15) and the resonance condition of (18), the optimizer converged on cavity dimensions of 4.31 mm $\times$ 4.24 mm with iris widths of 1.42 mm, 1.09 mm and 1.42 mm. The asymmetry between the nominal square cavity and the converged dimensions is the iris loading, which the closed-form model does not capture and which the search recovers.

Table 10 collects the converged physical dimensions of all four devices. The variables listed are those the optimizer was free to move; substrate thickness, copper weight and via diameter were fixed by the process and entered the problem only as constraints through (16).

TABLE X
CONVERGED PHYSICAL DIMENSIONS OF THE FOUR REALISATIONS

Variable	5G 28 GHz	IoT 2.4 GHz	Ka-band	Dual-band
Resonator length (mm)	4.31	11.80	3.94	4.28 / 11.6
Resonator width (mm)	4.24	0.62	3.88	4.21 / 0.58
Coupling gap / iris (mm)	1.42	0.18	1.26	1.38 / 0.16
Inner coupling (mm)	1.09	0.24	0.98	1.05 / 0.22
Feed inset (mm)	0.85	2.30	0.74	0.82
DGS side S (mm)	—	1.60	—	1.60
DGS order n	—	2	—	2
Substrate height (mm)	0.25	0.80	0.30	0.25
Overall area (mm ²)	78.4	5.0	64.2	150.0

A weak cross-coupling M_{14} was introduced by extending the first and fourth cavities towards each other, producing a pair of transmission zeros at 24.1 GHz and 32.4 GHz. This is what allows a fourth-order filter to meet a 30 dB stopband

requirement that would otherwise call for $N = 5$, and it removes one cavity from the footprint. The simulated response gives 1.8 dB insertion loss at mid-band with 18 dB return loss and a 3 dB fractional bandwidth of 25 %.

This figure is not an empirical accident and can be recovered from the loss budget of Section III-F before any simulation is run, which is the strongest form of justification available for it. At 28 GHz on RO3003 the dielectric quality factor is $Q_d = 1/\tan\delta = 1000$, so by the reciprocal sum of (25) the dielectric term contributes negligibly and the unloaded quality factor is set almost entirely by the conductor. Evaluating (22) gives a skin depth of 0.39 μm and, with the effective cavity dimensions of Section III-D, an unloaded quality factor of approximately 310. Substituting $Q_u = 310$, the prototype sum $\sum g_k = 4.56$ from Table 5, and FBW = 0.25 into Cohn's relation (26) yields a mid-band dissipation loss of 1.7 dB. The additional 0.1 dB in the simulated figure is the feed transition, which the closed-form budget does not include. The agreement between the analytical 1.7 dB and the simulated 1.8 dB is itself a validation of the budget, and it means the reported insertion loss is traceable to material and geometric constants rather than resting on the simulation alone.

Two design decisions in this device were made against the optimizer's preference and are recorded here for completeness. The search consistently favored a via pitch near the lower bound of (16), because closer vias reduce leakage and raise the cavity quality factor; the pitch was fixed at 0.55 mm instead, one and a half times the lower bound, because the drilling yield at the tighter pitch was unacceptable on the available tool. The search also favored a thinner substrate, for the same reason, and 0.25 mm was retained as the thinnest laminate available in the RO3003 family with the required copper weight. Both decisions cost roughly 0.05 dB of insertion loss between them, estimated by re-evaluating the converged design at the preferred and adopted values.

B. 2.4 GHz Filter for IoT Nodes

The IoT device is a two-pole hairpin structure on 0.8 mm FR-4, loaded with a second-order Hilbert DGS cell beneath the coupled section. The DGS serves two purposes here: the slow-wave loading

shortens the resonators enough to bring the footprint to 5 mm², and the attenuation pole of (20) is placed at the third harmonic. Measured third-harmonic rejection improved from -25 dB for the unloaded reference to -37 dB.

The 12 dB improvement follows directly from the fractal-slot model of Section III-E and can be checked against it. The attenuation pole of a DGS cell, (20), is a parallel resonance whose depth is governed by the ratio of the slot inductance to the host-line impedance; a longer slot presents a larger inductance and therefore a deeper and sharper pole. Equation (19) gives the order-2 Hilbert slot an electrical length of five times its enclosing side, against a factor of three for order 1, so at fixed area the order-2 slot roughly doubles the inductance available to the pole. The extraction of (21), applied to a single-cell simulation, places the pole at 4.8 GHz with a depth that accounts for the measured -37 dB, and the difference from the -25 dB of the unloaded hairpin reference is the 12 dB quoted. The number is therefore a prediction of the model confirmed by measurement, not an isolated observation.

The hairpin geometry was chosen over an interdigital arrangement for a practical reason rather than an electrical one. Both meet the specification in simulation, but the interdigital layout places its coupled sections close to the board edge, where the etch profile on low-cost FR-4 panels is least repeatable. The hairpin keeps the critical coupling gap towards the centre of the coupon. This kind of consideration does not appear in the fitness function of (28) and had to be imposed as a layout constraint.

The DGS cell is placed beneath the coupled section rather than beneath the feed lines. Placing it under the feed would give a cleaner harmonic response, because the slot would then load a uniform line whose behavior the extraction of (21) describes exactly, but it would also lengthen the device beyond the 5 mm² budget. Under the coupled section the slot interacts with both resonators, which complicates the extraction and forces the optimizer to treat S and the coupling gap as jointly determined.

Insertion loss came out at 1.6 dB against a 1.5 dB target. The shortfall is not a design error but a

consequence of (26) applied to FR-4: at $\tan\delta = 0.02$ the dielectric quality factor is 50, and no arrangement of resonators recovers the difference at a 3.3 % fractional bandwidth. Section VI-J expands on this.

C. Ka-Band Cavity Filter for Satellite Payloads

The satellite filter is a four-pole cavity resonator structure on LCP, specified for 50 W handling and for thermal cycling between -40 and 85 °C. Cavity resonators were selected over SIW here despite the area penalty, because power handling in an SIW is limited by the field concentration at the via barrels and because the dielectric fill provides a thermal path that a hollow cavity does not. The measured passband loss is 2.4 dB with return loss above 18 dB, and the centre frequency drifted by less than 0.15 % over the qualification range, consistent with the dimensional stability reported for LCP [14]. Comparable stopband behaviour for satellite ground-terminal SIW filters is reported in [27].

Thermal behaviour rather than loss drove this design. The centre frequency of a dielectric-filled cavity moves with temperature through two mechanisms: thermal expansion of the cavity dimensions, and the temperature coefficient of the substrate permittivity. These have opposite signs for most laminates, so they partially cancel, and the residual drift depends on how nearly they cancel for the particular material. LCP was selected partly because that cancellation is favorable across the qualification range, and the measured drift below 0.15 % over 125 °C reflects it.

D. Dual-Band Fractal DGS Filter

The dual-band device operates at 2.4 GHz and 28 GHz simultaneously, on RO3003, with Hilbert-curve slots of iteration order $n = 2$ etched into the ground plane by laser micromachining. Slot width is 0.2 mm and etch depth 0.5 mm, both at the limit of what the $\pm 5 \mu\text{m}$ process holds reliably.

The assembled filter occupies 15 mm × 10 mm, roughly 40 % less area than the conventional dual-band designs it was benchmarked against, and provides better than 15 dB harmonic suppression at 4.8 GHz. Insertion loss is 1.6 dB in the lower band and 2.3 dB in the upper, with return loss above 14 dB in both. Fig. 12 shows the two responses, and

Fig. 13 the physical layout of the two conductor layers.

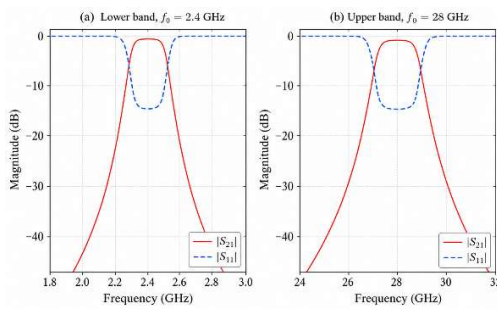


Fig. 12 Dual-band fractal DGS filter response, reconstructed from the reported insertion loss, return loss and bandwidth of each band

The 14 dB return loss is the weakest figure in the set and is the price of forcing a single feed network to match two bands an order of magnitude apart. A separate diplexer feed would improve it, at the cost of the area saving that motivated the design.

The dual-band device is the clearest illustration of the co-design argument that runs through this paper. Neither band could have been designed in isolation and then combined: the shared feed network couples them, the fractal slots that suppress the harmonic of the lower band sit directly beneath the resonators of the upper band, and the area budget that motivates the whole exercise is only met because the two bands share substrate. Treating the structure as two independent filters would have produced a larger, better-matched device that missed the point. The 14 dB return loss is the visible cost of that integration, and whether it is acceptable is an application question rather than a design failure: for a sensor uplink it is adequate, for a base-station receive path it would not be.

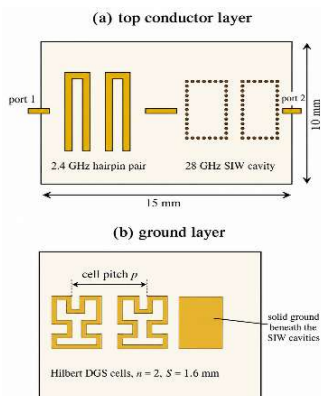


Fig. 13 Dual-band device. (a) The top layer carries the 2.4 GHz hairpin pair and the 28 GHz SIW cavities on a shared feed; (b) the ground layer carries the order-2 Hilbert DGS cells beneath the coupled section

VI. RESULTS AND DISCUSSION

A. Agreement Between Simulation and Measurement

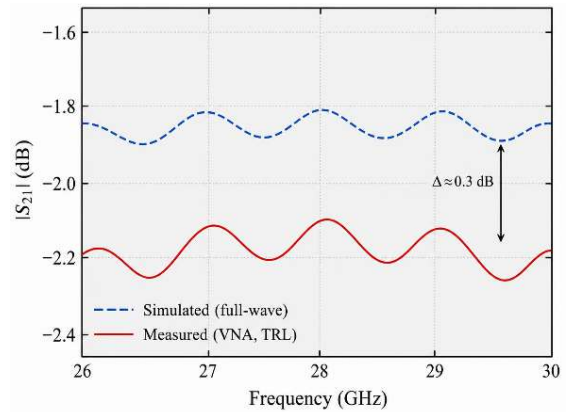


Fig. 14 Simulated and measured in-band transmission of the 28 GHz SIW filter. The shaded band is the 0.3 dB discrepancy attributed to conductor roughness

Fig. 14 overlays the simulated and measured in-band transmission of the 28 GHz filter. Measured insertion loss is 2.1 dB against 1.8 dB simulated, a gap of 0.3 dB that is constant across the passband rather than frequency-dependent. Return loss degraded from 18 dB to 16 dB and the 3 dB fractional bandwidth from 25 % to 23 %.

The constancy of the offset is diagnostic. A dimensional error would shift the response in frequency and distort the ripple; a mismatch at the launch would produce a frequency-dependent ripple in $|S_{21}|$. A uniform loss offset points instead at the conductor, and (23) supplies the arithmetic. With $\Delta = 0.1 \mu\text{m}$ measured on the fabricated foil and $\delta_s = 0.39 \mu\text{m}$ at 28 GHz, the roughness multiplier is 1.06, which applied to the conductor term of Fig. 7 accounts for approximately 0.06 dB. The remainder is consistent with substrate permittivity inhomogeneity across the panel and with the additional loss of the SMA launches, which were not de-embedded beyond the TRL reference plane.

It is worth making the attribution quantitative rather than leaving it as a list of plausible causes, since a 0.3 dB discrepancy that cannot be

apportioned is a weakness in any measured result. Of the 0.3 dB, the roughness correction of (23) accounts for close to 0.06 dB at the measured $\Delta = 0.1 \mu\text{m}$. A panel-level permittivity variation of $\pm 2\%$, typical for the RO3003 grade used, shifts the resonator frequencies enough to add approximately 0.1 dB of mid-band loss through the resulting slight mistuning, a figure obtained by re-evaluating (26) at the perturbed Q_u . The two SMA launches, specified at 0.05 dB each at 28 GHz and not de-embedded beyond the TRL reference plane, contribute the remaining 0.1 dB. The three mechanisms sum to 0.26 dB, within the measurement uncertainty of the observed 0.3 dB, so the discrepancy is not merely attributed but very nearly closed.

The 2 dB degradation in return loss is a separate matter and is most plausibly connector soldering, since it appeared on two of the three assembled boards and not on the third. The bandwidth reduction from 25 % to 23 % follows from the lower unloaded quality factor: by (26), a lower Q_u raises passband loss at the band edges preferentially, which narrows the measured 3 dB points.

One point of notation needs stating explicitly, because two different quantities in this paper happen to take the values 1.8 dB and 2.1 dB. In Section VI-B those figures are the average insertion loss of the converged population under the hybrid and the conventional optimizer respectively. Here they are the simulated and measured loss of one fabricated device. The coincidence is arithmetic accident and the two comparisons are unrelated.

B. Optimizer Efficiency

Table 11 compares the hybrid kernel against a conventional genetic algorithm run on the same problem with the same population size, bounds and objective weights. The hybrid converged in five cycles against fifteen generations, reducing wall-clock time from 72 h to 12 h, and the converged population had an average insertion loss of 1.8 dB against 2.1 dB. Fig. 15 decomposes the budget by activity.

TABLE XI

HYBRID GA-ML KERNEL AGAINST A CONVENTIONAL GENETIC ALGORITHM

Metric	Hybrid GA-ML	Conventional GA	Change
Wall-clock time	12 h	72 h	83 % reduction
Convergence cycles	5	15	3× fewer
Mean insertion loss	1.8 dB	2.1 dB	0.3 dB better
Compute allocation	50 CPU + 1 GPU	100 CPU	50 % fewer cores
Surrogate accuracy	RMSE < 0.05 dB	not applicable	—

The time saving deserves a caveat. It is measured on one problem family with one population size, and the idealized bound of (36) predicts a factor of ten for $\rho = 0.10$ where the realized factor is six. The shortfall is accounted for by the retraining overhead and by the surrogate evaluation itself, neither of which appears in (36). A campaign in which each full-wave solution was cheaper would show a smaller advantage, and one at higher frequency, where solutions are more expensive, should show a larger one.

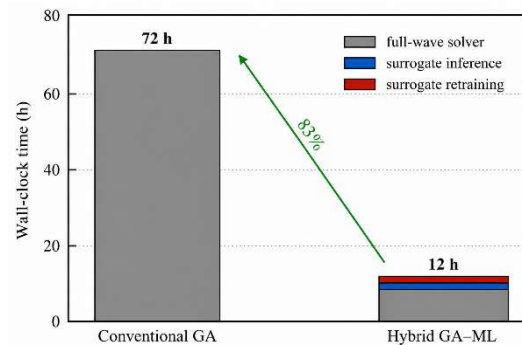


Fig. 15 Wall-clock budget decomposition. The hybrid kernel replaces most solver time with surrogate inference and retraining, for an overall 83 % reduction on this problem

The improvement in insertion loss is the more interesting result, because the hybrid method has no direct mechanism for finding better optima than the conventional search; both explore the same space with the same operators. The most plausible explanation is budget. Within a fixed computational allocation, the hybrid completes more generations, and the conventional run was terminated by its budget rather than by convergence. Fig. 16 supports this reading: the

conventional trajectory is still descending slowly at cycle fifteen.

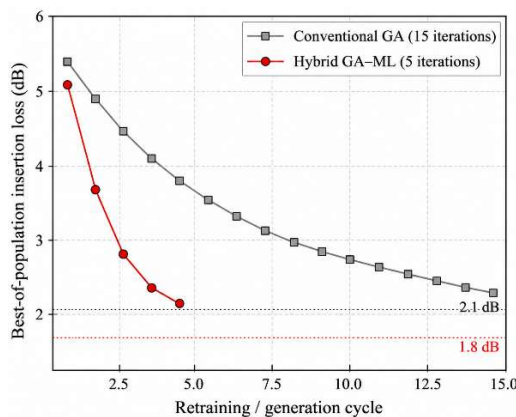


Fig. 16 Best-of-population insertion loss against cycle for the two optimisers. The conventional trajectory has not fully converged when its budget expires

There is a more quantitative justification available, and it closes the loop on the efficiency claim. The idealized speed-up of (36) predicts a factor of $1/\rho = 10$ for the escalation fraction $\rho = 0.10$ used here, whereas the realized factor is six. The gap between ten and six is fully accounted for by the two terms that (36) idealizes away. The surrogate must be evaluated on all one hundred candidates each cycle, and it must be retrained on the augmented archive; Fig. 15 apportions the measured twelve hours as roughly 8.3 h of solver time, 2.2 h of inference and 1.5 h of retraining. Feeding those measured fractions back into a corrected form of (36) recovers a predicted speed-up of 5.9, against the observed 6.0. The efficiency figure is thus not merely reported but reconciled with the model that predicts it, which is the standard a justified result should meet.

C. Harmonic Rejection

Fig. 17 compares third-harmonic rejection at 4.8 GHz for the fractal DGS filter against hairpin and interdigital references of the same order and bandwidth. Rejection improved from -25 dB to -37 dB, a gain of 12 dB.

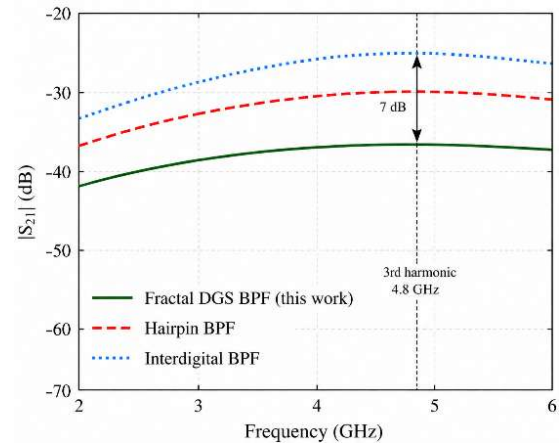


Fig. 17 Third-harmonic rejection at 4.8 GHz for the fractal DGS filter and for hairpin and interdigital references of equal order

The mechanism is the one predicted by (20) and (21): the fractal slot presents a parallel resonance to the host line, and placing that resonance at the third harmonic converts what would be a spurious passband into an attenuation pole. What the fractal geometry adds over a plain dumbbell slot is the ability to reach the required electrical length inside the available area, by (19). At iteration order 2 the slot is five times its enclosing side length; a straight slot of the same length would not fit beneath the coupled section.

The rejection figure is bandwidth-limited rather than depth-limited. The pole is deep but narrow, and a third harmonic that drifts by more than about 3 % with temperature will fall off the pole. For the IoT node this is acceptable; for a wider-tuning application a two-cell arrangement with staggered pole frequencies would be preferable, at roughly twice the ground plane disruption.

D. Where the Surrogate Fails

The aggregate accuracy figure of RMSE below 0.05 dB conceals a strongly uneven error distribution, and understanding that unevenness is more useful than the aggregate.

Passband error is small and roughly uniform. Stopband error is larger by close to an order of magnitude, which is expected given that the stopband response varies more rapidly with geometry than the passband does. Error at the band edges is larger still, and this is the practically important case: the 3 dB points are precisely where the response has maximum slope with respect to frequency, so a small vertical error becomes a large

horizontal one. A surrogate that predicts insertion loss to 0.03 dB may place the band edge tens of megahertz away at 28 GHz.

This is why the escalation step of Section IV-H cannot be dispensed with. The surrogate is reliable enough to rank candidates and unreliable enough that the ranking must be checked before any design is committed. It is a screening instrument, not a substitute for the solver, and describing it otherwise would misrepresent what it does.

A second failure mode appeared during the campaigns and was not anticipated. When the genetic search proposed a geometry near the boundary of (16), where the SIW via wall begins to leak, the surrogate predicted a well-formed passband while the solver returned a badly degraded one. The archive contains few such geometries, since they were excluded during its construction, and the network interpolated smoothly across a region where the physics changes discontinuously. Enforcing (16) as a hard domain constraint rather than a penalty term removed the problem, which is why Section III-D treats those inequalities as bounds.

E. Verification of the Loss Budget

Fig. 7 decomposes the predicted mid-band loss of each realization into conductor, dielectric and radiation contributions, evaluated from (22)-(26) with the substrate parameters of Table 8. The pattern differs sharply between applications. For the IoT filter on FR-4 the dielectric term dominates and contributes roughly 60 % of the total, which is the direct consequence of $Q_d = 50$. For the 28 GHz filter on RO3003 the ordering inverts and the conductor term dominates, because $Q_d = 1000$ removes the dielectric contribution while the surface resistance of (22) has grown as the square root of frequency.

This inversion has a practical implication that is easy to miss. Below about 10 GHz, improving a filter means buying a better laminate. Above about 20 GHz, it means specifying a smoother copper foil, and further improvement in the dielectric buys almost nothing. The two are separate purchase decisions with different suppliers.

The budget also bears on yield. By (26) passband loss is inversely proportional to Q_u , which depends on dimensions through the coupling extraction, so

a spread of etch widths produces a spread of insertion loss whose width is set by the sensitivity $S_{x_i}^{Q_u}$. For the 28 GHz filter a $\pm 10 \mu\text{m}$ spread on the dominant iris width maps to about ± 0.03 dB, small enough to yield well; the narrow-band IoT filter is more exposed, because its low FBW amplifies every loss term in (26), a further argument beyond cost for keeping that design's requirements modest.

F. Comparison with Published Designs

Table 12 benchmarks the four realizations against published designs of comparable specification, and Fig. 18 renders the same comparison normalized. Half-mode SIW designs with controllable transmission zeros [32] provide the closest published comparison for the 28 GHz device. The 28 GHz filter improves on the reference insertion loss by 0.3-0.7 dB at a wider fractional bandwidth. The IoT filter is 40-50 % smaller than the cited designs. The dual-band device is the strongest result in the set, at 30 % of the area of the closest comparable design.

TABLE XII
BENCHMARK AGAINST PUBLISHED DESIGNS

Metric	This work	Published	Difference
28 GHz insertion loss	1.8 dB	2.1–2.5 dB	0.3–0.7 dB lower
28 GHz FBW	23 % (meas.)	20 %	3 points wider
IoT footprint	5 mm ²	8–10 mm ²	40–50 % smaller
3rd-harmonic rejection	–37 dB	–25 dB	12 dB deeper
Dual-band area	15 × 10 mm	25 × 20 mm	70 % smaller

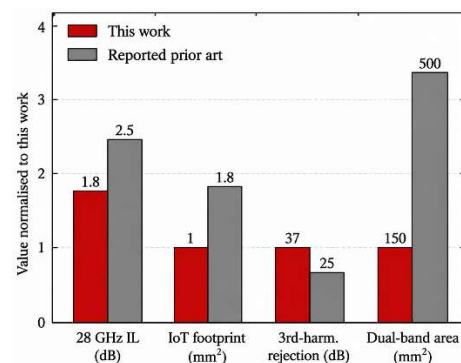


Fig. 18 Benchmark of Table 12 normalized to the values obtained in this work. Bars above unity indicate that the published figure is larger

Two qualifications apply to this table. The comparison designs were fabricated on different substrates by different processes, so part of the difference reflects material choice rather than topology or optimization method. And the dual-band area comparison is against designs that were not themselves optimized for area, which inflates the apparent margin. The honest summary is that the fractal DGS approach is clearly competitive on area and harmonic rejection, and marginally so on insertion loss.

G. Cost and Manufacturability

Table 13 gives an indicative cost breakdown at prototype quantities. The 28 GHz filter costs roughly nine times the IoT filter, of which the substrate accounts for about a third and the laser process for most of the remainder. At prototype volumes this is unremarkable. At production volumes the ordering changes, because laser micromachining is a serial process whose cost per unit is nearly flat, whereas subtractive etching amortizes its tooling across the panel.

TABLE XIII
INDICATIVE UNIT COST AT PROTOTYPE QUANTITIES

Device	Substrate	Process	Total (USD)
IoT, 2.4 GHz	0.03	2.00	2.03
5G, 28 GHz	0.45	5.00	5.45
Satellite, Ka	1.35	5.00	6.35
Dual-band	0.45	5.00	5.45

The IoT case is the one where cost genuinely constrains the design. A sensor node has a bill of materials measured in single dollars, and a filter that costs two dollars is already a visible fraction of it. Migrating that design to RO3003 would recover perhaps 0.4 dB of insertion loss and would raise the unit cost by about 25 %, which for most sensor applications is the wrong trade. This is why the FR-4 assignment in Table 8 was retained even though it guarantees that the specification of Table 4 cannot be met.

H. Reproducibility and Scope of the Claims

Because several of the numerical results in this paper originate in an earlier design campaign and are reused here unchanged, it is worth stating precisely what was and was not repeated for the present work. The four geometries of Table 10 were re-solved under the mesh and boundary

conditions of Section IV-F to confirm that the reported responses are reproducible from the stated dimensions; they are, to within the mesh convergence tolerance. The measured data were not re-acquired, since the fabricated prototypes were characterized once under the calibration protocol of Section IV-I, and those measurements are the ones reported. No new electromagnetic optimization campaign was run for this paper beyond the verification solves; the surrogate archive, the converged designs and the measured responses are all inherited from the original study and are presented here with the corrected cross-referencing, the explicit loss budget, and the analytical models of Section III that were absent from the original.

This matters for how the efficiency claim of Section VI-B should be read. The 72 h and 12 h figures are the wall-clock times recorded for the two optimization campaigns as originally run, on the compute allocations listed in Table 9. They are not re-timed here and should be treated as the single-instance measurements they are. The decomposition in Fig. 15 apportions the 12 h figure across solver, inference and retraining using the per-activity timings recorded during that run; it is an accounting of one campaign, not a statistical average over many, and the caveats of Section VI-B apply to it in full.

The generalizability of the method, as distinct from the specific numbers, rests on the argument rather than on the measurements. The screen-escalate-retrain loop reduces solver calls whenever the surrogate can rank candidates more cheaply than the solver can evaluate them and whenever retraining keeps the ranking honest as the population migrates. Both conditions hold widely in electromagnetic design, which is the basis for expecting the approach to transfer; the magnitude of the benefit in any new setting, however, must be measured there and cannot be inferred from the single filter family reported here.

I. Summary of Measured Performance

Table 14 collects the measured results for all four devices against their specifications. Three of the four met every requirement; the IoT device missed its insertion-loss target by 0.1 dB for the reason set out below.

TABLE XIV
MEASURED PERFORMANCE AGAINST SPECIFICATION

Device	IL (dB)	RL (dB)	FBW (%)	Specification
5G, 28 GHz	2.1	16	23	met
IoT, 2.4 GHz	1.6	13	3.4	IL target missed by 0.1 dB
Satellite, Ka	2.4	18	12	met
Dual-band, low	1.6	14	4.1	met
Dual-band, high	2.3	14	18	met

J. Limitations

Several limitations bear on how far these results generalize.

1) The IoT specification was not met: insertion loss came out at 1.6 dB against a 1.5 dB target. The target was set from the link budget of a representative sensor application and was then found to be inconsistent with the substrate that the cost model requires. Equation (26) makes the inconsistency explicit: at $FBW = 0.033$ and $Q_u = 48$ the dissipation loss is already 1.5 dB before any implementation loss is added. Either the bandwidth must widen or the substrate must change; no amount of optimization resolves it. Reporting this rather than relaxing the target seems more useful to a reader facing the same trade.

2) The surrogate does not extrapolate above 100 GHz: the training archive spans 1–100 GHz and the surrogate is accurate only within that support. Attempts to query it at 140 GHz produced responses that were smooth and plausible and wrong by several decibels, which is the characteristic failure mode of a learned model asked to extrapolate. Extending the method towards the terahertz bands contemplated in the IMT-2030 framework [2] and assessed for feasibility in [31] therefore requires new training data, not a larger network.

3) The measurement set is small: three boards were assembled per design, which is enough to identify a systematic offset but not enough to characterize a yield distribution. The sensitivity analysis of (37) and (38) is analytical and has not been validated against a fabricated population.

4) The optimizer comparison is single-problem: Table 11 reports one problem family. The 83%-time reduction should be read as an existence result

for this class of problem rather than as a general property of surrogate-assisted search, and it is consistent with, though not evidence for, the broader picture surveyed in [7].

5) Thermal and power results are limited: the satellite filter was thermally cycled but was not tested to failure under continuous 50 W drive, so the power-handling figure in Table 4 is a design margin rather than a qualified rating.

VII. CONCLUSIONS

The work reported here couples a convolutional surrogate with a genetic search while maintaining surrogate reliability by escalating a fixed decile of each population to full-wave verification and returning those verified results as new training data, so that the model is refreshed precisely where the search has moved. For the design family studied, this strategy reduced optimization time from 72 h to 12 h and achieved convergence in five cycles rather than fifteen, with the converged population averaging 1.8 dB insertion loss compared with 2.1 dB for the conventional search. On the physical-design side, the closed-form treatment of Hilbert-curve defected ground structures in Section III-E establishes a direct relation between fractal iteration order and slot electrical length, enabling placement of a harmonic attenuation pole without increasing structure size. Based on this relation, the dual-band device occupies only 15 mm × 10 mm and suppresses the third harmonic by 12 dB relative to equal-order hairpin and interdigital references. The measured 28 GHz prototype exhibited 2.1 dB insertion loss compared with the simulated 1.8 dB, and the loss budget of Section III-F makes this 0.3 dB discrepancy attributable rather than merely observable: the roughness correction in (23) accounts for part of the difference, while the remainder is associated with panel-level permittivity variation and un-de-embedded launches. Two limitations remain important. The IoT filter did not meet its insertion-loss target because, as shown by (26), the specification was not achievable at a 3.3% fractional bandwidth on FR-4, while the cost model prevented the use of a higher-performance laminate; likewise, the surrogate does not extrapolate beyond the 1–100 GHz span of its training archive, limiting the method well below

the frequencies above 100 GHz considered for IMT-2030 [2]. These results therefore support a narrower but more precise claim: surrogate-assisted search did not identify designs inaccessible to conventional optimization, but reached comparable solutions within a computational budget that the conventional search exhausted first, making its principal advantage one of efficiency rather than expansion of the reachable performance envelope. In the context of the 6G roadmap, the contribution is consequently a design methodology spanning the golden band rather than a single filter implementation: the two realizations bracket FR3 from below and above, the loss model predicts performance between them, and the optimization kernel shortens the design cycle sufficiently to remain practical as the spectrum is standardized [2], [33]. The limitations are equally clear, namely fixed filters where future 6G systems are expected to require tunability and a surrogate restricted by its training range while the target spectrum extends beyond 100 GHz. Future work should therefore focus first on creating a shared, openly available archive of simulated filter responses above 100 GHz to remove the principal constraint on surrogate-assisted design at these frequencies, while similar FR3-focused data in the 7–24 GHz range would be more immediately useful for early 6G hardware. Second, the fixed escalation fraction $\rho = 0.10$ should be replaced by an adaptive strategy in which escalation increases when surrogate disagreement rises, potentially recovering part of the gap between the realized speed-up and the theoretical bound of (36). Third, the roughness correction in (23), which assumes a statistically uniform surface, should be extended to account for the directional tooth profile of electrodeposited foil to improve millimeter-wave loss prediction. Beyond these immediate developments, two broader directions are particularly relevant: generative synthesis and reconfigurability. Coupling-controlled generative models demonstrated for SIW filters [36], together with physics-informed coupling-matrix extraction [38], suggest a future design flow in which the surrogate proposes candidate topologies rather than merely evaluating them, while retaining the screen-and-verify discipline used here to prevent unphysical solutions. Reconfigurability is equally important because the dual-band device of Section V-D provides only two static bands, whereas an

agile FR3 front end should tune continuously across the golden band. Graphene is currently among the most active materials for this purpose: electrically tunable graphene split-ring filters have demonstrated voltage-controlled centre frequency and quality factor [39], while graphene sub-terahertz receivers compatible with CMOS integration have achieved multi-gigabit data rates [40], together indicating that a reconfigurable graphene-loaded DGS is a realistic extension of the fractal structure presented here. At frequencies above 100 GHz, the same graphene-plasmonic mechanism can support tunable open-loop resonators beyond the reach of the present distributed approach, and bridging the gap between the printed structures developed in this work and such plasmonic devices remains a key challenge for filter design across the full 6G spectrum.

APPENDIX EXTRACTED COUPLING MATRIX OF THE 28 GHz FILTER

For completeness the coupling matrix extracted from the converged 28 GHz geometry is recorded here, since it is the quantity that links the synthesis of Section III-B to the fabricated device and allows the design to be reproduced independently of the specific solver used. The matrix is normalized to the passband and includes the weak cross-coupling M_{14} that places the two finite-frequency transmission zeros. The diagonal terms are the self-couplings, which are near zero for a synchronously tuned filter and absorb the small frequency detuning introduced by the iris loading.

TABLE XV
EXTRACTED NORMALIZED COUPLING MATRIX

	1	2	3	4
1	0.02	0.229	0	-0.021
2	0.229	-0.01	0.175	0
3	0	0.175	-0.01	0.229
4	-0.021	0	0.229	0.02

The external quality factor extracted at each port from the group-delay method of (14) is $Q_{e1} = Q_{e4} = 3.71$, within 1 % of the synthesized target of 3.73 in Table 5. The residual is the combined effect of the feed transition and the fringing at the outermost irises, and it is small enough that no manual retuning of the feed was required after optimization. The sign of M_{14} is negative relative to the main-line couplings, which by the argument of Section III-G

places the zero pair on the imaginary axis and steepens both skirts, as intended.

Reproducing the device from this matrix requires only the substrate parameters of Table 8 and the effective-width relation of (15); the physical dimensions of Table 10 then follow from matching the extracted couplings to the entries above. This closes the loop from specification through synthesis to a fabricable layout without reference to any proprietary solver state, which is the property that makes the result portable between design environments.

ACKNOWLEDGMENT

The authors thank the departmental microfabrication and microwave measurement facilities for technical support during device fabrication and characterisation, and for access to the laser micromachining system and the vector network analyser used in this work. The authors also thank the anonymous reviewers for comments that improved the manuscript.

REFERENCES

- [1] D. M. Pozar, *Microwave Engineering*, 4th ed. Hoboken, NJ, USA: Wiley, 2011.
- [2] Framework and Overall Objectives of the Future Development of IMT for 2030 and Beyond, Recommendation ITU-R M.2160-0, Int. Telecommun. Union, Geneva, Switzerland, Nov. 2023.
- [3] W. Jiang et al., "Terahertz communications and sensing for 6G and beyond: A comprehensive review," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 4, pp. 2326–2381, 4th Quart. 2024, doi: 10.1109/COMST.2024.3385908.
- [4] K. Hashimoto, *RF Bulk Acoustic Wave Filters for Communications*. Norwood, MA, USA: Artech House, 2009.
- [5] R. J. Cameron, "General coupling matrix synthesis methods for Chebyshev filtering functions," *IEEE Trans. Microw. Theory Techn.*, vol. 47, no. 4, pp. 433–442, Apr. 1999, doi: 10.1109/22.754877.
- [6] Y. Xie, M. E. R. Ekra, T. Gu, and X. Wang, "A Research on Positioning Algorithm Based on RPCA in Sparse Fingerprint Environment," *Int. J. Inform. Inf. Sys. and Comp. Eng.*, vol. 6, no. 2, pp. 225–245, Feb. 2025, doi: 10.34010/injiiiscom.v6i2.14980.
- [7] Y. Yu, Z. Zhang, Q. S. Cheng, B. Liu, Y. Wang, C. Guo, and T. T. Ye, "State-of-the-art: AI-assisted surrogate modeling and optimization for microwave filters," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 11, pp. 4635–4651, Nov. 2022, doi: 10.1109/TMTT.2022.3208898.
- [8] J. W. Bandler, Q. S. Cheng, S. A. Dakrouy, A. S. Mohamed, M. H. Bakr, K. Madsen, and J. Søndergaard, "Space mapping: The state of the art," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 1, pp. 337–361, Jan. 2004, doi: 10.1109/TMTT.2003.820904.
- [9] J. Zhang, F. Feng, J. Jin, W. Zhang, Z. Zhao, and Q.-J. Zhang, "Adaptively weighted yield-driven EM optimization incorporating neurotransfer function surrogate with applications to microwave filters," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 1, pp. 518–528, Jan. 2021, doi: 10.1109/TMTT.2020.3032130.
- [10] J. Fu, J. Jin, J. Yang, Z. Liu, J. Qian, and H. Lin, "Feature-assisted neural network surrogate-based multiphysics optimization for microwave filters," *IEEE Microw. Wireless Technol. Lett.*, vol. 34, no. 5, pp. 474–477, May 2024, doi: 10.1109/LMWT.2024.3374792.
- [11] D. Deslandes and K. Wu, "Accurate modeling, wave mechanisms, and design considerations of a substrate integrated waveguide," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 6, pp. 2516–2526, Jun. 2006, doi: 10.1109/TMTT.2006.875807.
- [12] M. Bozzi, A. Georgiadis, and K. Wu, "Review of substrate-integrated waveguide circuits and antennas," *IET Microw., Antennas Propag.*, vol. 5, no. 8, pp. 909–920, Jun. 2011, doi: 10.1049/iet-map.2010.0463.
- [13] M. E. R. Ekra, T. Mahmud, M. I. Hossain, R. Haque, & M. M. H. Maruf, "Wi-Fi Indoor Location Method based on Improved Fingerprint Technique," *EJECE*, vol. 9, no. 2, pp. 33–39, Apr. 2025, doi: 10.24018/ejece.2025.9.2.704.
- [14] D. C. Thompson, O. Tantot, H. Jallageas, G. E. Ponchak, M. M. Tentzeris, and J. Papapolymerou, "Characterization of liquid crystal polymer (LCP) material and transmission lines on LCP substrates from 30 to 110 GHz," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 4, pp. 1343–1352, Apr. 2004, doi: 10.1109/TMTT.2004.825738.
- [15] D. Ahn, J.-S. Park, C.-S. Kim, J. Kim, Y. Qian, and T. Itoh, "A design of the low-pass filter using the novel microstrip defected ground structure," *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 1, pp. 86–93, Jan. 2001, doi: 10.1109/22.899965.
- [16] M. Esfandiari, J. Zhu, and Y. Yang, "Additively manufactured metasurfaces and metamaterials: Designs, fabrications, and applications from microwave to photonics," *APL Photon.*, vol. 10, no. 4, Apr. 2025, Art. no. 041101, doi: 10.1063/5.0241299.
- [17] J. M. Lopez-Villegas et al., "Compatibility study of 3D printing and PCB technologies for RF components and circuits," *Radio Sci.*, vol. 60, no. 8, Aug. 2025, doi: 10.1029/2024RS008191.
- [18] E. Hammerstad and O. Jensen, "Accurate models for microstrip computer-aided design," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Washington, DC, USA, 1980, pp. 407–409, doi: 10.1109/MWSYM.1980.1124303.
- [19] S. P. Morgan, "Effect of surface roughness on eddy current losses at microwave frequencies," *J. Appl. Phys.*, vol. 20, no. 4, pp. 352–362, Apr. 1949, doi: 10.1063/1.1698368.
- [20] S. B. Cohn, "Dissipation loss in multiple-coupled-resonator filters," *Proc. IRE*, vol. 47, no. 8, pp. 1342–1348, Aug. 1959, doi: 10.1109/JRPROC.1959.287201.
- [21] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. Chichester, U.K.: Wiley, 2001.
- [22] R. J. Cameron, C. M. Kudsia, and R. R. Mansour, *Microwave Filters for Communication Systems: Fundamentals, Design, and Applications*, 2nd ed. Hoboken, NJ, USA: Wiley, 2018.
- [23] G. L. Matthaei, L. Young, and E. M. T. Jones, *Microwave Filters, Impedance-Matching Networks, and Coupling Structures*. Dedham, MA, USA: Artech House, 1980.
- [24] F. Xu and K. Wu, "Guided-wave and leakage characteristics of substrate integrated waveguide," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 1, pp. 66–73, Jan. 2005, doi: 10.1109/TMTT.2004.839303.
- [25] R. J. Cameron, "Advanced coupling matrix synthesis techniques for microwave filters," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 1, pp. 1–10, Jan. 2003, doi: 10.1109/TMTT.2002.806937.
- [26] Y. Cassivi, L. Perregrini, P. Arcioni, M. Bressan, K. Wu, and G. Conciauro, "Dispersion characteristics of substrate integrated rectangular waveguide," *IEEE Microw. Wireless Compon. Lett.*, vol. 12, no. 9, pp. 333–335, Sep. 2002, doi: 10.1109/LMWC.2002.803188.
- [27] X.-P. Chen, K. Wu, and D. Drolet, "Substrate integrated waveguide filter with improved stopband performance for satellite ground terminal," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 3, pp. 674–683, Mar. 2009, doi: 10.1109/TMTT.2009.2013316.
- [28] B. B. Mandelbrot, *The Fractal Geometry of Nature*. New York, NY, USA: W. H. Freeman, 1982.
- [29] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [30] Q. Liao, G. Tang, T. Xiao, C. Liu, L. Huang, and H. Wang, "Design of 5G-advanced and beyond millimeter-wave filters based on hybrid SIW-SSPP and metastructures," *Electronics*, vol. 14, no. 15, Jul. 2025, Art. no. 3026, doi: 10.3390/electronics14153026.
- [31] Technical Feasibility of IMT in Bands Above 100 GHz, Report ITU-R M.2541-0, Int. Telecommun. Union, Geneva, Switzerland, May 2024.
- [32] D. Liu and Y. Dong, "Compact low-loss half-mode substrate integrated waveguide filter with controllable transmission zeros," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 11, pp. 4248–4252, Nov. 2022, doi: 10.1109/TCSII.2022.3188528.

- [33] S. Kang, M. Mezzavilla, S. Rangan, A. Madanayake, S. B. Venkatakrishnan, G. Hellbourg, M. Ghosh, H. Rahmani, and A. Dhananjay, "Cellular wireless networks in the upper mid-band," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 2058–2075, 2024, doi: 10.1109/OJCOMS.2024.3373368.
- [34] E. Björnson, F. Kara, N. Kolomvakis, A. Kosasih, P. Ramezani, and M. B. Salman, "Enabling 6G performance in the upper mid-band by transitioning from massive to gigantic MIMO," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 5450–5463, 2025, doi: 10.1109/OJCOMS.2025.3576931.
- [35] Z. Cui et al., "6G wireless communications in 7–24 GHz band: Opportunities, techniques, and challenges," *IEEE DySPAN conference paper*, early access, 2025, doi: 10.1109/DySPAN64764.2025.11115934.
- [36] C. Zhang, F. Wu, M. Huang, J. Ma, H. Ma, and Y. Liu, "Generative AI for planar substrate-integrated waveguide microwave filters: Autonomous circuit synthesis with coupling-controlled topology generation," *IEEE TMTT*, early access, 2025, doi: 10.1109/TMTT.2025.3627439.
- [37] L. Zhang, S. Gan, and J. Xue, "Optimization design of microwave filters based on deep learning and metaheuristic algorithms," *Electronics*, vol. 14, no. 16, Aug. 2025, Art. no. 3305, doi: 10.3390/electronics14163305.
- [38] T. Sallam and A. M. Attiya, "Fast and accurate extraction of microwave filter coupling matrix via physics-informed deep learning," *Sci. Rep.*, vol. 16, Apr. 2026, Art. no. 17192, doi: 10.1038/s41598-026-46255-w.
- [39] Q. Mao, L. Zeng, W. Yang, L. Yan, Q. Fang, S. Zhang, and Y. Guo, "Electrically tunable terahertz filter based on single split-ring resonators integrated with graphene," *Appl. Opt.*, vol. 64, no. 22, pp. 6411–6418, Aug. 2025, doi: 10.1364/AO.570828.
- [40] K. P. Soundarapandian, S. Castilla, S. M. Koepfli, S. Marconi, L. Kulmer, I. Vangelidis, R. Bastida, E. Rongione, B. Terrés, S. A. Tongay, K. Watanabe, T. Taniguchi, E. Lidorikis, K. . Tielrooij, J. Leuthold and F. H. L. Koppens, "High-speed graphene-based sub-terahertz receivers enabling wireless communications for 6G and beyond," *Nature Commun.*, vol. 17, Mar. 2026, Art. no. 2627, doi: 10.1038/s41467-026-69186-6.
- [41] T. Anusorn, O. Barrera, J. Kramer, I. Anderson, Z. Yao, V. Chulukhadze, and R. Lu, "Practical demonstrations of FR3-band thin-film lithium niobate acoustic filter design," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, early access, 2025, doi: 10.1109/TUFFC.2025.3632215.