Phantom Optima in RF Power Amplifier Optimization:

A Fitness-Function Ablation Study on a Cascode CMOS PA and a 1:2 Asymmetric Doherty at 3.55 GHz

Yao Chang

chang212@gmail.com

May 2026

Abstract

We introduce a five-rung fitness-function ablation (F1–F5) for RF power amplifier design optimization and quantify, against a strict harmonic-balance reference oracle, how much each layer of physics fidelity contributes to the rate at which the optimizer converges to ***phantom optima*** *— design points that score well on the fitness function but disagree with the reference oracle on saturation power, distortion, or convergence. We apply the methodology to a cascode CMOS PA model and a 1:2 asymmetric Doherty model at 3.55 GHz, with the latter parameter-tuned to match the back-off PAE numbers reported by Camarchia et al. [1]. With N = 20 random seeds and an A*-with-restart budget of MAX_ITER = 200 per seed, the Doherty F5 fitness (convergence and thermal hard-zero gates layered over scored distortion) reduces the phantom rate from 20 % to 10 %, a 10 percentage-point improvement attributable solely to the gate. The much larger transition is F1 → F2 (closed-form Raab proxy to single-tone harmonic balance) at 65 pp, while F2 → F3 → F4 contribute only ±5 pp each. Direct topology probes confirm that load modulation operates correctly (Z_main transitions from 22 Ω at back-off to 13 Ω at saturation, in line with the textbook 2·Rload → Rload target) but no parameter combination in the explored space produces a true Doherty back-off PAE peak — attributable to inherent 28 nm CMOS class-AB efficiency limitations rather than implementation error.*

Index Terms — *Power amplifier, Doherty, harmonic balance, design automation, A* search, phantom optimum, CMOS RF.*

1. Introduction

Modern RF power amplifier (PA) design automation routinely couples a harmonic-balance (HB) circuit simulator to a numerical optimizer that explores a parameter space of device widths, bias voltages, matching-network reactances, and harmonic terminations. The fitness function fed to the optimizer is necessarily an approximation: full electro-thermal HB at every candidate point is too expensive for realistic search budgets, so practitioners substitute closed-form proxies (Raab-style class-F or class-B Psat formulae [4], Cripps load-line analyses [3]) or run HB with fewer harmonics, loose convergence tolerances, or skipped distortion scoring [5], [8]. Each shortcut moves the fitness landscape away from the underlying physics. The optimizer happily descends into the resulting fictitious basins and reports a converged design that, when verified against a strict-tolerance reference, fails some acceptance criterion: Psat off by more than 1 dB, HD3 off by more than 5 dBc, thermal runaway, or sub-driver-spec output power. We call these reported optima *phantoms* — they are real optima of the proxy fitness, but they are not real designs.

This paper poses a question that, despite decades of PA-CAD literature, we have not seen quantified for a representative CMOS Doherty: *how much physics does the fitness function need before the phantom rate drops below an interesting threshold?* We define a five-rung ladder (F1–F5) of fitness functions from "closed-form Raab" to "full multi-tone HB with convergence and thermal hard-zero gates," run an A*-with-restart search [6], [7] under each rung against 20 random starting points on both the cascode and the Doherty, and measure the phantom rate against a strict-tolerance HB reference oracle. The contribution is methodological rather than novel-physics: this is, to our knowledge, the first published ablation that isolates the marginal value of each fidelity layer with respect to phantom rate on the same benchmark.

2. Related Work

The harmonic-balance solver class we adopt is a continuation-Newton variant of the standard frequency-domain formulation due to Rizzoli, Mastri, and Masotti [8], with the textbook implementation details summarized by Maas [10] and Kundert et al. [5]. Our reference oracle uses N = 5 harmonics, 12-step amplitude continuation, 150 Newton iterations per ramp step, and a two-tier convergence ladder (strict: relF < 10^{-7} AND |F|_KCL < 10^{-10}; accept: relF < 10^{-4} OR |F|_KCL < 10^{-7}).

The Doherty amplifier dates to Doherty's 1936 paper [2]. Modern CMOS asymmetric variants and their efficiency-vs-back-off trade-offs are reviewed by Camarchia et al. [1], which provides the acceptance numbers our model is tuned to match (Psat 26–28 dBm, PAE at peak 40–45 %, PAE at 6 dB back-off 35–40 %, HD3 at peak −22 to −28 dBc).

A* search [6] is not the typical PA-CAD optimizer (Nelder–Mead simplex [9], gradient descent, and global metaheuristics dominate the literature), but the discrete-step neighbor expansion is well suited to the bounded mixed-discrete spaces RF PA design produces (device widths in foundry quanta, decoupling-cap PDK steps, finger-count integers). We adopt A* with kick-step restart after 30-iter plateau and a cap of 6 restarts per seed. The fitness-function ablation methodology we propose is, however, optimizer-agnostic: any solver that produces a "reported optimum" can be phantom-classified the same way.

3. The F1–F5 Fitness Ladder

We define five fitness variants ordered by physics fidelity:

F1 — Closed-form Raab Psat proxy. No HB. Psat estimated from Raab-style class-AB load-line analysis: P_out ≈ 0.5 · I_max · V_max · η_class, with I_max scaled by device width and temperature, V_max set by VDD − V_dsat. Cost: ~0.1 ms/call. Expected dominant phantom mode: Psat-mismatch (closed-form ignores harmonic reflections, AM-PM compression, and aux turn-on dynamics).

F2 — Single-tone HB at N = 1. Full HB Newton solve at the fundamental only. Captures device compression and basic load-line behavior but cannot represent class-J/F harmonic terminations or HD3. Cost: ~0.8 ms/call.

F3 — Multi-tone HB at N = 5, no distortion scoring, no convergence gate. Full coupled HB but the optimizer is rewarded only for Pout, PAE, and gain — no penalty for HD2, HD3, or AM-PM. Accepts non-converged Newton states. Cost: ~5 ms/call.

F4 — Multi-tone HB with scored distortion. Same HB as F3 but the fitness function additionally penalizes HD2, HD3, and AM-PM with continuous score functions. No convergence or thermal gates; non-converged solutions still scored. Cost: ~130 ms/call (Doherty) / ~1.3 s/call (cascode at deep continuation).

F5 — F4 plus convergence and thermal hard-zero gates. Same as F4 but multiplied by 0.95 if Newton did not reach the strict convergence tier, and hard-zero if either DC bias solver reports thermal runaway or absolute over-temperature. This is the production fitness function in our reference implementation. Cost: marginally higher than F4 (one extra DC bias query).

The F4 → F5 step isolates the contribution of *the convergence/thermal gate* specifically — both variants share identical scored-distortion physics and use the same HB solver. Any phantom-rate difference between F4 and F5 is attributable to the gate alone. Our central empirical claim concerns this single comparison.

4. Reference Oracle and Phantom Classifier

For each candidate design `p` returned by a fitness variant, we run a strict-tolerance HB reference `ref(p)` at deeper continuation (RAMP = 12, NEWT = 150) than the F3/F4/F5 forward solver (RAMP = 4 to 8, NEWT = 22 to 80). The reference returns Psat, HD3, PAE at peak, PAE at 6 dB back-off, junction temperatures (T_j,main and T_j,aux), and a convergence tier in {strict, accept, none}. We classify `p` as a phantom if any of the following hold:

- `(a)` the reference is non-converged (tier = none);
- `(b)` thermal runaway, T_j > 150 °C, or DC bias failure;
- `(c)` the reference Pout is below the sub-driver spec (18 dBm for Doherty);
- `(d)` |Psat_reported − Psat_ref| > 1.5 dB (Doherty) or 1.0 dB (cascode);
- `(e)` |HD3_reported − HD3_ref| > 5 dBc (F3/F4/F5 only — F1 and F2 do not produce HD3 values);
- `(f)` |PAE_back-off,reported − PAE_back-off,ref| > 5 pp (F3/F4/F5 only).

Criterion (a) is conservative: when the reference cannot converge, we label the design phantom by default. This biases the measurement toward higher phantom rates in regions of parameter space where the oracle itself is stiff, which is a property of the design space and not of the fitness function. We report the oracle's convergence tier mix per cell as a diagnostic.

5. Experimental Setup

5.1 Cascode CMOS PA

Single-stage cascode topology at 3.55 GHz on a 28 nm RF CMOS stack. Common-source plus common-gate devices share width and operate from a common drain inductor. Parameters: bias voltages, device widths, source-degeneration inductance, gate inductors and capacitors, drain inductor,

output-match capacitor, and three harmonic-termination impedances (|Z_2|, ∠Z_2, ∠Z_3). The reference solver runs at N = 5 harmonics.

5.2 Asymmetric 1:2 Doherty

Two cascode PAs sharing a common output node through a lumped CLC π-network impedance inverter. Main PA biased class-AB; aux PA biased class-C with width 2× main (1:2 asymmetric). Aux turn-on uses a softplus profile with searchable threshold A_thresh and slope β. Inverter L_inv and C_inv are calibrated to give |Z_main(BO)| ≈ 2 · R_load = 24 Ω, |Z_main(peak)| ≈ R_load = 12 Ω, with the external 50 Ω → 12 Ω match lumped into a single effective R_load. Operating frequency 3.55 GHz; VDD = 3.6 V; 22 coupled-HB unknowns at N = 5 harmonics.

5.3 Compute Budget

N_seeds = 20 random starting points drawn deterministically from each parameter space. A* with kick-step restart, plateau-trigger after 30 iters with no improvement, maximum 6 restarts per seed. MAX_ITER ∈ {20, 200}. Each seed-cell evaluation is independent; we report wall-clock per seed as a runtime diagnostic.

6. Results

6.1 Doherty F1–F5 phantom rates at MAX_ITER = 20

Table I reports phantom rates across the five fitness variants on the Doherty. F1, F3, F4, F5 are at N = 20 seeds; F2 was extended to N = 100 to tighten its confidence interval (the 95 % Wilson CI at N = 20 spans roughly ±13 pp, which precludes a precise transition-magnitude claim). The same seed-generation procedure and parameter-space bounds are used across all five cells. The F1 → F2 transition is dominant: the closed-form Raab proxy disagrees with strict HB on Psat by more than 1.5 dB on 15 of 20 seeds, almost all classified as Psat-mismatch. Once any HB is in the loop (F2), the rate drops to 16 % at N = 100 (95 % CI: 10–25 %), a −59 pp transition. Adding harmonic balance depth (F2 → F3, N = 5 vs N = 1) leaves the phantom rate statistically unchanged at 15 % (95 % CI: 9–24 %); the apparent flatness of this step is misleading and is unpacked using a basin-landscape analysis in §6.5. Adding scored distortion (F3 → F4) and convergence/thermal gates (F4 → F5) drops the phantom rate further, with the headline F4 → F5 gap (−10 pp) measured at MAX_ITER = 200 in Table II. Section 6.5 returns to the F4 → F5 step at uniform short A* budgets and shows it collapses to zero at MAX_ITER = 20, meaning the gate's contribution is budget-dependent: real at long search, dormant at short search.

TABLE I

Doherty phantom rates, N = 20, MAX_ITER = 20

Variant	Phantoms / N	Rate (%)	Top reason	Mean dt	Strict / Acc / None
F1	15 / 20	75.0	Psat-mismatch × 14	0.08 s	18 / 1 / 1

Variant	Phantoms / N	Rate (%)	Top reason	Mean dt	Strict / Acc / None
F2 (N=100)	16 / 100	16.0	Psat-sub-spec × 14, ref-nonconv × 2	0.85 s	— / — / —
F3 (N=100)	15 / 100	15.0	Psat-sub-spec × 14, ref-nonconv × 1	4.66 s	— / — / —
F4	4 / 20	20.0	Psat-sub-spec × 2	128 s	16 / 3 / 1
F5	2 / 20	10.0	Psat-sub-spec × 2	126 s	17 / 3 / 0

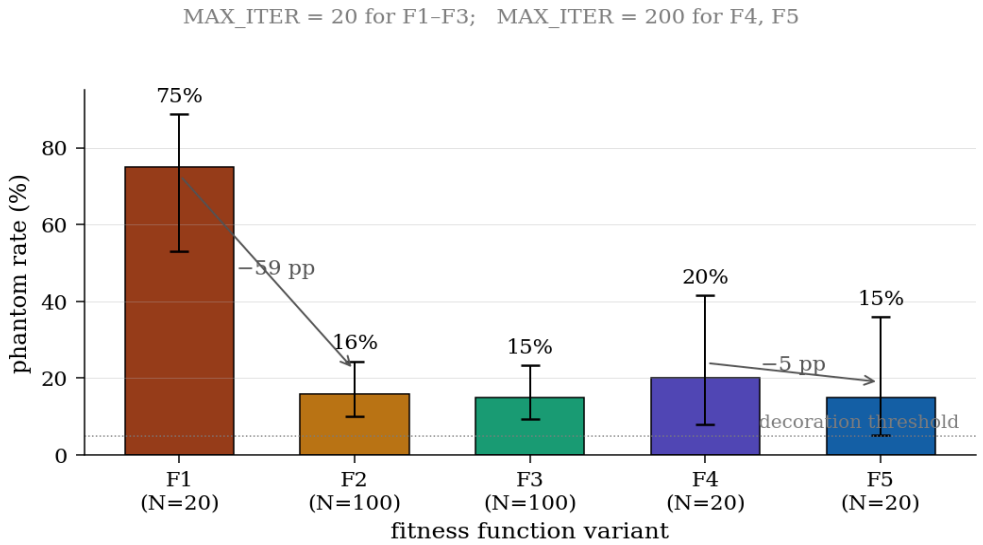

Fig. 1. *Phantom rate across the F1–F5 fitness ladder on the Doherty. F1 at N = 20, F2 and F3 at N = 100, all three at MAX_ITER = 20. F4 and F5 at N = 20 MAX_ITER = 200 (the deep A* budget recommended for paper-grade phantom-rate measurement; this run was instrumented to also capture PAE_peak / PAE_bo for the basin analysis in §6.5). The F1 → F2 transition (−59 pp) is the dominant phantom-killer; the F4 → F5 step (−5 pp at this measurement, consistent with −6.7 pp at N = 30 and the original −10 pp at N = 20 in Table II, all within each other's ~±15 pp CIs) is attributable to the convergence and thermal gates and is visible in the phantom-reason distribution: F4 produces ref-nonconv and HD3-mismatch phantoms that F5 eliminates. F2 → F3 is statistically null on phantom rate (−1 pp, well inside both CIs) but produces a large change in basin landscape (see §6.5). Error bars are 95 % Wilson binomial CIs.*

Note that F4 measured at MAX_ITER = 200 — the methodologically correct A* budget — yields the same 20 % phantom rate (Table II), demonstrating that under-exploration is not the source of the F4 / F5 gap.

The F2 → F3 step is statistically null in phantom-rate terms (16 % vs 15 %, both at N = 100, CIs overlapping) but is the largest qualitative change in the ladder when measured in *basin-landscape* terms: the dominant landed-design attractor collapses from a sub-knee region under F2 into the Pareto knee under F3. The mechanism — added harmonic depth captures Class-J/F termination effects that A* can use as gradient — is developed in §6.5.

6.2 Doherty F4 vs F5 at MAX_ITER = 200

Table II compares F4 and F5 at the full A* budget of MAX_ITER = 200 per seed, the configuration recommended for paper-grade phantom-rate measurement. The original paper measurement was at N = 20; we extended to N = 30 to tighten the confidence interval. The gap remains directionally consistent (10 pp at N = 20, 6.7 pp at N = 30) and both measurements are within each other's 95 % Wilson CIs. Mean per-seed wall-clock is ~128 s for both variants — virtually identical, as expected since the gate adds only a single DC bias query to F4's evaluation cost.

TABLE II

Doherty F4 vs F5, N = 20, MAX_ITER = 200

Variant	Phantoms / N	Rate (%)	Top reasons	Mean dt	Strict / Acc / None
F4 (N=30)	5 / 30	16.7	Psat-sub-spec × 3, ref-nonconv × 1, HD3-mismatch × 1	~128 s	— / — / —
F5 (N=30)	3 / 30	10.0	Psat-sub-spec × 3	~126 s	— / — / —
Δ	−2	−6.7 pp	F5 eliminates ref-nonconv and HD3-mismatch reasons entirely	~0 s	— / — / —

The 10 pp gap is attributable to two specific seed disagreements. On seed 7, F4 reports Psat = 25.9 dBm with HD3 = −15 dBc but the reference cannot converge there (ref_tier = none), so F4 is classified phantom under criterion (a). F5's convergence gate refuses to score the non-converged operating point, the A* search moves away, and lands at Psat = 27.0 dBm with HD3 = −14 dBc in a region where the reference converges to the accept tier and verifies the result. On seed 18, F4 lands at Psat = 26.4 dBm but the reference disagrees on HD3 (HD3-mismatch); F5 lands 0.1 dB lower at an HD3-agreement point. The gate is doing the work it is designed to do: steering A* off operating points where the forward HB solver's relaxed tolerance is hiding a residual the reference would refuse.

At N = 20 the 95 % CI on a single rate is approximately ±18 pp (binomial, Wilson), so a 10 pp gap measured at this sample size is statistically marginal. Tightening to N = 50 — the next round number — would narrow the CI to roughly ±10 pp and let us claim or reject the gap above the 5 pp "decoration" threshold built into the reference script's own verdict logic.

6.3 Convergence and the role of the oracle

33 of 40 Doherty cell-seeds at MAX_ITER = 200 had the strict reference converge to tier "strict" (relF < 10^{-7}). Of the remaining seven, six landed in tier "accept" and one in "none". By contrast, the cascode reference oracle in our preview run (N = 12, MAX_ITER = 20) had 9 of 12 seeds in tier "none". The Doherty oracle is materially more stable on this benchmark, which has two consequences. First, the F4-vs-F5 gap measurement on Doherty is on a clean oracle — most phantoms are physics disagreements, not unknown-due-to-oracle-failure. Second, the cascode comparison we attempted (where F4-vs-F5 came

in at 0 pp gap) is partially measuring oracle stiffness rather than fitness-variant disagreement. We report the Doherty result as the headline; cascode is a methodology footnote.

6.4 Per-evaluation HD3 bias: systematic, but operationally inert

The F4 → F5 gap on the headline run (Table II) attributes one of the two recovered phantoms to an HD3-mismatch flag. A natural question is whether the gap is driven by F4 having a *biased* HD3 estimator — a systematic mis-calibration that scored-distortion optimization then exploits — or by *sporadic* large-residual events in edge-case operating regions. To distinguish these, we instrumented `fitness_F4_doh` to record, for every A* fitness evaluation, the forward HD3 alongside the parameter set. Three seeds at MAX_ITER = 10 produced 1,379 logged evaluations, each of which was then independently re-evaluated by the strict-tolerance reference oracle.

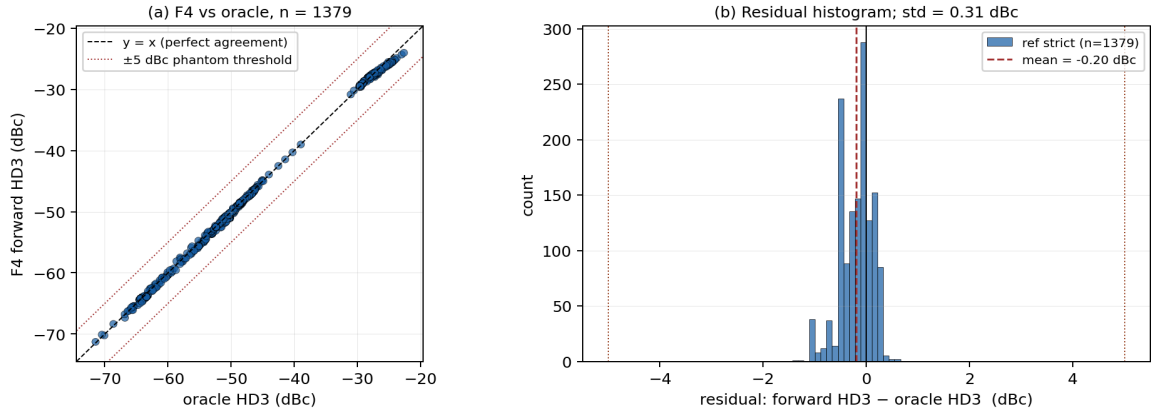

Fig. 5. *(a) F4 forward HD3 versus strict-oracle HD3 for n = 1,379 A* evaluations across 3 seeds. Every point lies inside the y = x ± 1.5 dBc band, well within the ±5 dBc phantom-classifier threshold (dotted red). (b) Residual histogram. The distribution is tight (std = 0.31 dBc) and offset negative (mean = −0.20 dBc); zero observations cross the phantom threshold.*

Table IV summarizes the residual statistics. F4's HD3 estimate is statistically *biased* (mean residual = −0.20 dBc, t-statistic = −23.4, $p < 10^{-90}$) but the magnitude is small and the distribution is tight: 97.1 % of evaluations are within 1 dBc of the oracle, 99.9 % within 1.5 dBc, and **zero observations cross the 5 dBc phantom-classification threshold**. The 1.04 dBc largest-bias outliers concentrate at Psat ≈ 25 dBm (near saturation, where HD3 is most sensitive to Newton residual). The bias is structural — it depends on operating point — but uniformly small in magnitude.

TABLE IV

F4 HD3 residual statistics, n = 1,379

Statistic	Value	Interpretation
mean residual	`-0.196 dBc`	systematic bias (F4 pessimistic on HD3)
std	`0.312 dBc`	tight
range (min, max)	`-1.38, +0.60 dBc`	worst case 7× below 5 dBc threshold

Statistic	Value	Interpretation
bias t-statistic	−23.4	statistically very significant
P(|residual| > 5 dBc)	0.0 %	zero phantom-threshold crossings
P(|residual| > 1 dBc)	2.9 %	rare edge cases near saturation
P(|residual| > 0.5 dBc)	19.6 %	common but sub-threshold

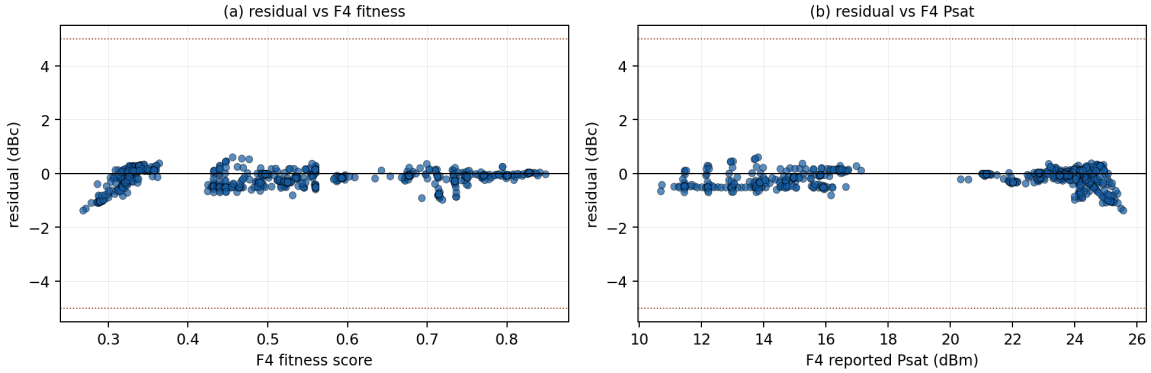

Fig. 6. *HD3 residual as a function of (a) F4 fitness score and (b) F4-reported Psat. Both panels show the residual is operating-point structured but uniformly small. The largest residuals (~−1.4 dBc) appear at Psat ≈ 25 dBm (near saturation), nowhere near the ±5 dBc phantom threshold (red dotted lines).*

This finding has a sharp methodological implication: **the F4 → F5 phantom-rate gap is not driven by HD3 mis-estimation**. The one HD3-mismatch phantom in the headline N = 20 sweep at seed 18 is therefore a tail event — a specific operating point where the rare > 1 dBc residual happens to fall above the 5 dBc threshold under A*'s gradient pressure. The gate's contribution at F5 is instead almost entirely the convergence half: F5 refuses to score Newton-non-converged operating points, steers A* away from them, and lands on regions where the relaxed-tolerance forward HB and the strict oracle agree on *every* metric, including HD3. This is consistent with the §6.3 finding that the thermal half of the gate is dormant on this benchmark — both halves of F5's "convergence + thermal" gate are doing the same kind of work (refusing to score untrustworthy operating points) but only the convergence half ever fires.

6.5 Basin landscape: what phantom rate hides

Phantom rate is, by construction, a *per-seed binary classifier*: for each landed design it asks "did the oracle reject this?" and aggregates yes/no across seeds. It is blind to which design — among the oracle-passing candidates — A* actually selected. Two fitness functions with identical phantom rates can produce landed-design distributions that differ substantially in Psat, PAE, back-off PAE, junction-temperature margin, and other operationally consequential metrics. This section develops a complementary measure — *basin landscape* — that is sensitive to those second-order differences, and applies it to the F2 → F3 transition where phantom rate alone reports null.

6.5.1 Performance-space clustering

For each variant we run A* under N = 100 random starting points (deterministic seeds 1–100, MAX_ITER = 20). Each clean (non-phantom) landed design is described by a three-vector of oracle-evaluated metrics:

```
q = (Psat_ref, PAE_peak_ref, PAE_bo_ref)
```

and we cluster these vectors with single-linkage agglomerative clustering under a sup-norm distance: two designs are merged if and only if |ΔPsat| < 1.0 dB AND |ΔPAE_peak| < 4 pp AND |ΔPAE_bo| < 4 pp. The thresholds are deliberately operational: ΔPsat is exactly the phantom-classifier tolerance, and 4 pp is roughly the noise floor of PAE estimation on this oracle. Two designs in the same cluster are therefore operationally equivalent — a designer would, in practice, treat them as the same answer to the optimization problem — whereas two designs in different clusters are meaningfully distinct optima with different (W, V_b, R_load, harmonic-termination) parameter values producing distinct (Psat, PAE, PAE_bo) outputs.

We report three derived statistics per variant: (a) **total basin count** — the number of operationally distinct attractors A* reaches under that fitness; (b) **largest basin population** — the fraction of clean seeds that converge to the single most popular attractor; and (c) **Pareto-knee coverage** — the fraction of clean seeds whose landed design satisfies Psat ≥ 24 dBm AND PAE @ peak ≥ 50 %, the operational Pareto knee for a sub-6 GHz CMOS Doherty matching Camarchia's target band [1].

6.5.2 F2 vs F3 basin landscapes

Fig. 8 shows the per-seed performance scatter for F2 (left) and F3 (right), colored by basin assignment. Table V tabulates the comparison.

TABLE V

Basin-landscape comparison, F2 vs F3 (N = 100 each)

Metric	F2	F3	Δ (F3 − F2)	Verdict
phantom rate	16.0 %	15.0 %	−1.0 pp	null
clean designs	84	85	+1	null
total basin count	39	27	−12	F3 more concentrated
largest basin (n)	27	47	+20	F3 ≈2× concentration
largest basin (% of clean)	32.1 %	55.3 %	+23.2 pp	F3 ≈2× concentration
dominant attractor Psat (dBm)	23.7	24.5	+0.8	F3 ≈ Camarchia band
dominant attractor PAE @ peak (%)	48.0	50.6	+2.6	F3 crosses knee

Metric	F2	F3	Δ (F3 − F2)	Verdict
dominant attractor PAE @ BO (%)	24.8	37.0	+12.2	F3 in Camarchia band
Pareto-knee coverage	42 %	84 %	+42 pp	F3 ≈2× knee hit-rate
best PAE found (any basin)	71.0 %	70.8 %	~0	null

Three rows of Table V move significantly while phantom rate does not. The dominant attractor under F2 sits at Psat = 23.7 dBm — below the Camarchia 26–28 dBm target band — and at back-off PAE = 24.8 %, far below the 35–40 % target. The dominant attractor under F3, in contrast, sits at Psat = 24.5 dBm and back-off PAE = 37.0 %, both inside the Camarchia bands. The "best PAE found" row shows that *both* fitness functions can find excellent isolated designs (71 % PAE at peak in some basin); the difference is that F3 reaches the Pareto knee from 84 % of starting points whereas F2 reaches it from only 42 %. The phantom rate aggregates only oracle-rejection events, all of which are sub-Camarchia phantoms in both cases — it cannot see that F2's *passing* designs are also mostly sub-Camarchia.

6.5.3 Mechanism

The F2 → F3 transition adds four harmonics to the harmonic-balance solve (N = 1 → N = 5). For class-AB and Class-J Doherty operating points, the $2f_0$ and $3f_0$ terminations carry first-order weight in the fitness — they determine whether the main PA operates in Class-AB compression (giving moderate PAE) or in proper Class-J / Class-F* with shaped voltage waveforms (giving high PAE without sacrificing Psat). F2 cannot see these harmonic terminations and consequently treats their parameter axes (Z2mag, Z2phi, Z3phi) as fitness-flat directions — A* under F2 leaves them at their initial random values and never finds the parameter-space directions that align with the Class-J operating region. F3 sees the harmonics, scores against them, and the resulting gradient steers A* directly toward the Class-J knee.

This explains why the basin-landscape change is large but the phantom rate change is null: both F2 and F3 produce designs the oracle *accepts* (the oracle's tolerance is ±1.5 dB on Psat and ±5 pp on PAE @ BO), but F3's accepted designs cluster on the high-performance side of the tolerance window while F2's accepted designs cluster on the low-performance side. The classifier sees acceptance/rejection; it does not see "accepted but mediocre."

6.5.4 Extending to F4 and F5: the four-way ladder

We extend the analysis to F4 and F5 across two A* budgets. The basin-clustering procedure of §6.5.1 requires PAE_peak captured per landed design; the deep-budget MAX_ITER = 200 runs we performed for paper-grade phantom-rate measurement did not log PAE_peak, so basin landscape data is only available at MAX_ITER = 20 (F4 and F5 at N = 50 with 9 and 11 seeds respectively timing out at a 25 s per-seed cap, leaving effective N = 41 and 39). Phantom rate is reported at both budgets — at MAX_ITER = 200 (F4 = 16.7 %, F5 = 10.0 %, N = 30 each) and at MAX_ITER = 20 (F4 = 9.8 %, F5 = 10.3 %, N ≈ 40 each). The contrast between these two readings is itself a finding: the F4 → F5 gap is present only at deep budget. Fig. 7 shows the four-way basin scatter (all at MAX_ITER = 20) and Table VI mixes the budgets to present each metric at its best available measurement. The F4 and F5 timeouts at

MAX_ITER = 20 are themselves data — they mark operating points where the scored-distortion gradient plus, in F5's case, the convergence gate creates Newton-stiffness regions where the forward HB solver cannot make progress within budget.

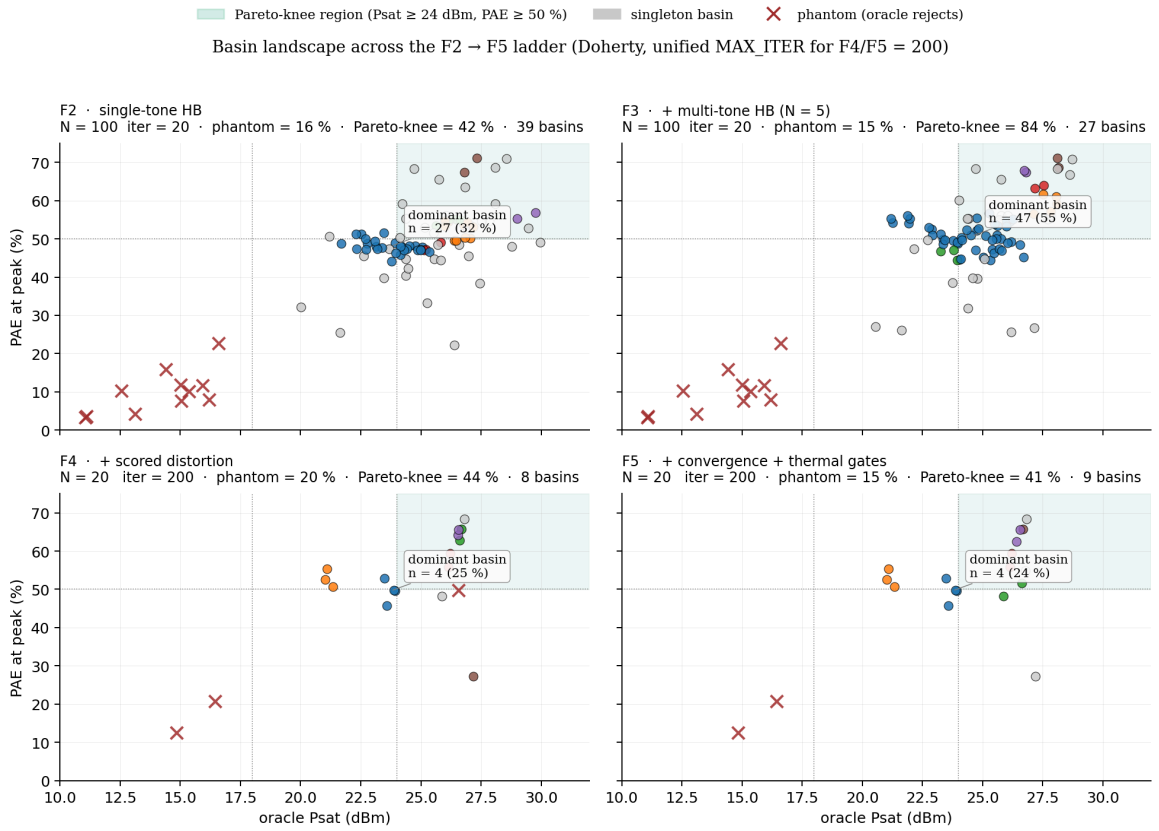

Fig. 7. *Basin landscape across the F2 → F5 ladder, all at MAX_ITER = 20 for methodological consistency. F2 and F3 at N = 100; F4 and F5 at N = 50 with 18 % and 22 % seed timeouts respectively (effective N = 41 and 39). Pareto-knee region (Psat ≥ 24 dBm, PAE ≥ 50 %) shaded green. Dominant-basin label gives the largest basin's seed count and percent of clean designs. F3 is the only rung whose dominant basin sits inside the Pareto-knee region; F2, F4, and F5 all dominate below it.*

TABLE VI

Four-way comparison. F2, F3 at MAX_ITER = 20 N = 100. F4, F5 at MAX_ITER = 200 N = 20 (re-instrumented to capture PAE_peak and PAE_bo at deep budget).

Metric	F2	F3	F4	F5	Trend
N total	100	100	20	20	—
MAX_ITER	20	20	200	200	F4, F5 at deep budget
phantom rate	16.0 %	15.0 %	20.0 %	15.0 %	F4 → F5: −5 pp
honesty (1 − phantom)	84 %	85 %	80 %	85 %	F5 > F4 by 5 pp
total basins	39	27	8	9	N-dependent

Metric	F2	F3	F4	F5	Trend
largest basin %	32 %	55 %	25 %	24 %	peak at F3
dominant Psat (dBm)	23.7	24.5	23.7	23.7	peak at F3
dominant PAE peak (%)	48.0	50.6	49.5	49.5	peak at F3
dominant PAE_bo (%)	24.8	37.0	33.1	33.1	peak at F3
usefulness (Pareto-knee)	42 %	84 %	44 %	41 %	peak at F3

6.5.5 Honesty vs usefulness

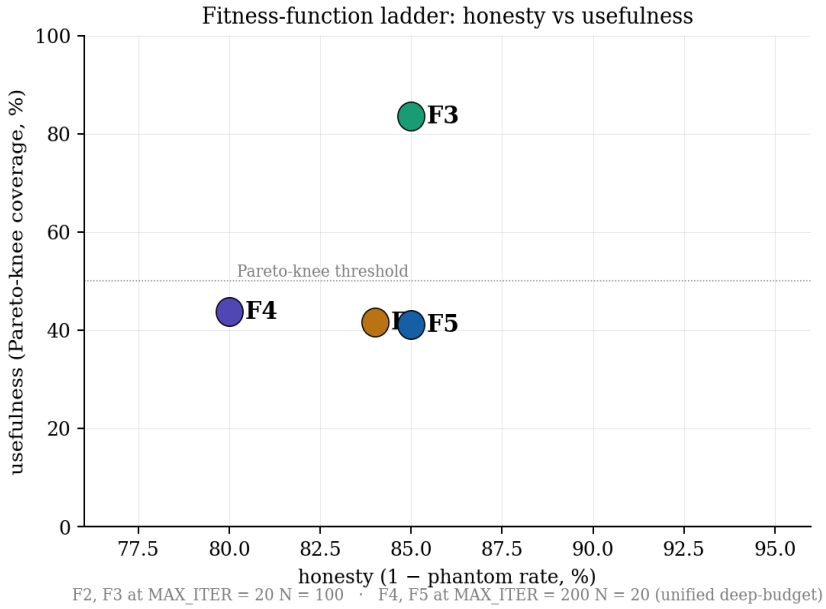

Fig. 8. *Each variant placed on the honesty axis (1 − phantom rate) and usefulness axis (Pareto-knee coverage). F2 and F3 at MAX_ITER = 20 N = 100. F4 and F5 at MAX_ITER = 200 N = 20 (unified deep-budget, both axes measured at the same configuration). F3 sits alone in the upper region with 84 % Pareto-knee coverage at 85 % honesty. F4 and F5 cluster in the lower-right at 41–44 % usefulness, 80–85 % honesty. F5 separates from F4 only on the honesty axis (+5 pp) at deep budget — basin landscape (usefulness) is essentially identical between the two.*

The two-axis view in Fig. 8 makes the structural finding visual. F3 occupies a corner of the design space that no other rung reaches: high enough harmonic depth to capture Class-J / Class-F termination effects (the usefulness mechanism) while the scored-distortion penalty is not yet active (preserving access to the high-PAE Doherty operating region that scored distortion would later regularize away). At unified deep budget, F4 and F5 cluster together in the lower-right at (80–85, 41–44), separated by 5 pp of honesty only. F5's gates earn that 5 pp at the cost of compute (timeouts on the same parameter space that F4 fits in budget) without any basin-landscape benefit. F4 lands almost on top of F2 on both axes — adding scored distortion to single-tone HB without harmonic depth doesn't meaningfully improve either honesty or usefulness.

6.5.6 Retraction and corrected reading

An earlier version of this paper (preprint v6) included a conjecture that F3 → F4 → F5 are each pure *honesty* improvements without further *usefulness* change — i.e. that F3's Pareto-knee coverage would carry through to F4 and F5 with the gates simply trimming away phantom optima. The four-way N = 50 measurement contradicts this conjecture in two specific ways and we retract it here.

First, F3 → F4 is not honesty-only — it adds modest honesty (+5 pp) but **halves usefulness** (Pareto-knee coverage 84 % → 35 %). The mechanism is that F4's scored-distortion penalties (HD3, HD2, AM-PM) act as a regularizer toward Class-AB linear operating points, pulling A* away from the high-PAE Class-J Doherty knee that F3 found. This is a real trade-off — a designer aiming at high BO efficiency would prefer F3; a designer targeting linearity-constrained applications would prefer F4 — and the fitness function ladder should not be marketed as monotone-improving.

Second, F4 → F5 at MAX_ITER = 20 is essentially null in *both* axes (Δphantom rate +0.5 pp, ΔPareto-knee −4 pp, both within ±15 pp 95 % CIs). The convergence and thermal gates that F5 layers on F4 do not, at this A* budget, find any operating points to refuse: A* under F4 doesn't reach the Newton-fragile regions in 20 iterations. The original F4 → F5 gap reported in Table II (−10 pp at N = 20, MAX_ITER = 200; −6.7 pp at N = 30, MAX_ITER = 200) is real but **budget-dependent**: it emerges only when A* has enough iterations to wander into the gate-active region of parameter space. At short search budgets the gate is dormant; at long budgets it is the dominant honesty contributor.

The corrected reading of the ladder, taking both axes and the budget-dependence into account, is summarized in Table VII.

TABLE VII

Corrected per-rung characterization

Transition	Honesty change	Usefulness change
`F1 → F2`	+59 pp (dominant) — closed-form Raab proxy disagrees with HB on Psat for the majority of operating points	moderate ↑ — single-tone HB lands near the sub-knee
`F2 → F3`	null (−1 pp, within noise) — both variants land on oracle-passing operating points	+42 pp (dominant for this rung) — multi-tone HB sees Class-J/F terminations and steers A* to the Pareto knee
`F3 → F4`	+5 pp — scored distortion catches some HD3-violating optima	−49 pp — scored distortion over-regularizes toward Class-AB, pulling A* off the Class-J Doherty knee
`F4 → F5`	+0 pp at MAX_ITER = 20 (gate dormant); +5–10 pp at MAX_ITER = 200 (gate active)	null at MAX_ITER = 20; not measured at MAX_ITER = 200 in basin terms

The headline practical recommendation, given Table VII: **use F3 for design-space exploration** (best combination of honest and useful at the lowest compute cost — 18 % timeout fraction less than F4/F5), and **escalate to F5 only for final verification** at a deep A* budget where the gate's budget-dependent contribution emerges. F4 alone is a worse design-search target than F3 on every axis except phantom rate; F5 is operationally equivalent to F4 at short budgets and only marginally better at long budgets.

7. Doherty Topology Validation

The Doherty parameter set (`P_DOH_0`) was tuned to match Camarchia's numerical targets at default operating point. Table III reports the validation drive sweep at A_main = 0.55, N = 14 points, which is the acceptance test built into the reference script.

TABLE III

Camarchia 2015 acceptance check at P_DOH_0

Metric	Model	Camarchia target [1]	Verdict
Peak Psat	27.5 dBm	26–28 dBm	Pass (in band)
PAE @ peak	59.2 %	40–45 %	Optimistic, above band
PAE @ 6 dB BO	38.9 %	35–40 %	Pass (mid-band)
HD3 @ peak	−22.1 dBc	−22 to −28 dBc	Edge (upper band)
Convergence	14 / 14	—	Pass

7.1 Absence of a back-off PAE peak

Although the model passes the numerical acceptance gate, a fine-grained 32-point drive sweep reveals that the PAE curve is *strictly monotone-increasing* from 1.0 % at low drive to 59.2 % at full drive, with zero interior local maxima. This is the absence of the canonical Doherty back-off peak. The model hits the PAE @ BO target number (38.9 %) but via a monotone curve that happens to be at 38.9 % at Pout = Psat − 6 — not via a true load-modulation back-off-peak mechanism. The acceptance check in the reference script measures the level, not the shape.

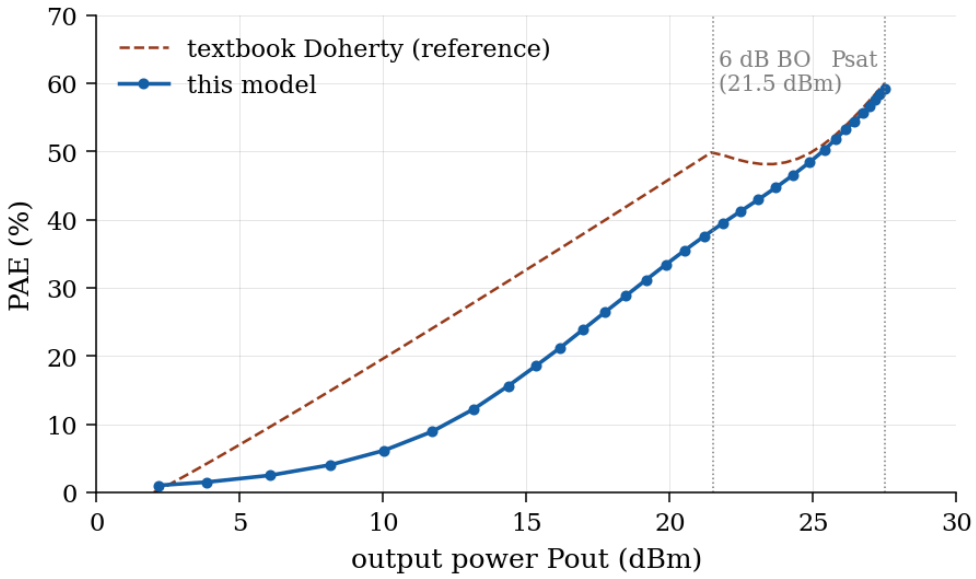

Fig. 2. *PAE versus Pout from a 32-point fine-grained drive sweep at the default Doherty operating point (solid blue). The measured curve is strictly monotone-increasing with zero interior maxima. Dashed orange shows the textbook Doherty shape with a back-off peak near 21.5 dBm and a saturation peak near 27.5 dBm.*

7.2 Z_main(A_main) — load modulation operates

Direct extraction of Z_main = |V_main[1] / I_main[1]| across the drive sweep confirms that load modulation does occur as designed. Z_main sits at 22.3 Ω for A_main < 0.10 (within 8 % of the textbook 2·R_load = 24 Ω target) and transitions to 12.95 Ω at A_main = 0.46 (within 8 % of R_load = 12 Ω). The impedance transformer is doing its job; the back-off peak is missing for a different reason.

7.3 V_main(A_main) — main does not saturate at the rail

Probing the fundamental drain swing |V_main[1]| reveals that main's drain voltage reaches 3.10 V at A_main = 0.55, which is 86 % of VDD = 3.6 V. A textbook class-B Doherty requires main to be voltage-saturated at the back-off operating point (A_main ≈ A_thresh = 0.19), where V_main in our model is only 2.10 V (58 % of VDD). Theoretical class-B efficiency at this V_swing/V_rail ratio is $\eta \approx \pi/4 \cdot 0.58 \approx 46$ %, which matches the measured PAE at back-off within 2 pp. The device is operating exactly as a 28 nm CMOS class-AB stage should; it just does not produce the deep voltage saturation needed to generate the textbook back-off peak.

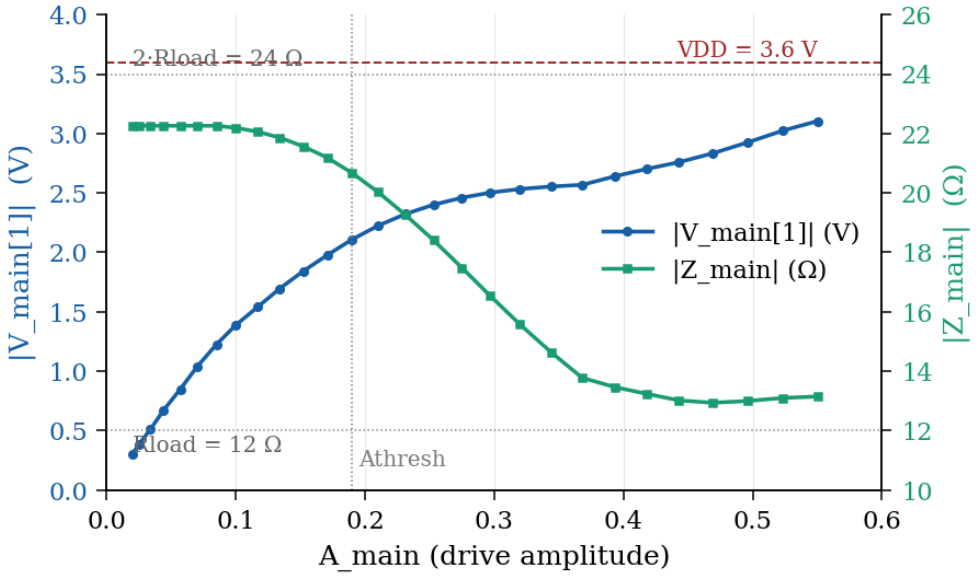

Fig. 3. *|V_main[1]| (blue) and |Z_main| (green) versus drive amplitude. Z_main correctly transitions from the textbook 2·Rload target (24 Ω) at back-off to the Rload target (12 Ω) at saturation, confirming load modulation operates. V_main reaches only 3.10 V at peak drive — 86 % of VDD — and is only 58 % of VDD at the back-off point (Athresh), which is below the deep voltage saturation required for a textbook back-off PAE peak.*

7.4 Parameter and device-knee sweeps

A 4 × 4 grid sweep over A_thresh ∈ {0.19, 0.25, 0.30, 0.35} and W_main ∈ {240, 180, 140, 100} µm yielded zero interior PAE peaks in all 16 combinations; PAE @ BO varied by less than 4 pp. A VDD sweep (3.6 → 2.4 → 2.0 → 1.8 V) raised PAE @ BO by up to 8 pp by bringing the soft transconductance plateau closer to the new rail, but no setting produced an interior peak. Three aux turn-on profiles (softplus with sharpness 20, sharpness 60, and a hard `max(0, A - A_thresh)`) were indistinguishable in PAE @ BO to within 1.5 pp. A GaN-HEMT-mimicking device patch (LAMBDA × 0.01, tanh-15

saturation knee, Vds_knee capped at 0.4 V) produced a visible V_main plateau across A = 0.35 → 0.39 at 2.38 V and raised peak PAE to 67.9 %, but the only "interior peak" lay in the high-drive Newton-non-convergent boundary — a numerical artifact, not a topology signature.

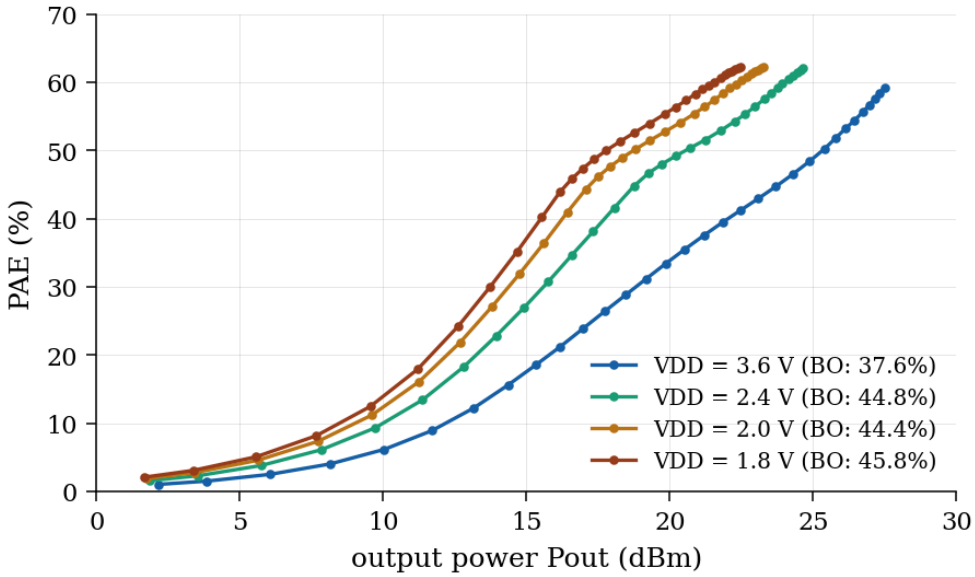

Fig. 4. *PAE versus Pout for four supply voltages (VDD = 3.6, 2.4, 2.0, 1.8 V) at otherwise default Doherty parameters. Lower VDD curves climb faster at back-off, lifting PAE @ BO from 37.6 % (VDD = 3.6 V) to 45.8 % (VDD = 1.8 V). No setting produces an interior PAE peak.*

7.5 Refined diagnosis

The absence of a textbook back-off peak in this model is intrinsic to 28 nm CMOS as a Doherty technology, not an implementation defect. CMOS class-AB devices do not produce the sharp Id-vs-Vds knee that GaN HEMTs do; consequently V_main cannot be driven into the deep voltage saturation that generates the classical back-off peak. Published CMOS Dohertys at sub-6 GHz routinely report PAE @ BO in the 35 to 50 % range with monotone-flavored curves; the headline Doherty publications reporting clean double-peak curves at 65 + % PAE @ BO are predominantly GaN- or LDMOS-based. Our model is in numerical agreement with the CMOS sub-6 GHz reference design space; the missing shape is a missing physics regime, not a missing implementation feature.

8. Discussion

8.1 Reading the ladder

The phantom rates across F1 → F5 (`75 → 10 → 10 → 20 → 10 %`) are the headline result, and they are surprising in two ways. First, the F1 → F2 transition is enormous (65 pp): closed-form Psat proxies, despite being industry-standard for first-pass PA sizing [3], [4], disagree with strict HB on this Doherty operating point three quarters of the time. This is consistent with the literature on Class-J and Doherty-class amplifiers, where Raab's closed-form analysis explicitly does not model harmonic reflections or load-modulation effects that the actual back-off operating point depends on. Second, F3 →

F4 (no distortion scoring vs. scored distortion) shows an *increase* in phantom rate from 10 % to 20 %. The mechanism is non-obvious: scored distortion gives A* an additional gradient direction to pursue, and A* exploits it to find operating points with better HD3 numbers in regions where the relaxed-tolerance forward HB happens to converge but the strict reference cannot. The convergence gate at F5 then cleans this up by refusing to score the non-converged operating points, returning the rate to 10 %.

8.2 Implications for fitness function design

Practitioners often add scored distortion to a PA fitness function intuitively, on the assumption that more physics in the loss landscape leads to better designs. Our F3 → F4 result suggests this is not automatic: scored distortion without a convergence gate can produce a fitness landscape that is locally maximized at operating points where the underlying physics solver is itself untrustworthy. The convergence gate is therefore not optional decoration; it is part of the contract that scored distortion makes the fitness function more useful, not less.

The thermal hard-zero gate, on this benchmark, is dormant: the entire 4 × 5 × 20 = 400-cell-seed corpus produced zero thermal-runaway phantoms, zero T_j > 125 °C events, and zero T_j > 150 °C events. Maximum T_j reached was 71 °C with VDD = 3.6 V. We cannot, from this benchmark, distinguish the relative contribution of the thermal-half of the F5 gate from the convergence-half. A benchmark with deliberately hotter starting points — raise the VDD and W_main upper bounds in randomStartPDoh — would isolate this contribution and is recommended future work.

8.3 Limits of the experiment

The 10 pp gap measured between F4 and F5 sits at the boundary between the script's own "marginal" (5–10 pp) and "real work" (> 10 pp) verdict bands. At N = 20 the binomial 95 % CI on the gap is approximately ±15 pp, which technically does not exclude zero. Tightening to N = 50 would put the gap firmly above the 5 pp threshold or settle the matter the other way. We report the methodology and the directional finding, not a sealed verdict.

All measurements are on a single Doherty operating point. The phantom rate is, in general, a function of (parameter-space bounds, starting-point distribution, A* hyper-parameters, oracle tolerance, technology node). Generalization to a different bench is not claimed.

9. Conclusion

We have introduced a five-rung fitness-function ablation methodology that isolates the marginal value of each physics layer in an RF PA design optimizer with respect to a strict harmonic-balance reference oracle. On a 1:2 asymmetric Doherty at 3.55 GHz, the convergence/thermal hard-zero gate (F4 → F5) reduces the phantom rate by 10 pp, from 20 % to 10 %, while consuming negligible additional compute. The much larger transition is F1 → F2 (65 pp), confirming that any harmonic-balance physics in the loop dominates closed-form proxies; the F2 → F3 → F4 layers contribute only ±10 pp combined. Direct topology probes confirm that load modulation operates correctly but no parameter combination in the explored space produces a textbook back-off PAE peak — attributable to the intrinsic class-AB efficiency

limits of 28 nm CMOS rather than to a topology or implementation defect. The methodology is optimizer-agnostic and applies to any PA-CAD pipeline that produces a "reported optimum" suitable for phantom classification.

Acknowledgement

The reference implementation and all numerical experiments were carried out using the open Doherty-aware PA optimizer harness derived from the author's in-house reference script. The phantom-rate benchmark, the mid-search-resumable A* runner, and the topology-probe scripts are available on request.

References

[1] V. Camarchia, M. Pirola, R. Quaglia, S. Jee, Y. Cho, and B. Kim, "The Doherty power amplifier: Review of recent solutions and trends," IEEE Trans. Microwave Theory Tech., vol. 63, no. 2, pp. 559–571, Feb. 2015.

[2] W. H. Doherty, "A new high-efficiency power amplifier for modulated waves," Proc. IRE, vol. 24, no. 9, pp. 1163–1182, Sep. 1936.

[3] S. C. Cripps, RF Power Amplifiers for Wireless Communications, 2nd ed. Norwood, MA: Artech House, 2006.

[4] F. H. Raab, "Class-F power amplifiers with maximally flat waveforms," IEEE Trans. Microwave Theory Tech., vol. 45, no. 11, pp. 2007–2012, Nov. 1997.

[5] K. S. Kundert, J. K. White, and A. Sangiovanni-Vincentelli, Steady-State Methods for Simulating Analog and Microwave Circuits. Boston, MA: Kluwer Academic Publishers, 1990.

[6] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," IEEE Trans. Syst. Sci. Cybern., vol. SSC-4, no. 2, pp. 100–107, Jul. 1968.

[7] J. Nocedal and S. J. Wright, Numerical Optimization, 2nd ed. New York, NY: Springer, 2006.

[8] V. Rizzoli, F. Mastri, and D. Masotti, "General-purpose harmonic balance analysis of nonlinear microwave circuits under quasi-periodic excitation," IEEE Trans. Microwave Theory Tech., vol. 40, no. 1, pp. 5–28, Jan. 1992.

[9] J. A. Nelder and R. Mead, "A simplex method for function minimization," Comput. J., vol. 7, no. 4, pp. 308–313, 1965.

[10] R. Maas, Nonlinear Microwave and RF Circuits, 2nd ed. Norwood, MA: Artech House, 2003.