QUANTIFYING NONLINEAR BEHAVIOR IN DIGITAL MOOG LADDER FILTERS: CROSS-IMPLEMENTATION COMPARISON AND COMMON-CORE ABLATION

Hiroyuki Oyama

Independent Researcher
Tokyo, Japan
oyama@module.jp

ABSTRACT

Digital Moog ladder filters are often compared by their linear frequency response, even though musically important differences emerge under nonlinear operation. This paper presents an open, reproducible SPICE-referenced evaluation framework that combines two components: a cross-implementation comparison of five digital ladder-filter implementations against the same SPICE reference, and a controlled common-core ablation. The cross-implementation comparison shows that close linear agreement can mask substantial nonlinear divergence, while the ablation shows that ladder nonlinear behavior depends not only on saturator choice and nonlinearity placement, but also on the non-uniform contribution of different ladder stages: in partial ablations of a SPICE-referenced TPT/ZDF core, retaining earlier-stage nonlinearities preserves harmonic behavior better than retaining later-stage nonlinearities alone. Beyond the specific comparisons reported here, the framework provides a reproducible basis for evaluating fidelity–cost trade-offs when nonlinear structure must be simplified.

1. INTRODUCTION

The Moog ladder filter remains a canonical target of virtual-analog audio DSP. Its musical identity depends not only on resonant low-pass behavior, but also on self-oscillation, drive-dependent cutoff motion, and input-dependent harmonic coloration. Digital ladder filters are nevertheless still compared primarily through linear frequency-response plots, even though nonlinear circuit behavior is often crucial to the resulting sonic character [1, 2]. That view is useful, but incomplete: two implementations can look very close linearly while diverging once nonlinear stages become active.

This paper uses a common SPICE-simulated ladder circuit as a reference across all tests. Rather than proposing another ladder algorithm, it focuses on cross-implementation evaluation of canonical literature models and widely used open-source implementations under the same SPICE baseline. The work makes two contributions. First, it provides a reproducible SPICE-referenced evaluation framework for comparing practical ladder-style implementations within a shared quantitative frame, centering nonlinear behavior through complementary harmonic, cutoff-shift, and self-oscillation views. Second, it uses a controlled structure/function ablation on a common ladder core to show that nonlinear behavior depends not only on saturator choice and placement, but also on the non-uniform contribution of different ladder stages.

The same framework also serves as a practical aid for fidelity–cost trade-offs: the metrics quantify which analog-referenced behaviors are retained and which are lost when nonlinear structure is simplified, since cutoff-shift tracking, self-oscillation, and harmonic spectra need not all be preserved equally well.

2. RELATED WORK

Digital Moog-ladder modeling has a long history. The underlying analog topology goes back to Moog's early voltage-controlled filter work [3]. Rossum showed early on that placing saturation nonlinearity within the feedback path of a digital filter is a practical route to analog-like overload behavior [4]. Stilson and Smith framed the ladder as a digital implementation problem in its own right [5]. Huovilainen presented a widely cited nonlinear ladder formulation based on a cascade interpretation [6]. Fontana proposed a structure-preserving digital reformulation of the Moog VCF [7]. Zavalishin's treatment of topology-preserving and zero-delay-feedback filters provided a highly influential practical framework for virtual-analog design [8]. D'Angelo and Välimäki later refined nonlinear ladder modeling and compared an improved model against a SPICE reference [9].

Most prior work has focused on designing better models rather than on applying a unified benchmark across structurally different implementations. While individual models are typically validated against SPICE or hardware measurements in their own publications [6, 9], systematic cross-implementation evaluation of production library code against a common circuit reference is less commonly reported in the ladder-filter literature. Related ideas also appear in guitar-effects research, where physically informed simplifications and SPICE-compared circuit models have been used for distortion and pedal modeling [10, 11, 12]. The focus here is a compact comparison framework that places practical open-source implementations and a controlled common-core ablation within one shared SPICE-referenced quantitative frame.

3. METHOD

3.1. Reference and Compared Implementations

The reference is a ladder-filter circuit simulated in SPICE. The analog baseline is generated with ngspice 46, chosen as a free, scriptable, open-source simulator for reproducibility. Since SPICE integrates on a non-uniform adaptive grid rather than at a fixed sample rate, the maximum time step is capped at $1/8$ of the benchmark sample period and the output is resampled onto the uniform grid by linear interpolation. The reference keeps the

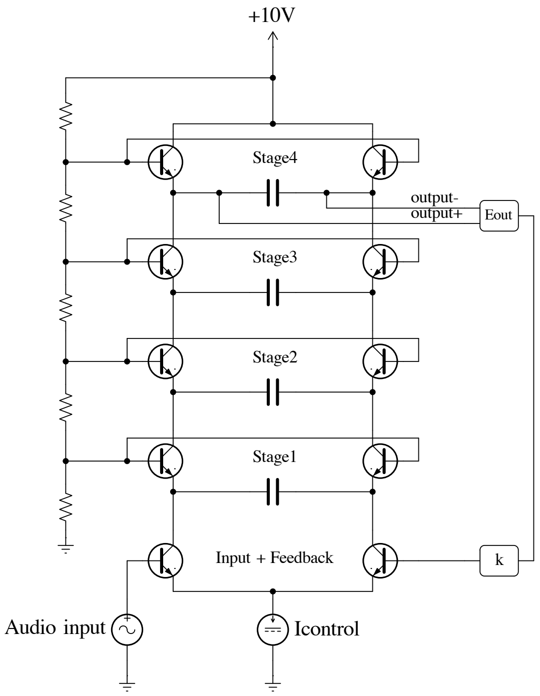


Figure 1: SPICE reference circuit used as the benchmark baseline. The schematic makes explicit the input differential pair, the four stacked NPN differential-pair stages, the differential output observation block E_{out}, and the resonance feedback gain k. Transistors use a generic idealized NPN model ($\beta_F = 10000$, $I_S = 1 \times 10^{-14}\,\mathrm{A}$, $V_{AF} = 100\,\mathrm{V}$, zero junction capacitance); supply $10\,\mathrm{V}$, ladder capacitors $68\,\mathrm{nF}$.

core Moog-ladder path explicit (the input differential pair and four stacked ladder stages) while idealized sources set cutoff (tail current $I_{\mathrm{control}} = 8\pi V_T C f_c$, $V_T = 26\,\mathrm{mV}$), resonance (gain k), and differential output. The transistors use a generic idealized NPN model rather than a specific device, so saturation follows the emitter-coupled tanh alone. This gives five saturation points in a fully nonlinear implementation, referred to throughout as "1+4": one at the input differential pair and one per ladder stage. This count is distinct from the four filter poles that define the ladder's low-pass topology, and that distinction becomes consequential in the structure/function ablation. The complete SPICE netlist is provided in the accompanying repository. The present work follows established practice in using SPICE as an analog baseline [9, 12], but extends it by applying the same reference and the same quantitative metrics across multiple implementations. The reference circuit is shown in Fig. 1.

The comparison set is chosen to maximize structural contrast while retaining two canonical literature models. Huovilainen 2004 [6] and D'Angelo 2014 [9] are included as those canonical references. The main figures therefore use SPICE, Huovilainen 2004, D'Angelo 2014, Csound, VCV Rack, and JUCE. Zavalishin 2012 is reserved as the ablation platform: fixing a single architectural core removes model-to-model structural variability, so that saturator function and nonlinear placement can be varied one factor at a time.

3.2. Compared Model Characteristics

The comparison set deliberately spans models with different provenance and different notions of what counts as a "ladder" implementation. Two canonical literature implementations anchor the set: Huovilainen 2004 [6] and D'Angelo 2014 [9]. The remaining models are practical implementations drawn from publicly available open-source projects, chosen to broaden the structural contrast within the cross-implementation comparison.

Huovilainen 2004 and D'Angelo 2014 are used directly from their published formulations. The three open-source library implementations are evaluated at specific tagged releases; the source code at each tag was inspected to confirm the internal structure and parameter conventions. In brief, Csound `6.18.1` is represented by its built-in `moogladder` opcode, VCV Rack `v2.6.4` by the Fundamental package's `VCF` module implementation, and JUCE `8.0.13` by `juce::dsp::LadderFilter`. Csound and VCV Rack both retain a distributed nonlinearity across the ladder, whereas JUCE applies `tanh` only at the signal input and in the feedback path, leaving the four filter sections otherwise linear; this topology is referred to here as boundary saturation. As documented in the Csound manual, the `moogladder` opcode follows the Huovilainen cascade lineage, consistent with its close agreement here; VCV Rack's `VCF`, by contrast, cites no specific model in its source and integrates a distributed Padé-`tanh` ladder ODE with RK4 (Table 1).

3.3. Common Interface Conditions

All digital models are evaluated through the same sample-by-sample interface with cutoff in Hz, resonance parameter k, and scalar input/output samples. This interface is only a boundary convention for the benchmark. When an implementation assumes a different internal signal scale from the analog reference, the boundary layer applies the input/output scaling needed to place the model in a comparable operating domain. When an implementation uses a different native resonance range, the boundary layer maps it to a shared nominal convention in which self-oscillation begins near $k \approx 4$, as an empirically observed onset region in the SPICE reference. These boundary conventions are summarized in Table 1.

The purpose is to compare model cores rather than host conventions: cutoff, resonance, and level are aligned only at the external boundary, while each model's internal state update, nonlinearity placement, and feedback structure are left untouched. Huovilainen 2004 and D'Angelo 2014 are left in their native formulations. Csound receives an input pre-drive, matching output trim, and the resonance mapping $k \mapsto k/4$; the resulting scale factor is $23.48 = 1.22070\,\mathrm{V}/(2V_T)$ from the `moogladder` normalization constant at $2V_T = 0.052\,\mathrm{V}$. VCV Rack applies $k \mapsto \sqrt{k/10}$ and maps the $\pm 5\,\mathrm{V}$ I/O convention to the benchmark's $2V_T$ domain. JUCE normalizes I/O to the $2V_T$ domain, cancels the internal `outputGain` of 1.2, and applies a DC passband correction $(1+4r)/\big((1+k)(1+2r)\big)$ where $r = 0.1+0.9\,(k/4.5)$ is JUCE's internal scaled resonance;[1] its resonance mapping is $k \mapsto k/4.5$.

[1]From `jmap(resonance,0.1,1.0)` in `updateResonance()`, JUCE 8.0.13 `juce_LadderFilter.h`; the benchmark maps $k \mapsto k/4.5$.

Table 1: Boundary conventions used to place each model in a common operating domain. These are input/output and control-range normalizations only; the internal DSP structures are left unchanged.

Model	Core character	I/O scale at boundary	Resonance mapping
Huovilainen 2004	1+4 stage `tanh` cascade, 2× OS	unity / unity	direct k
D'Angelo 2014	1+4 stage `tanh`, delay-free loop	unity / unity	direct k
Csound	1+4 stage `tanh` cascade, 2× OS	$\times 23.48$ / $\times(1/23.48)$	$k \mapsto k/4$
VCV Rack	1+4 stage Padé-approx. `tanh`, implicit 1-sample FB delay (RK4)	$\times 5/(2V_T)$ / $\times(2V_T)/5$	$k \mapsto \sqrt{k/10}$
JUCE	`tanh` at signal input and FB path; four sections linear	$\times 1/(2V_T)$ / $\times 2V_T$; gain cancel; DC corr.	$k \mapsto k/4.5$

Table 2: Main test conditions used in the benchmark.

Test	Conditions
Linear response	1 s logarithmically swept sine, 0.01 V, 10 Hz–20 kHz, $k = 2$, 96 kHz
Self-oscillation	1 s excitation at 1570.8 Hz, input 1 V, cutoff 1000 Hz, $k = 4.3$, final 0.5 s RMS, 96 kHz
Cutoff shift	Sine sweep around resonance, cutoff 1000 Hz, $k = 2$, amplitude sweep 0.001 V–0.104 V, 96 kHz
Harmonic spectra	Farina exponential swept sine (ESS), cutoff 1000 Hz, $k = 4$, input 0.1 V, 384 kHz
Cutoff sweep	Sawtooth 130.813 Hz (approx. C3), cutoff 20 Hz–15 kHz, $k = 3.5$, 96 kHz
Drive sweep	Sawtooth 65.41 Hz (approx. C2), cutoff 1000 Hz, $k = 0$, input 0.001 V–0.39 V, 96 kHz

3.4. Test Cases and Metrics

The linear frequency-response plot is derived from the response to the same one-second logarithmically swept sine of fixed amplitude 0.01 V from 10 Hz to 20 kHz at 96 kHz, using swept-sine deconvolution to recover a common small-signal transfer-function estimate across the comparison set [13]. All tests except the harmonic spectra measurement use 96 kHz as a shared operating rate to keep the comparison conditions consistent while staying well above the 20 kHz band of interest, as detailed in Table 2.

The harmonic spectra test likewise uses Farina-style swept-sine deconvolution to recover nonlinear harmonic spectra [13]: after deconvolution the n-th harmonic component is time-gated at a known $\log(n)$ offset, recovering H2–H10 without cross-contamination. This test alone runs at 384 kHz so that the wideband ESS retains enough margin through H10; H2–H9 is reported for consistency with the cutoff-sweep and drive-sweep views, which cap at H9. For the cutoff-sweep and drive-sweep views, harmonic levels are read directly from steady-state output spectra under the stated sawtooth drive conditions, then converted to H2–H9 levels relative to the fundamental before comparison against SPICE.

Self-oscillation sustain is measured after a one-second excitation is removed, with RMS taken over the final 0.5 seconds. The excitation frequency 1570.8 Hz differs from the self-oscillation frequency near 1000 Hz cutoff to keep the driven and oscillating states visually distinct; the test runs at $k = 4.3$, a margin above onset so that all nonlinear models sustain reliably. Large-signal cutoff shift is measured by sweeping the sinusoidal drive frequency around the resonance region for each input level and tracking the fitted peak frequency and peak gain. The input-level range for the drive-sweep test spans from the nearly linear regime into the heavily saturated region while staying within the stability limits of the reference circuit.

Aliasing is not treated here as an independent comparison axis: alias rejection changes substantially with sampling rate, internal oversampling, and host detail, so a fair aliasing comparison would need a separate design. Internal oversampling is a documented property of each implementation, left untouched. All models are evaluated through the same external 96 kHz interface (with the SPICE baseline simulated at 8× oversampling before resampling), so the benchmark fixes the operating condition rather than imposing an oversampling ratio. Spot checks above 96 kHz on the harmonic, cutoff-sweep, and drive-sweep conditions showed broadly stable rankings: the grouping between distributed 1+4 stage and boundary-saturation models was preserved, with only minor swaps among the closest models. Audio-rate modulation and alias-rejection trade-offs therefore remain outside the present scope.

For the harmonic plots, the benchmark intentionally avoids reducing each run to a single scalar too early. Harmonics are first tracked per order and per operating point; within each run the harmonic levels are normalized to that run's own H1, and only then are the H2–H9 differences against SPICE formed and summarized as RMSE. This preserves enough structure to distinguish three practically distinct failure modes: a model may match one static spectrum while drifting under cutoff motion, enter its nonlinear regime at the wrong level, or produce the right amount of distortion with the wrong harmonic balance.

The main narrative uses one linear overview, self-oscillation sustain, nonlinear cutoff shift, and three nonlinear harmonic views: static structure versus output frequency at fixed cutoff, resonance, and drive; harmonic change during a cutoff sweep at $k = 3.5$ (close to self-oscillation onset without crossing into sustained oscillation); and harmonic growth with input amplitude at $k = 0$, so that resonant feedback does not complicate attribution to the nonlinear stages.

The metric design follows the same logic. Each test is first reduced to a directly comparable observable relative to SPICE, then summarized by a compact error measure suited to that observable: a dB-scale magnitude RMSE from SPICE for linear response, final-window RMS for self-oscillation sustain, peak-tracking error for cutoff shift, and H2–H9 H1-normalized RMSE for the harmonic tests. These summaries keep qualitatively different plots connected through the same reference frame rather than collapsing the comparison to a single ranking; the figures remain primary.

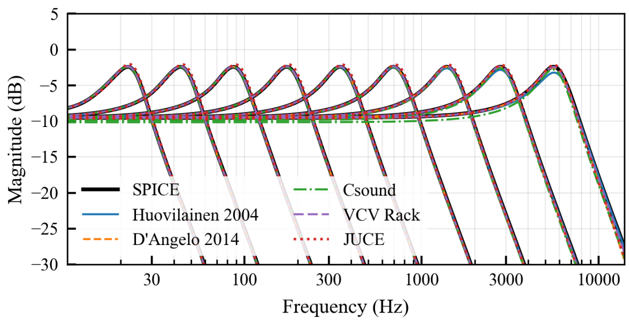


Figure 2: Linear frequency response for the comparison set: SPICE, Huovilainen 2004, D'Angelo 2014, Csound, VCV Rack, and JUCE. Responses are measured from a 1-s logarithmically swept sine signal at $k = 2$, input 0.01 V (Table 2).

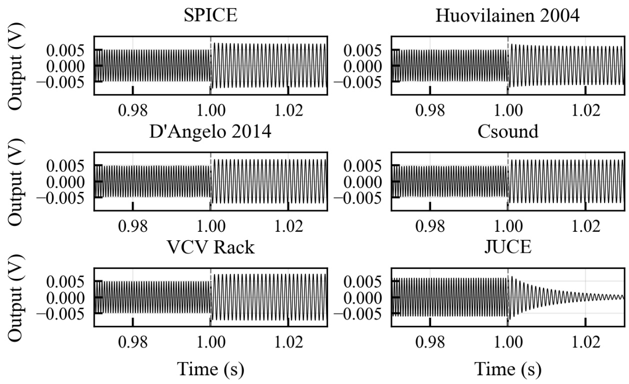


Figure 3: Self-oscillation sustain for SPICE, Huovilainen 2004, D'Angelo 2014, Csound, VCV Rack, and JUCE. A sinusoidal excitation with filter cutoff 1000 Hz, $k = 4.3$, input 1 V is applied and then removed; the residual oscillation level is measured over the final 0.5 seconds (Table 2).

4. RESULTS

4.1. Linear Frequency Response

Fig. 2 shows that several structurally different implementations remain very close in the linear domain. Huovilainen 2004 and VCV Rack are the closest non-SPICE models (0.116 dB and 0.162 dB linear RMSE), with JUCE, D'Angelo 2014, and Csound still sharing essentially the same small-signal low-pass response. The linear domain alone therefore has limited discriminative power between these implementations. JUCE remains competitive in linear RMSE while diverging much more clearly in the nonlinear tests below.

4.2. Self-Oscillation Sustain

Fig. 3 shows the same subset near the oscillation boundary. The test uses $k = 4.3$: near onset the oscillation amplitude is highly sensitive to small differences in onset threshold across models, so a margin above $k = 4$ is chosen to provide a stable comparison point. D'Angelo 2014, Csound, and VCV Rack remain close to the SPICE sustain level (4.803 mV), indicating that their self-oscillation onset is close to that of SPICE at this operating point. Huovilainen 2004 sustains at a lower amplitude (4.262 mV), consistent with its omitted resonance-correction polynomial (the amplitude/resonance-compensation factor $\mathrm{acr} \approx 1.016$ at this operating point), which slightly reduces effective resonance gain and raises the onset threshold. JUCE returns to silence after the excitation is removed. In JUCE's boundary-saturation topology, the four filter sections remain linear and the resonance gain passes through a `tanh`-limited feedback path. For small-signal onset this reduces to a loop-gain condition that JUCE's resonance mapping places at the marginal threshold only at $k = 4.5$, keeping it below unity throughout the range tested here ($k \leq 4.3$).

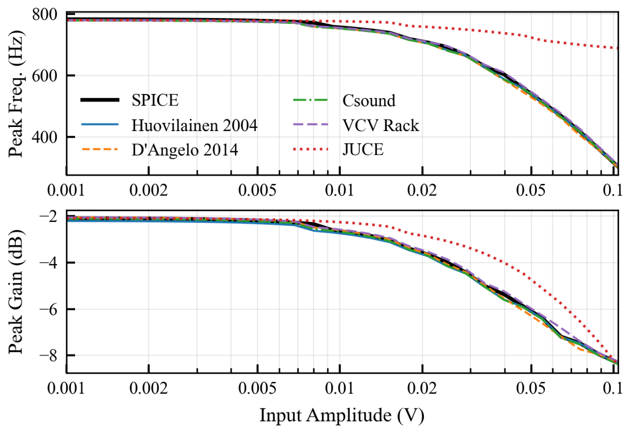


Figure 4: Large-signal cutoff shift for SPICE, Huovilainen 2004, D'Angelo 2014, Csound, VCV Rack, and JUCE. Peak frequency and peak gain are tracked as input level increases at cutoff 1000 Hz, $k = 2$, under steady-state sinusoidal excitation swept around the 1000 Hz resonance region (Table 2).

4.3. Large-Signal Cutoff Shift

Fig. 4 measures how the apparent resonance peak moves as input level increases, making the operating-point dependence more explicit than the linear overview alone. Huovilainen 2004, D'Angelo 2014, Csound, and VCV Rack track the SPICE baseline within roughly 20 cents and 0.11 dB in peak frequency and peak gain, whereas JUCE departs by 427 cents and 0.70 dB. The figure therefore serves as an intermediate view between the linear overview and the later harmonic plots: it isolates large-signal peak motion before turning to full harmonic structure.

4.4. Nonlinear Harmonic Structure

Figs. 5–7 establish the cross-implementation comparison in the nonlinear domain, showing the divergence from three angles: static harmonic structure, structure during a cutoff sweep, and harmonic growth during a drive sweep. Across all three views, Huovilainen 2004, D'Angelo 2014, Csound, and VCV Rack remain closer to SPICE than JUCE: implementations that share a broad linear identity can nonetheless occupy very different nonlinear operating regimes.

Table 3 makes the same ordering explicit. The four 1+4 stage implementations cluster closely in the spectra metric at roughly 0.04 % to 0.10 % of H1, while JUCE departs more clearly at

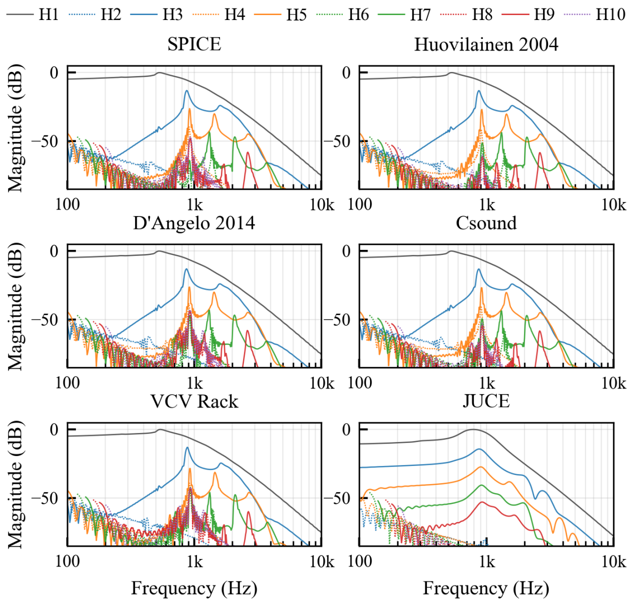


Figure 5: Nonlinear harmonic structure versus output frequency for the comparison set: SPICE, Huovilainen 2004, D'Angelo 2014, Csound, VCV Rack, and JUCE. Harmonic levels are measured relative to the fundamental at cutoff 1000 Hz, $k = 4$, input 0.1 V, 384 kHz (Table 2).

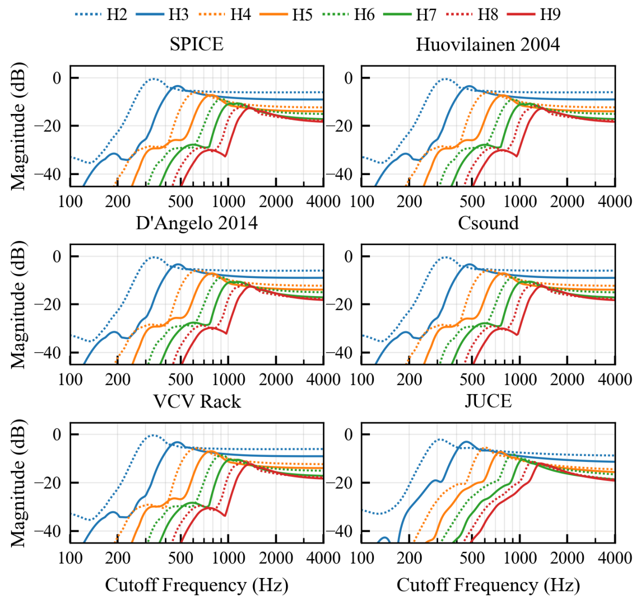


Figure 6: Nonlinear harmonic structure during a cutoff sweep for SPICE, Huovilainen 2004, D'Angelo 2014, Csound, VCV Rack, and JUCE. 130.813 Hz sawtooth (approx. C3), $k = 3.5$, input 0.1 V; shows harmonic levels relative to the fundamental across the cutoff sweep (Table 2).

0.65 %. Within that clustered group, Huovilainen 2004 is the closest overall and VCV Rack the farthest, but both remain far below JUCE in all three harmonic summaries. The per-harmonic breakdown shows that this separation within the comparison set is already visible in the low- to mid-order components, especially H3 and H5, rather than being driven only by the weakest high-order terms.

The static spectra (Fig. 5) are the most compact entry point, the cutoff-sweep view (Fig. 6) shows how harmonic balance reorganizes as the cutoff moves through the band, and the drive-sweep view (Fig. 7) separates onset level from harmonic-growth rate.

4.5. Controlled Structure/Function Ablation

The broader cross-implementation comparison shows strong nonlinear separation but does not by itself explain which design choices produce it. We therefore turn to a controlled ablation on the Zavalishin 2012 core (Figs. 8a, 8b, and 9) to convert that separation into structural explanation and design insight. Here, "Zavalishin-style ladder core" denotes a common topology-preserving-transform (TPT), zero-delay-feedback (ZDF) ladder structure whose state-update and feedback topology are fixed while only saturator type and nonlinearity placement are changed. Because the implicit equation is solved by Newton–Raphson, changing the saturator requires only a local change to the nonlinear function and its derivative. As in the cross-implementation comparison, both panels remain SPICE-referenced and use the same 96 kHz/384 kHz split as the earlier tests. The distributed `tanh` baseline is compared against two distributed saturator substitutions (`atan`, soft clip) and two boundary variants ($\tanh(\text{in}) + \tanh(\text{FB})$, $\tanh(\text{in} + \text{FB})$); partial stage-linearization variants are then examined separately in Fig. 9. In this benchmark the baseline `tanh` stage is $2V_T \tanh(v/2V_T)$; the `atan` stage is $\frac{4V_T}{\pi}\arctan\left(\frac{\pi}{2}\frac{v}{2V_T}\right)$ and the soft clip is $2V_T\, u/(1 + |u|)$ with $u = v/2V_T$. All three preserve unit slope at the origin and the same $\pm 2V_T$ asymptotic limits, differing only in curvature between those limits.

Figs. 8a and 8b turn that separation into a structural explanation. The linear overview first shows that the distributed `atan` substitution leaves the small-signal identity largely intact, whereas the distributed `soft clip` already reduces the resonance peak height more visibly while leaving the passband gain largely unchanged. This is expected rather than contradictory: the linear overview is a finite-amplitude swept-sine measurement (0.01 V), so near the high-Q resonance the signal is amplified into the curved region of each saturator. With identical origin slopes, the earlier-bending soft clip therefore lowers the resonant peak first. In the harmonic spectra view, the $\tanh(\text{in}) + \tanh(\text{FB})$ and $\tanh(\text{in} + \text{FB})$ variants sit clearly above the distributed `tanh` baseline, indicating that collapsing the nonlinearity out of the four ladder sections materially changes the measured response. The distributed `soft clip` variant is the strongest departure overall, while the distributed `atan` variant shows that changing saturator function alone can also shift the harmonic balance substantially even when distributed placement is retained.

The saturator-substitution and boundary-placement variants vary saturator type or collapse nonlinearity to the filter boundaries while keeping the common core fixed. By contrast, the partial-ablation rows in Table 4 retain `tanh` and vary which ladder stages are linearized. Fig. 9 isolates those ten partial-ablation variants and makes the stage-removal pattern directly readable from the spectra.

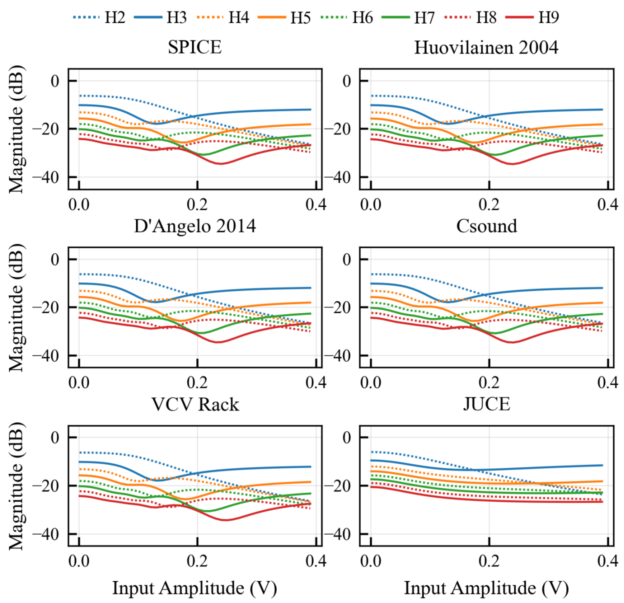


Figure 7: Nonlinear harmonic structure during a drive sweep for the comparison set: SPICE, Huovilainen 2004, D'Angelo 2014, Csound, VCV Rack, and JUCE. 65.41 Hz sawtooth (approx. C2), cutoff 1000 Hz, $k = 0$; shows harmonic growth across the drive sweep (Table 2).

Table 4 summarizes the same effect numerically across the same four views used for the comparison set. In 'Spectra', the unablated distributed `tanh` core (the within-core reference, against which each variant is read in addition to SPICE) sits at only 0.05 % of H1. Among the saturator-substitution and boundary-placement variants, every alternative rises to at least 0.29 %; among the partial-ablation variants, the values range from 0.17 % to 0.34 %. Collapsing nonlinearity to the filter boundaries raises the harmonic-domain summaries well above that baseline, with the $\tanh(\text{in}) + \tanh(\text{FB})$ and $\tanh(\text{in} + \text{FB})$ variants remaining close to each other across all three nonlinear views. The distributed `atan` variant sits above both boundary variants in 'Spectra' and 'Cutoff' but below them in 'Drive,' showing that saturator function and placement do not perturb every operating view to the same degree. The distributed `soft clip` variant departs further still, especially in 'Cutoff' (7.04 %). Fig. 9 extends that explanation into a stage-level design result: when only part of the ladder can remain nonlinear, the placement of the remaining nonlinear stages matters systematically.

Saturator choice and placement act as partly independent levers (Table 4): the distributed `soft clip` variant produces larger departures than either boundary variant despite retaining distributed placement. Changing the saturator is therefore not equivalent to moving the nonlinearity to the filter boundaries when the goal is to preserve ladder-like nonlinear behavior. This structural interpretation is established by the common-core ablation itself; the cross-implementation comparison is then useful as a check that the same pattern appears in practical implementations.

Table 3: Compact quantitative summary of the comparison set. Lower is better in every column. 'Linear' reports $\sqrt{\text{bias}^2 + \sigma^2}$ of the magnitude deviation from SPICE across 20 Hz–20 kHz in dB, while 'Spectra,' 'Cutoff,' and 'Drive' report average H2–H9 RMSE in H1-normalized linear amplitude, shown here as percent of H1 (the spectra figure extends to H10; H2–H9 is used here for consistency across all harmonic views).

Model	Linear (dB)	Spectra (% H1)	Cutoff (% H1)	Drive (% H1)
Huovilainen 2004	0.116	0.04	0.17	0.05
D'Angelo 2014	0.363	0.06	0.22	0.18
Csound	0.403	0.06	0.19	0.09
VCV Rack	0.162	0.10	0.33	0.24
JUCE	0.304	0.65	3.39	3.03

Table 4: Compact quantitative summary of the structure/function ablation set, using the same summary metrics as Table 3. Lower is better in every column. "1+4 stage" denotes one input stage plus four ladder sections; the final ten rows list all partial-ablation variants in terms of which ladder stages were linearized.

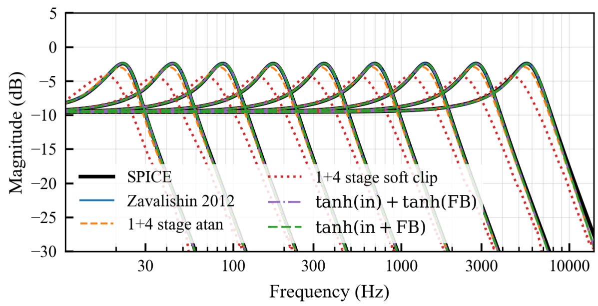
Variant	Linear (dB)	Spectra (% H1)	Cutoff (% H1)	Drive (% H1)
1+4 stage `tanh`	0.344	0.05	0.14	0.06
1+4 stage `atan`	0.483	0.59	3.07	1.33
1+4 stage `soft clip`	1.936	0.96	7.04	1.98
$\tanh(\text{in}) + \tanh(\text{FB})$	0.338	0.29	1.98	1.71
$\tanh(\text{in} + \text{FB})$	0.338	0.38	2.17	1.71
stages 1, 2 linearized	0.340	0.26	1.74	1.90
stages 1, 3 linearized	0.341	0.26	1.51	1.87
stages 1, 4 linearized	0.339	0.17	1.16	1.50
stages 2, 3 linearized	0.342	0.25	1.17	1.87
stages 2, 4 linearized	0.344	0.18	1.09	1.48
stages 3, 4 linearized	0.344	0.20	1.11	1.49
stages 1, 2, 3 linearized	0.341	0.34	2.33	2.92
stages 1, 2, 4 linearized	0.340	0.28	1.68	1.60
stages 1, 3, 4 linearized	0.340	0.29	1.56	1.56
stages 2, 3, 4 linearized	0.341	0.30	1.43	1.55

5. DISCUSSION AND LIMITATIONS

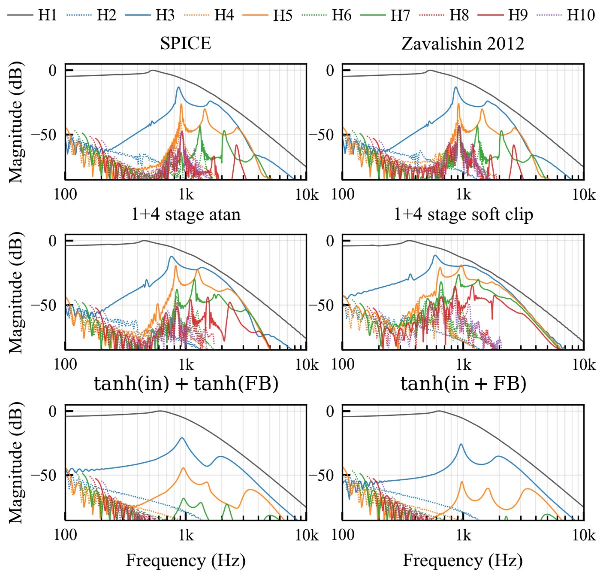
The shared SPICE reference places linear and nonlinear agreement on the same quantitative footing. It thereby gives an analog-referenced basis for judging which aspects of behavior a given implementation or simplification preserves and which it loses.

The boundary variants illustrate a broader caution: a boundary-saturation design can look acceptable linearly and in averaged summaries while still missing the analog-referenced nonlinear operating behavior visible in the harmonic-growth trajectories.

A similar caution applies when the nonlinear budget is reduced. Starting from the full 1+4 stage `tanh` ladder, we ablated selected ladder-stage nonlinearities while retaining the nonlinear input, and compared which stage removals remained closest to SPICE (Table 4, Fig. 9). When two ladder-stage nonlinearities were ablated, removing the later stages (i.e., retaining the earlier ones) often stayed closest to SPICE, whereas removing the earlier



(a) Linear frequency response.



(b) Nonlinear harmonic spectra.

Figure 8: Controlled structure/function ablation on a common Zavalishin-style ladder core, comparing 1+4 stage `tanh`, 1+4 stage `atan`, 1+4 stage `soft clip`, $\tanh(\text{in}) + \tanh(\text{FB})$, and $\tanh(\text{in} + \text{FB})$; SPICE in black. Upper panel: linear response (96 kHz); lower panel: nonlinear harmonic spectra (384 kHz, $k = 4$).

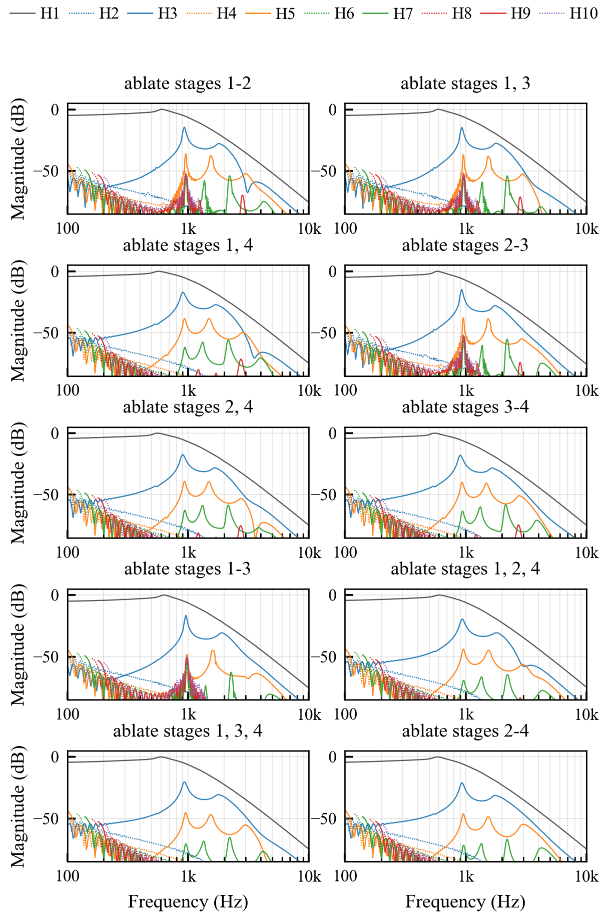


Figure 9: Static nonlinear harmonic spectra for the ten partial stage-linearization variants of the common Zavalishin-style ladder core. Each panel keeps the nonlinear input stage and linearizes only the listed ladder sections; harmonic levels are measured relative to the fundamental at cutoff 1000 Hz, $k = 4$, input 0.1 V, 384 kHz.

stages diverged more in ‘Spectra’, ‘Cutoff’, and ‘Drive’. The same tendency is visible directly in Fig. 9 and persists when three stages are linearized: retaining earlier nonlinear stages stays closer to the low-error end of the set, whereas linearizing the earliest stages pushes the low- to mid-order harmonics higher. This is consistent with the earlier stages seeing larger signal excursions before the low-pass sections attenuate the signal, so their nonlinear contribution to the overall harmonic balance is greater. For simplified ladders, the implication is direct: if only part of the ladder can remain nonlinear, preserving earlier-stage nonlinearity is usually a better approximation than preserving later-stage nonlinearity alone, and several such variants match or improve on the boundary-saturation summaries at comparable nonlinear count. All such variants remained similar in the self-oscillation check. A systematic analysis of optimal stage-removal strategies under explicit resource constraints is left as future work.

All four 1+4 stage implementations cluster together because they preserve distributed nonlinearity despite different saturator approximations and solver structures (VCV Rack, for example, stays in this cluster); within this benchmark, preserving distributed placement is at least as important as matching the exact saturator law.

JUCE’s boundary-saturation topology provides a practical example of the same pattern isolated by the ablation. Its nonlinear departure is used here as a structural contrast rather than as a standalone judgment of JUCE itself. The $\tanh(\text{in}) + \tanh(\text{FB})$ Zavalishin variant, which mirrors JUCE’s topology, shows comparable departure across all three harmonic views (Table 4).

The SPICE reference is still only a simulation baseline and has not been validated against hardware measurements; for modeling and measurement of a physical Moog VCF, see [14]. It models the filter core only; peripheral circuitry, component tolerances, transistor mismatch, thermal drift of V_T, and parasitic capacitances are all absent, and these can shift cutoff frequency, self-oscillation threshold, and nonlinear character in real hardware. The boundary normalizations align only external conventions, leaving each model's internal algorithms unchanged. The study focuses on static and quasi-static conditions; aliasing is not isolated as a separate variable, since its relative ranking can depend strongly on sampling rate and oversampling strategy. Within those limits, the benchmark checks whether a digital ladder implementation tracks a common analog-style baseline across musically relevant conditions and separates structural nonlinear differences from control-range mismatches. It does not, however, address perceptual preference, audio-rate modulation, or optimization under explicit CPU or polyphony constraints. Instead, it provides a reproducible evaluation framework for such future work.

The H2–H9 RMSE summaries are therefore best read as comparative error measures, not direct audibility thresholds; prior perceptual work likewise shows that simple THD-style measures do not map reliably onto perceived distortion [15, 16, 17, 18]. Tightening that link remains a natural next step, together with a hardware spot-check to anchor the absolute scale, a second ablation core to test whether the stage-importance ordering generalizes beyond this TPT/ZDF structure, perceptually weighted summaries, audio-rate modulation tests, and listening studies.

6. CONCLUSION

A reproducible SPICE-referenced evaluation framework was applied to five practical ladder-filter implementations and a controlled common-core ablation. The cross-implementation comparison confirms that close linear agreement does not predict nonlinear fidelity, while the ablation shows that harmonic response depends on saturator function, nonlinearity placement, and which ladder stages remain nonlinear. In the partial ablations, retaining earlier-stage nonlinearities preserved the SPICE-referenced harmonic structure better than retaining later-stage nonlinearities alone. Three practical rules follow: boundary saturation can diverge strongly in nonlinear behavior despite competitive linear response, nonlinear-budget reduction is stage-dependent rather than interchangeable, and preserving distributed placement can matter as much as matching the exact saturator. Nonlinear tests therefore belong in the core evaluation set when implementations differ in nonlinearity placement or saturator count, and simplified ladder designs should treat nonlinear placement as a primary design variable.

The benchmark, evaluation pipeline, and audio samples are archived at `https://doi.org/10.5281/zenodo.20735847` (source: `https://github.com/oyama/ladder-filter-testbench`).

7. REFERENCES

[1] V. Välimäki and A. Huovilainen, "Oscillator and filter algorithms for virtual analog synthesis," *Computer Music Journal*, vol. 30, no. 2, pp. 19–31, 2006.

[2] J. Pakarinen and D. T. Yeh, "A review of digital techniques for modeling vacuum-tube guitar amplifiers," *Computer Music Journal*, vol. 33, no. 2, pp. 85–100, 2009.

[3] R. A. Moog, "A voltage-controlled low-pass high-pass filter for audio signal processing," in *17th AES Convention*, New York, USA, 1965.

[4] D. Rossum, "Making digital filters sound "analog"," in *Proceedings of the International Computer Music Conference (ICMC)*, San Jose, CA, USA, 1992, pp. 30–34.

[5] T. S. Stilson and J. O. Smith, "Analyzing the Moog VCF with considerations for digital implementation," in *Proceedings of the International Computer Music Conference*, Hong Kong, 1996.

[6] A. Huovilainen, "Non-linear digital implementation of the Moog ladder filter," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Naples, Italy, 2004, pp. 61–64.

[7] F. Fontana, "Preserving the structure of the Moog VCF in the digital domain," in *Proceedings of the International Computer Music Conference (ICMC)*, Copenhagen, Denmark, 2007.

[8] V. Zavalishin, "The art of VA filter design," 2012, Native Instruments GmbH white paper, https://www.native-instruments.com/fileadmin/ni_media/downloads/pdf/VAFilterDesign_2.1.2.pdf.

[9] S. D'Angelo and V. Välimäki, "Generalized Moog ladder filter: Part II—explicit nonlinear model through a novel delay-free loop implementation method," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 1873–1883, 2014.

[10] D. T. Yeh, J. S. Abel, and J. O. Smith, "Simplified, physically-informed models of distortion and overdrive guitar effects pedals," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Bordeaux, France, 2007, pp. 189–196.

[11] B. Holmes and M. van Walstijn, "Physical model parameter optimisation for calibrated emulation of the dallas rangemaster treble booster guitar pedal," in *Proceedings of the 19th International Conference on Digital Audio Effects (DAFx)*, Brno, Czech Republic, 2016, pp. 47–54.

[12] K. J. Werner, V. Nangia, A. Bernardini, J. O. Smith, and A. Sarti, "An improved and generalized diode clipper model for wave digital filters," in *139th AES Convention*, New York, USA, 2015.

[13] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *108th AES Convention*, Paris, France, 2000, AES Paper 5093.

[14] E. M. Paschou, F. Esqueda, V. Välimäki, and J. Mourjopoulos, "Modeling and measuring a Moog voltage-controlled filter," in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Kuala Lumpur, Malaysia, 2017.

[15] E. R. Geddes and L. W. Lee, "Auditory perception of nonlinear distortion," in *115th AES Convention*, New York, USA, 2003, AES Paper 5891.

[16] S. Temme, P. Brunet, and P. Qarabaqi, "Measurement of harmonic distortion audibility using a simplified psychoacoustic model," in *133rd AES Convention*, San Francisco, USA, 2012.

[17] J. Breebaart, "Evaluation of statistical inference tests applied to subjective audio quality data with small sample size," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 5, pp. 887–897, 2015.

[18] S. Zieliński, F. Rumsey, and S. Bech, "On some biases encountered in modern audio quality listening tests—A review," *Journal of the Audio Engineering Society*, vol. 56, no. 6, pp. 427–451, 2008.