telecom

MDPI

Article

Channel Estimation for Active Reconfigurable Intelligent Surfaces Under Amplifier Compression

Abdalrahman Basheer Ibrahim * and Ibrahim Khider Eltahir

Department of Electronics Engineering, Faculty of Engineering, Sudan University of Science and Technology (SUST), Khartoum 00407, Sudan; ibrahim_khider@hotmail.com
* Correspondence: ahmidi2022@gmail.com

Abstract

Active reconfigurable intelligent surfaces amplify the reflected signal using amplifiers integrated into their elements. The signal model adopted throughout the literature treats each amplifier as a constant gain applied jointly to the incident signal and to the thermal noise of the element. Its experimental validation was performed at a single, very low incident power, far below the level at which an amplifier compresses. This article examines channel estimation when the amplifier is driven near saturation. Adopting an established nonlinear surface model, we show that the equivalent gain depends on the pilot power, so the closed-form training results of the linear model no longer hold: the conventional estimator becomes biased, the published training rules lose their optimality, and the estimation error exhibits a floor. We split the amplifier distortion into a static per-element component and a fluctuating component, and derive closed-form expressions for the error variance, the estimator bias, and the resulting floors, together with a Gaussian-equivalent variance bound. The expressions agree with Monte Carlo simulation to within 0.25 dB over twelve configurations. A numerical Bayes-MMSE benchmark that uses the exact Rapp response is at most 1.3 dB better than the Bussgang-based linear estimator, so the linearization costs little. Three consequences follow: the conventional estimator degrades once the pilot power passes a threshold; additional receive antennas, surface elements and training slots do not remove the floor in the simulated regime, because its dominant component is static, and this holds for rank-one, Rician and Rayleigh backward channels alike; and the published rule for the training amplification points the wrong way unless the receiver noise is high or the amplifier headroom is large. We then propose an online gain-tracking scheme that identifies the operating back-off from the ratio of two estimates taken at different pilot powers, needing no baseband processing at the surface, and we quantify its sensitivity to finite samples, channel aging, parameter drift and shape mismatch. Finally, we characterize robustness across amplifier models, element spread, and system scale and identify the region in which compensation helps.

Keywords: active reconfigurable intelligent surface; channel estimation; amplifier compression; Bussgang decomposition; Cramér–Rao bound; gain tracking; 6G

Check for updates

Academic Editor: Minseok Kim

Received: 8 September 2026
Revised: 22 October 2026
Accepted: 26 October 2026
Published: 2 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Active RISs overcome the multiplicative-fading bottleneck of passive surfaces by integrating a reflection-type amplifier into each element, at the cost of non-negligible amplified thermal noise [1,2]. Throughout this paper, *active* denotes reflection-type amplification; it

does not refer to surfaces equipped with a few receive RF chains for sensing, for which the same adjective is often used.

The signal model on which the active-RIS literature rests treats the amplification matrix as a constant real gain. Its experimental validation was carried out with the network-analyzer input set to $P_x = -50$ dBm [3]. That measurement confirms that the noise power grows linearly with the reflection gain. It does not, however, test whether the gain remains constant as the *incident* power grows. A varactor-based parametric reflection amplifier saturates like any other amplifier [4], and third-order intermodulation has been measured directly on RIS elements with embedded varactor diodes [5,6], so the nonlinearity is documented at the component level rather than merely presumed. The consequences of that saturation for the radiated field have been characterized in [7,8]; its consequences for channel estimation have not, to our knowledge, been examined analytically for active RIS. Digital predistortion, used to linearize large transmit arrays [9], is unavailable here: a reflective surface has no baseband chain in which to apply it, so the distortion must be accounted for in the estimator rather than removed at the source.

1.1. Related Work

Channel estimation for active RIS has been studied under the linear model, using either an optimized training reflection pattern [10] or a convolutional estimator [11]. Data-driven estimators have also been reported under hardware impairments, for RIS-assisted full-duplex MIMO [12] and for massive-MIMO SEFDM [13]. As described in its abstract, the former learns the impairment-induced distortion from data without an explicit model, so it does not rely on an explicit amplifier model such as the one studied here; the latter involves neither a reflecting surface nor amplifier compression. An LS estimator attaining the Cramér–Rao lower bound (CRLB) has also been derived for an amplifying surface [14]. All of these assume a constant amplifier gain.

Three related impairments must be distinguished. First, the self-interference of [1] is a nonlinearity *in the reflection matrix*, caused by imperfect inter-element isolation, and is independent of the signal level; the effect studied here is a function of the incident amplitude and vanishes at low drive. Second, the framework of [15] addresses non-i.i.d. Gaussian perturbations and non-uniform phase variation, not amplifier compression. As described in its abstract, that framework treats spatially correlated Gaussian channel perturbations and non-uniform phase variation, and it does not describe a saturating amplifier or a gain that depends on the incident power; its full formulation was not available to us, so we make no claim about its equations. In the present problem the surface output is the memoryless nonlinear map $g(u)$ of Equation (2), whose effective gain $\kappa(P_{\mathrm{T}})$ and distortion variance $\sigma_d^2(P_{\mathrm{T}})$ depend on the incident power, so that $\partial \mathbf{\Phi}_k / \partial P_{\mathrm{T}} \neq 0$ in Equation (5). Third, amplifier nonlinearity in RIS-assisted systems has been treated for beamforming and spectral efficiency [16,17], but under perfect or separately acquired CSI; to our knowledge, the estimation stage for a saturating amplifier with a pilot-power-dependent gain has not been revisited analytically. The Bussgang linear MMSE (BLMMSE) estimator is an established benchmark for channel estimation under amplifier nonlinearity in massive MIMO [18] and is included among our baselines.

Two groups of studies must be kept apart. Estimators for nonlinear hardware in massive MIMO, such as [18], address a link that contains no reflecting surface. Here the nonlinearity sits inside the cascaded channel, between the user–RIS and RIS–BS links, so the quantity to be estimated enters the amplifier input and the operating point, hence the effective gain, is unknown at the receiver. Conversely, the active-RIS estimators [10,11,14] assume a linear amplifier. The claim of this paper is therefore limited to channel estimation

for an active RIS whose elements saturate; we do not claim to be the first to estimate channels under amplifier nonlinearity in general.

Table 1 positions this work against the closest prior art. The distinguishing feature is not the nonlinear model, which is adopted from [7], but the fact that, to our knowledge, every published active-RIS channel estimator assumes the amplifier gain to be a known constant, whereas the nonlinear estimators for other hardware, such as [18], do not involve a reflecting surface.

Table 1. Positioning against the most closely related work.

Work	Amplifier Model	Ch. Est.	Bound	Limitation Addressed Here
Chen et al. [10]	Linear, constant gain	Yes	Variance	Training rule derived under linearity
Wang et al. [11]	Linear, constant gain	Yes	None	No bound; labels taken from LS
Pan et al. [14]	Linear, amplifying surface	Yes	CRLB	Bound not valid under compression
Kolomvakis et al. [7]	Third-order polynomial	No	None	Estimation stage never revisited
Sharini et al. [15]	Not an amplifier model (perturbations, phase)	Yes	—	Different impairment (perturbations, phase)
Arjas and Atzeni [18]	Nonlinear, massive MIMO	Yes	None	Not a reflecting surface
Lawal et al. [12]	Hardware impairments, learned from data	Yes	—	No explicit amplifier model
This work	**Saturating (Rapp, Saleh)**	**Yes**	**Gauss.-equiv.; Bayes-MMSE**	**—**

The first three entries assume the amplifier gain to be a known constant. The remaining entries address distortion, perturbations or hardware impairments in other settings. This work tracks the gain of a saturating active RIS online. The last row, in bold, is the present work.

1.2. Contributions

- **C1:** We show that, under compression, the training observation matrix becomes a function of the pilot power, so the results of the linear model that hold for every pilot power do not carry over: conventional LS is biased by the factor κ/p, and the closed-form training optima of the linear analysis become biased.
- **C2:** We derive closed forms for the error variance, the bias, and both high-power floors, together with a Gaussian-equivalent variance bound that extends the CRLB known for amplifying surfaces [14]. The distortion is split into a static per-element component and a fluctuating component. The static component, which dominates in the simulated regime, is exactly independent of the number of base-station antennas, RIS elements, training slots and of the backward-channel geometry, so none of these mitigates it. A numerical Bayes-MMSE benchmark with the exact Rapp response shows that the Bussgang-based estimators lose at most 1.3 dB.
- **C3:** We show that the training-amplification rule obtained under the linear model is not merely suboptimal but *directionally wrong* once the amplifier compresses, unless the receiver noise is high or the headroom is large; a numerical map delimits where the rule reverses.
- **C4:** We propose an online gain-tracking scheme based on dual-power probing. It identifies the operating back-off from the ratio of two channel estimates acquired at different pilot powers. It requires neither the saturation level nor the channel gain to be known, adds one probe frame, and needs no baseband processing at the surface. We establish its well-posedness and quantify its sensitivity to finite samples, channel aging, drift and shape mismatch.

- **C5:** We characterize robustness across amplifier models, element spread, and system scale, and identify the operating region in which compensation is beneficial. All results are reported against input back-off, so they do not depend on the absolute saturation level of any particular device.

1.3. Scope

This work is analytical and simulation-based. Its premise is that a widely used model was validated only in the small-signal regime; the same standard applies here, and the present results are therefore predictions of a component-level behavioral model rather than measurements of a fabricated surface. Experimental validation on an active-RIS prototype is left as necessary future work.

Practical validity of the model. The Rapp and Saleh responses are memoryless, per-element behavioral models. They capture AM/AM compression and, for Saleh, AM/PM conversion, and Section 5 quantifies both. They do not capture mutual coupling, which makes the effective input of an element depend on the state of its neighbors, nor bias-dependent nonlinearity, in which the saturation level and the small-signal gain move with the DC operating point. Slow drift of the saturation level or of the gain is covered by the tracker of Section 4, which identifies the operating back-off rather than the device parameters; a coupling- or bias-induced change in the *shape* of the response is covered only to the extent of the smoothness-mismatch study of Section 5. We did not find published AM/AM curves of active-RIS elements against which the parameters could be fitted: the prototypes we reviewed report gain or received-power improvement at fixed drive [19] or analyze the element in its linear domain [20], without compression characteristics. We therefore express every result against input back-off. For orientation, the simulated element has 20 dB small-signal gain, an input 1-dB compression point of −41.2 dBm, an output 1-dB compression point of −22.2 dBm and a saturated output of −20 dBm.

2. System Model

An N-element active RIS assists uplink transmission from a single-antenna user to a K-antenna base station (BS) over T training slots. Let $\mathbf{b} \in \mathbb{C}^N$ be the user–RIS channel, $\mathbf{g}_k \in \mathbb{C}^N$ the RIS–BS channel seen by antenna k with $\mathbf{G}_k = \mathrm{diag}(\mathbf{g}_k)$, and $h_{\mathrm{d},k}$ the direct link. Under the conventional model the surface output is

$$\mathbf{y} = \mathbf{P}\boldsymbol{\Theta}\mathbf{x} + \mathbf{P}\boldsymbol{\Theta}\mathbf{v}, \tag{1}$$

with $\mathbf{P} = \mathrm{diag}(p_1, \ldots, p_N)$ real, $\boldsymbol{\Theta}$ unit-modulus diagonal, and $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \sigma_1^2 \mathbf{I}_N)$ the amplified thermal noise. Static noise is neglected, consistent with the reported measurements [3].

2.1. Amplifier Response

During training the signal entering the amplifier of element n is $u_n = \sqrt{P_\mathrm{T}}\, b_n s + v_n$, where s is the pilot symbol; signal and noise are amplified jointly. We model the element by the soft-limiting response [21]

$$g_n(u) = \frac{p_n u}{\left[1 + (p_n |u| / A_{\mathrm{sat}})^{2}\right]^{1/(2)}}, \tag{2}$$

where A_{sat} is the saturation amplitude and controls the sharpness of the transition. As $A_{\mathrm{sat}} \to \infty$, (2) reduces exactly to (1), so the conventional model is a special case rather than a competing alternative. A saturating response is required: the third-order polynomial commonly used to characterize the radiated distortion [7,22] is a truncated expansion that diverges rather than saturating and is therefore not valid in the compression regime studied here. An AM/PM model is treated in Section 5.

2.2. Bussgang Decomposition

For zero-mean u_n [23,24],

$$g_n(u_n) = \kappa_n u_n + d_n, \qquad \mathbb{E}\{u_n^* d_n\} = 0, \tag{3}$$

$$\kappa_n = \frac{\mathbb{E}\{u_n^* g_n(u_n)\}}{\mathbb{E}\{|u_n|^2\}}, \quad \sigma_{d,n}^2 = \mathbb{E}\{|g_n(u_n)|^2\} - |\kappa_n|^2 \sigma_u^2. \tag{4}$$

Here $\sigma_u^2 = \mathbb{E}\{|u_n|^2\}$, so both κ_n and $\sigma_{d,n}^2$ depend on the transmit power. Assuming identical elements we drop the index n; the non-identical case is treated in Section 5.

Stochastic model. Two roles of $\mathbf{b}$ must be separated. During one training block $\mathbf{b}$ is a fixed unknown vector: it is the quantity to be estimated. The Bussgang quantities κ and σ_d^2 are instead *ensemble* moments. They average over the channel prior $b_n \sim \mathcal{CN}(0, \varrho_b)$ and over the thermal noise, so that $u_n \sim \mathcal{CN}(0, \sigma_u^2)$ with $\sigma_u^2 = P_\mathrm{T}\varrho_b + \sigma_1^2$, and the estimation error is averaged over the same prior. This is the standard use of the decomposition [24]. It is applied consistently in Sections 2 and 3 and in the simulations, which draw a fresh $\mathbf{b}$ per realization and keep it fixed over the T training slots.

Static and fluctuating distortion. Because $\mathbf{b}$ and the pilot amplitude are constant during training, the distortion splits as $d_n(t) = \bar{d}_n + \tilde{d}_n(t)$. Here $\bar{d}_n = \mathbb{E}\{d_n \,|\, b_n\}$ is a deterministic function of b_n, static over the T slots, and $\tilde{d}_n(t)$ is the part driven by the thermal noise $v_n(t)$. Define $\sigma_s^2 = \mathbb{E}\{|\bar{d}_n|^2\}$ and $\sigma_f^2 = |\kappa|^2\sigma_1^2 + \mathbb{E}\{|\tilde{d}_n|^2\}$, so that $|\kappa|^2\sigma_1^2 + \sigma_d^2 = \sigma_s^2 + \sigma_f^2$; both follow from Equation (2) by quadrature or Monte Carlo. Stein's lemma gives $\mathbb{E}\{b_n^* d_n\} = 0$: the static part is uncorrelated with the channel, as the Bussgang decomposition requires, although it is not independent of it. When $\sigma_1^2 \ll P_\mathrm{T}\varrho_b$, as in Table 2, nearly all of the distortion is static ($\sigma_f^2/\sigma_s^2 \approx -30$ dB).

Table 2. Simulation parameters.

Parameter	Symbol	Value
RIS elements	N	16
BS antennas	K	16
Training slots	T	$N+1$
Small-signal gain	p	10 (20 dB)
Saturation amplitude	A_sat	$0.10\ \sqrt{\mathrm{mW}}$
Rapp smoothness		2
User–RIS gain	ϱ_b	10^{-6}
RIS–BS gain	ϱ_g	$10^{-3}/50^2$
Direct-link gain	ϱ_d	10^{-7}
RIS thermal noise	σ_1^2	−70 dBm
BS noise	σ_2^2	−80 dBm
Probe ratio	ϱ	10 dB

3. Channel Estimation Under Compression

3.1. Collapse of the Linear Model

Substituting (3) into the training observation and stacking over T slots gives

$$\mathbf{y}_k = \sqrt{P_\mathrm{T}}\,\mathbf{S}\,\boldsymbol{\Phi}_k(P_\mathrm{T})\,\mathbf{h}_k + \tilde{\mathbf{z}}_k, \tag{5}$$

with $\mathbf{h}_k = [h_{\mathrm{d},k}, \mathbf{b}^\mathrm{T}]^\mathrm{T}$, $\mathbf{S} = \mathrm{diag}(s(1), \ldots, s(T))$ the pilot matrix, and an effective noise comprising the amplified thermal noise, the distortion and the receiver noise.

Proposition 1. *The observation matrix $\boldsymbol{\Phi}_k(P_\mathrm{T})$ in (5) depends on the pilot power through κ. For a fixed pilot power and a known κ, (5) is linear in $\mathbf{h}_k$, and the compression-aware LS estimator, which*

uses $\mathbf{\Phi}_k(P_T)$, is the correctly specified estimator [25]. The conventional LS estimator, which assumes $\kappa = p$, is misspecified and returns κ/p times the compression-aware one. Because $\mathbf{\Phi}_k$ depends on the design variable, the closed-form optimal training scaling derived under the linear model [10] is no longer optimal, and the family of linear-model results indexed by P_T does not carry over.

Remark 1. *The central comparison of this paper is therefore not between LS and a better estimator of the same model, but between a conventional estimator that ignores the dependence of $\mathbf{\Phi}_k$ on P_T and a compression-aware estimator that requires $\kappa(P_T)$, that is, the operating point. Section 4 addresses how that operating point is obtained.*

3.2. Training Design

Two requirements are worth stating explicitly. The RIS patterns must use columns $1, \dots, N$ of the T-point DFT, with the DC column reserved for the direct link. Including the DC column in the RIS block would make it collinear with the direct-link column, rendering the LS problem singular. The RIS–BS link is taken to be line-of-sight with constant modulus per element, $|g_{k,n}|^2 = \varrho_g$, since a Rayleigh backward link produces near-null columns that destroy conditioning. With these choices the Gram matrix is diagonal,

$$\mathbf{\Phi}_k^H \mathbf{\Phi}_k = T\,\mathrm{diag}\big(1, |\kappa|^2\varrho_g, \dots, |\kappa|^2\varrho_g\big). \tag{6}$$

3.3. Error Variance and Bound

Let $\sigma_{\mathrm{eff}}^2 = N\varrho_g\,\sigma_f^2 + \sigma_2^2$ denote the power per slot of the fluctuating disturbance plus receiver noise. For a given κ, consider the surrogate linear-Gaussian model in which the fluctuating disturbance is Gaussian and independent of $\mathbf{b}$. Its Fisher information is $\mathcal{J}(\mathbf{h}_k) = (P_T/\sigma_{\mathrm{eff}}^2)\mathbf{\Phi}_k^H\mathbf{\Phi}_k$, which with (6) gives

$$\mathrm{CRLB}(b_n) = \frac{\sigma_{\mathrm{eff}}^2}{P_T T|\kappa|^2\varrho_g}. \tag{7}$$

Remark 2. *Equation (7) is a Gaussian-equivalent variance bound, not the exact CRLB of the nonlinear observation model. Three points qualify it. (i) The fluctuating disturbance is not Gaussian. Since Gaussian noise minimizes the Fisher information for a given variance, (7) is conservative with respect to that component. (ii) Its variance depends on the unknown channel through the local slope of g, because a strongly driven element fluctuates less; σ_f^2 is therefore an ensemble average, and the exact Fisher information differs from element to element. (iii) The static component $\bar{d}_n$ is a deterministic function of b_n and appears as a bias, which a bound for unbiased estimators does not cover. For a random channel the appropriate lower bound is the Bayesian MMSE, which Section 6 evaluates numerically with the exact Rapp response. Equation (7) should be read as a benchmark for the variance of the fluctuating component, and the gap to the Bayes-MMSE measures what the Bussgang linearization costs.*

Combining across the K BS antennas (Appendix A) yields, for the rank-one line-of-sight backward link, the per-element error variance

$$\varepsilon_b = \frac{\sigma_s^2 + (N/T)\,\sigma_f^2}{P_T|\kappa|^2} + \frac{\sigma_2^2}{K\,P_T T|\kappa|^2\varrho_g}. \tag{8}$$

For independent Rayleigh backward channels with maximal-ratio combining ($K \geq 2$) the corresponding expression is

$$\varepsilon_b^{\mathrm{Ray}} = \frac{\sigma_s^2 + \big(1 + \frac{N-1}{K-1}\big)\sigma_f^2/T}{P_T|\kappa|^2} + \frac{\sigma_2^2}{(K-1)\,P_T T|\kappa|^2\varrho_g}. \tag{9}$$

https://doi.org/10.3390/telecom7060141

The previous form $N(|\kappa|^2\sigma_1^2 + \sigma_d^2)/(P_T T|\kappa|^2)$ is the special case in which all of the distortion is white across the training slots, $\sigma_s^2 = 0$.

Remark 3. *Antennas, elements and training slots act on the two components differently. The static component enters the estimate of b_n only through the n-th element's own correlation with the n-th DFT column; the orthogonality of the columns removes the contribution of every other element, for any backward channel. Its contribution $\sigma_s^2/(P_T|\kappa|^2)$ is therefore independent of K, N, T and of the geometry of the backward channel. The fluctuating component leaks from all N elements, scales as N/T, and is divided by K only when the backward channel scatters, Equation (9); for rank-one line-of-sight the ratio $g_{k,m}/g_{k,n}$ does not depend on k and K gives no reduction. Only the receiver-noise term is divided by K (or $K-1$) in all cases. Since $\sigma_s^2 \gg \sigma_f^2$ whenever the surface noise is far below the pilot signal at the amplifier input, adding BS antennas does not mitigate the distortion for rank-one, Rician and Rayleigh backward channels alike (Section 6.6). If instead the surface noise is comparable to the input signal, the fluctuating component dominates and both K and T reduce it (Section 6.6).*

3.4. Bias and Error Floors

Writing $\varrho_b = \mathbb{E}\{|b_n|^2\}$, the conventional estimator, which assumes gain p, and the compensated estimator, which uses κ, satisfy

$$\mathrm{NMSE}_{\mathrm{conv}} = \left|1 - \frac{\kappa}{p}\right|^2 + \frac{|\kappa|^2}{p^2}\frac{\varepsilon_b}{\varrho_b}, \qquad \mathrm{NMSE}_{\mathrm{comp}} = \frac{\varepsilon_b}{\varrho_b}. \tag{10}$$

Defining the distortion-to-signal ratios $\gamma_d \triangleq \sigma_d^2/(|\kappa|^2\sigma_u^2)$, $\gamma_s \triangleq \sigma_s^2/(|\kappa|^2\sigma_u^2)$ and $\gamma_f \triangleq \sigma_f^2/(|\kappa|^2\sigma_u^2)$, and using $\sigma_u^2 \to P_T\varrho_b$ at high power, (8) gives $\mathrm{NMSE}_{\mathrm{comp}} \approx \gamma_s + (N/T)\gamma_f$ plus the receiver term. As $P_T \to \infty$ the fluctuating part vanishes, $\gamma_f \to 0$, because a saturated element output carries only the phase of its input. The static part tends to its hard-limiter value for Gaussian input,

$$\lim_{P_T\to\infty} \mathrm{NMSE}_{\mathrm{comp}} = \left(\frac{4}{\pi} - 1\right) + \frac{4\,\sigma_2^2}{\pi K T A_{\mathrm{sat}}^2 \varrho_g}, \tag{11}$$

$$\lim_{P_T\to\infty} \mathrm{NMSE}_{\mathrm{conv}} = 1. \tag{12}$$

The first term is -5.63 dB, independent of p, A_{sat} and ; it follows from $|g| \to A_{\mathrm{sat}}$ and $\mathbb{E}\{|u|\} = \sigma_u\sqrt{\pi}/2$. The second term of (11) is the receiver noise referred to the saturated surface output and is negligible whenever the amplified distortion dominates it. With the values of Table 2 it equals -19.3 dB, 13.7 dB below the distortion term, so the invariance of the floor with respect to K, N and T is exact only in this distortion-dominated regime. The receiver term scales as $1/(KT)$ and equals the distortion term at $\sigma_2^2 \approx -66$ dBm for $K = 16$ and $T = 17$. In deep compression $\kappa \to 0$, so the conventional estimate collapses toward zero and its NMSE approaches unity: the estimate becomes uninformative rather than merely degraded.

Corollary 1. *The static term of (8) does not depend on N, T or K. With the minimal training length $T = N + 1$ the fluctuating term is at most γ_f, which is negligible whenever $\sigma_1^2 \ll P_T\varrho_b$. The compensated floor is therefore asymptotically independent of the surface size and of the training length: enlarging the RIS or lengthening the training does not escape the distortion floor.*

3.5. Bussgang-LMMSE Estimator

Since the compensated estimator is unbiased on average over the channel prior but not MSE-optimal, we also consider the BLMMSE estimator, which additionally requires the

channel prior. With $\mathbf{b} \sim \mathcal{CN}(\mathbf{0}, \varrho_b \mathbf{I})$ and (6), the LMMSE solution is a scalar shrinkage of the compensated estimate by $\varrho_b/(\varrho_b + \varepsilon_b)$, giving

$$\mathrm{NMSE}_{\mathrm{BLMMSE}} = \frac{\varepsilon_b}{\varrho_b + \varepsilon_b} < \mathrm{NMSE}_{\mathrm{comp}}. \tag{13}$$

3.6. Computational Complexity

Because (6) is diagonal, none of the estimators requires a matrix inversion. Per BS antenna, forming $\mathbf{\Phi}_k^{\mathrm{H}} \mathbf{y}_k$ costs $O(TN)$ complex multiplications, so the total is $O(KTN)$, or $O(KN^2)$ with $T = N + 1$. A generic LS solution without the DFT structure would instead cost $O(K(N+1)^3)$ for the inversion alone. The compensated estimator differs from the conventional one only in the scalar κ, and BLMMSE adds a single scalar multiplication. The Bussgang coefficients in (4) depend on σ_u^2 alone and can be tabulated offline against input back-off, so they contribute no run-time cost. Compensation is therefore obtained at essentially no computational penalty; what it requires is knowledge of the operating point, which motivates Section 4.

4. Online Gain Tracking

4.1. Identifiability

The surface has no baseband, so κ cannot be measured locally. It is, however, observable from the estimates themselves. The conventional estimator returns $\hat{\mathbf{b}}(P_{\mathrm{T}}) = \mathbf{c}/(p\sqrt{P_{\mathrm{T}}})$, where $\mathbf{c}$ estimates the mean surface output. Its squared norm concentrates around $N\,\mathbb{E}\{|g(u)|^2\}/(p^2 P_{\mathrm{T}}) = N\sigma_u^2 h^2/P_{\mathrm{T}}$, where

$$h(\mathrm{IBO}) \triangleq \frac{1}{p}\sqrt{|\kappa|^2 + \sigma_d^2/\sigma_u^2} = \frac{1}{p}\sqrt{\mathbb{E}\{|g(u)|^2\}/\sigma_u^2} \tag{14}$$

is the *RMS gain*, a universal function of the input back-off $\mathrm{IBO} = A_{\mathrm{sat}}^2/(p^2\sigma_u^2)$ for a given . With $\sigma_u^2 \approx P_{\mathrm{T}}\varrho_b$ in the useful regime, the ratio of back-offs at two pilot powers $P_1 < P_2$ is

$$\frac{\mathrm{IBO}_1}{\mathrm{IBO}_2} \approx \frac{P_2}{P_1} \triangleq \varrho, \tag{15}$$

which is *known by construction*. Measuring

$$r = \frac{\|\hat{\mathbf{b}}(P_2)\|}{\|\hat{\mathbf{b}}(P_1)\|} = \frac{h(\mathrm{IBO}_2)}{h(\varrho\,\mathrm{IBO}_2)} \tag{16}$$

then gives one equation in one unknown, IBO_2. Neither A_{sat} nor ϱ_b is required separately: only the operating back-off, which is exactly what the estimator needs. The estimate follows as $\hat{\kappa} = p f(\widehat{\mathrm{IBO}}_2)$ with $f(\mathrm{IBO}) = \kappa/p$ taken from the tabulated Bussgang curve. A natural but inexact alternative takes $f = |\kappa|/p$ in place of h in (16). Because σ_d^2/σ_u^2 is not negligible under compression, that choice biases $\hat{\kappa}$ by 4% to 16% for IBO between −10 and 0 dB (Section 6.9); all tracker results use the exact table h.

4.2. Probe Design

The probe ratio ϱ trades observability against noise. For ϱ small, $\kappa(P_1) \approx \kappa(P_2)$, $r \approx 1$, and inverting (16) is ill-conditioned. For ϱ large, P_1 falls into the noise and $\|\hat{\mathbf{b}}(P_1)\|$ becomes unreliable. An interior optimum therefore exists, on a plateau; it is located numerically in Section 6.

Remark 4. *Only the magnitude of the ratio is needed. Once* IBO *is identified, the complete κ—including its phase for an AM/PM amplifier—follows from the tabulated curve. The phase is inferred, not measured, and Section 5 shows that inverting (16) on the phase instead is markedly worse, because the phase response saturates while the magnitude response does not.*

The overhead is one probe frame at reduced power, and the added complexity is a scalar root find, negligible against the $O(KTN)$ estimator cost.

4.3. Well-Posedness and Uncertainty

Uniqueness and conditioning. One equation in one unknown has a unique root only if the map $r(\mathrm{IBO}_2)$ is injective, and we do not claim this in general. For the Rapp response with = 2 and ϱ = 10 dB we evaluated r on a logarithmic grid of 201 points over $\mathrm{IBO}_2 \in [10^{-5}, 10^{5}]$ and found it strictly increasing, with range from $\varrho^{-1/2} = 0.316$ in deep compression to 1 in the linear regime. We repeated the test on a 401-point grid for $\in \{0.5, 1, 2, 3, 5, 10\}$ and $\varrho \in \{3, 6, 10, 15, 20\}$ dB, and for the Saleh response, whose AM/AM curve is itself non-monotone; in all 35 combinations r was non-decreasing over the full range and over the guard window. This is numerical evidence on a finite grid for the response families considered here. It is not a proof, and it does not cover an arbitrary measured device curve. For such a curve r should be tabulated and checked for monotonicity before use; if it is not monotone the root is no longer unique, and the window should be narrowed to a monotone branch.

The guard window addresses ill-conditioning, not non-uniqueness. The logarithmic sensitivity $\mathrm{d}\ln r/\mathrm{d}\ln \mathrm{IBO}_2$ peaks at 0.29 near 0 dB and falls below 0.05 outside $[-18, +7.5]$ dB, where r is flat and the root is ill-conditioned. The estimate is therefore clipped to a guard window $[-12, +8]$ dB, so that outside the informative region the scheme degrades gracefully to the nearest edge of the window. Beyond the window the loss stays within about ± 1 dB, against 20 to 30 dB without the guard (Section 6.9), and a mismatch of the shape costs 0.9 to 1.3 dB for a true of 3 to 10 (Section 6.9). The guard does not make the estimate accurate where r is flat, and it cannot repair a table whose shape is wrong enough to make the mapping non-monotone. As an operating rule, which we have not evaluated, an estimate that lands on a window edge should be treated as unreliable, and the conservative calibration of Section 6.4 used instead.

Finite-sample uncertainty. The statistic r is the ratio of two norms of N-element estimates. By the delta method $\sigma(\ln \widehat{\mathrm{IBO}}) \approx \sigma(\ln r)/|\mathrm{d}\ln r/\mathrm{d}\ln \mathrm{IBO}|$, so the uncertainty is amplified where the sensitivity is small. Monte Carlo results are given in Section 6.

5. Robustness

5.1. AM/PM Conversion

The Saleh model adds phase distortion,

$$|g(u)| = \frac{p|u|}{1 + \beta_a(|u|/u_s)^2}, \qquad \arg g = \arg u + \frac{\alpha_p(|u|/u_s)^2}{1 + \beta_p(|u|/u_s)^2}, \tag{17}$$

with u_s chosen so the peak output amplitude equals A_{sat}. The Bussgang gain is then *complex*, so the estimate carries a phase rotation in addition to a scale error. Two consequences are quantified in Section 6: correcting the magnitude alone is insufficient, and because the AM/AM response of (17) is non-monotone, compensation has a validity limit beyond which it is counterproductive.

5.2. Per-Element Spread

Real surfaces have element-to-element variation. Let $A_{\mathrm{sat}}^{(n)} = A_{\mathrm{sat}} 10^{\epsilon_n/20}$ with ϵ_n zero-mean Gaussian of standard deviation σ_A dB. The Gram matrix (6) remains diagonal with entries $T|\kappa_n|^2\varrho_g$, so per-element compensation remains exact; the loss is incurred only by approximating $\{\kappa_n\}$ with a single scalar.

6. Numerical Results

All curves are obtained by Monte Carlo simulation of (2) or (17); closed forms are overlaid where stated. Parameters are listed in Table 2. Results are reported against input back-off, so they do not depend on the absolute saturation level of any particular device.

Simulation protocol. The pilot symbol is $s(t) = 1$, so P_{T} is the pilot power. In every realization $\mathbf{b} \sim \mathcal{CN}(\mathbf{0}, \varrho_b \mathbf{I})$, $\mathbf{h}_{\mathrm{d}} \sim \mathcal{CN}(\mathbf{0}, \varrho_d \mathbf{I})$, $\mathbf{v}$ and the receiver noise are drawn afresh, and the backward channel is a rank-one line-of-sight link of constant modulus with angles of arrival and departure drawn uniformly (Rician and Rayleigh backward channels are used only where stated). The Saleh parameters are $\beta_a = 1.1517$, $\alpha_p = 4.0033$ and $\beta_p = 9.1040$.

6.1. Error Floor and Non-Monotonicity

Figure 1 shows the NMSE of the user–RIS estimate.

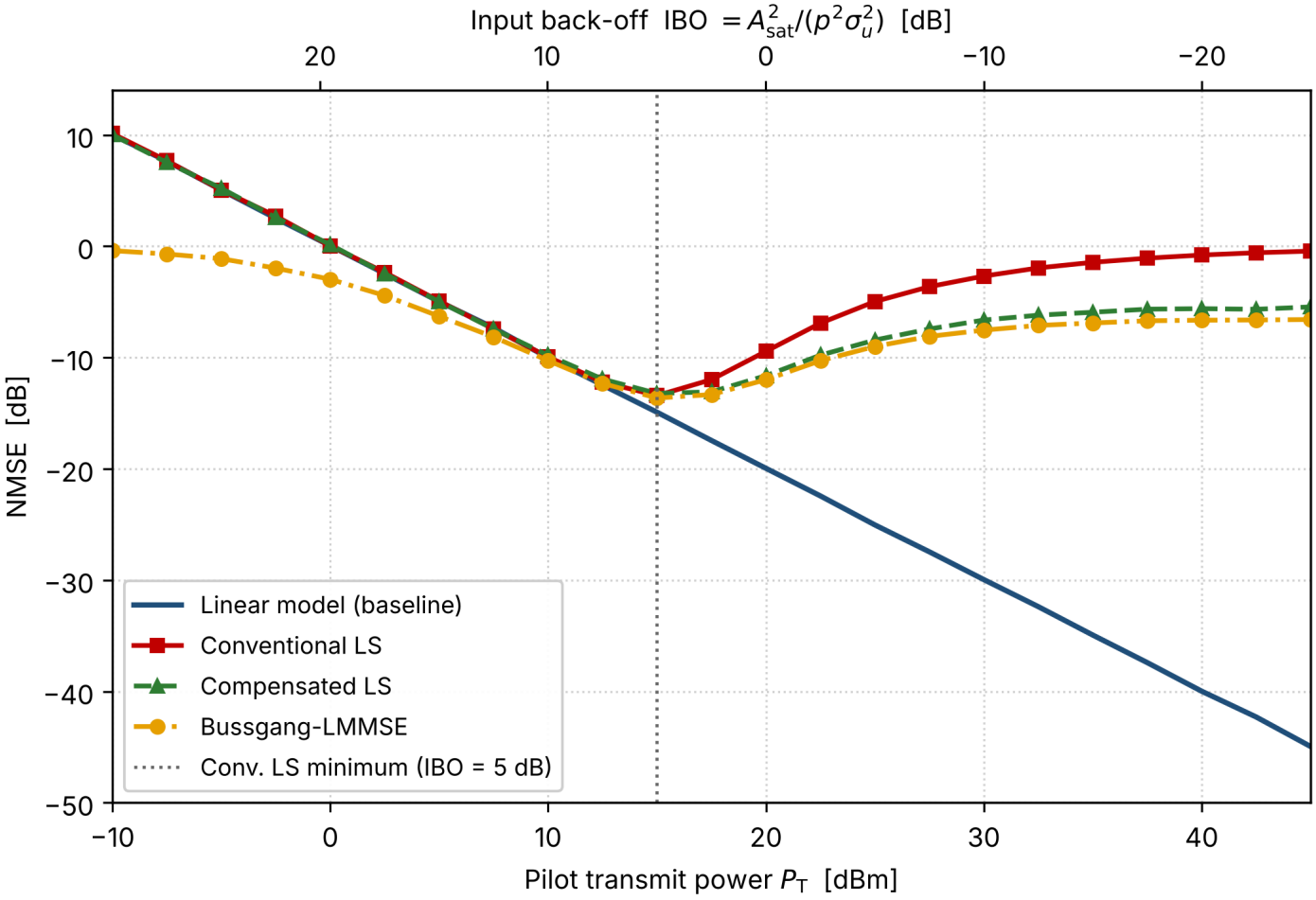


Figure 1. NMSE versus pilot power, with input back-off on the upper axis.

The linear model predicts a 10 dB/decade slope down to −45 dB. The conventional estimator tracks it until the compression knee at IBO ≈ 5 dB, reaches a minimum of −13.4 dB, then reverses and degrades to −0.44 dB, a gap of 44.5 dB from the linear-model prediction. The compensated estimator floors at −5.5 dB and restores monotonicity; BLMMSE gains a further 1 dB at 45 dBm, confirming (13). Its much larger advantage at low pilot power reflects shrinkage toward zero in a regime where the estimate carries little information. The compensated floor is the hard-limiter value of Equation (11), $4/\pi - 1 = -5.63$ dB, plus the receiver term, −5.45 dB in total. Equation (8) matches simulation to within 0.25 dB over the twelve configurations of Section 6.6 and Table 3.

Table 3. Compensated NMSE [dB]: simulation (95% CI) against the corrected Equation (8) (or (9)) and the previous form; $P_T = 30$ dBm unless noted; the last row uses $P_T = 20$ dBm.

T	σ_1^2 [dBm]	K	Backward	Simulation	Corrected	Previous
17	−70	16	rank-one	−6.64 [−6.80, −6.46]	−6.66	−6.91
34	−70	16	rank-one	−6.70 [−6.85, −6.56]	−6.79	−9.92
68	−70	16	rank-one	−6.80 [−6.96, −6.65]	−6.86	−12.93
136	−70	16	rank-one	−6.80 [−6.98, −6.65]	−6.89	−15.94
17	−30	16	rank-one	1.57 [1.35, 1.77]	1.48	−3.34
34	−30	16	rank-one	−1.27 [−1.48, −1.07]	−1.43	−6.35
68	−30	16	rank-one	−4.14 [−4.34, −3.96]	−4.24	−9.36
136	−30	16	rank-one	−7.07 [−7.26, −6.88]	−6.88	−12.37
17	−30	64	Rayleigh	−8.51 [−8.68, −8.33]	−8.42	−3.52
68	−30	64	Rayleigh	−12.09 [−12.27, −11.93]	−12.15	−9.54
17	−70	64	Rayleigh	−6.61 [−6.78, −6.45]	−6.85	−7.11
17	−50	16	rank-one	−7.59 [−7.76, −7.41]	−7.58	−11.09

6.2. The Prior-Art Training Rule

Figure 2 evaluates the training amplification chosen under the linear model. Sum-NMSE over $[h_d; \mathbf{b}]$ is used because the amplitude trade-off in the prior art originates in the direct link. Under the linear model the sum-NMSE decreases monotonically over the admissible range—the unconstrained stationary point lies at $a \approx 222$—so the rule reduces to using maximum gain. Under compression the curve turns at $a = 3.5$; operating at $a = a_{\max} = 10$ costs 3.4 dB. The published rule does not merely lose accuracy; it points in the wrong direction.

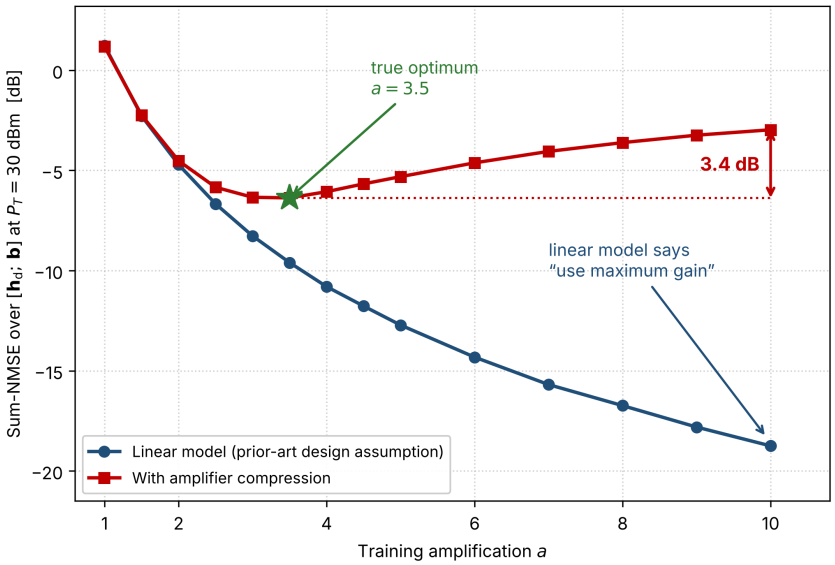


Figure 2. Sum-NMSE versus training amplification.

6.3. Scaling in Antennas and Elements

Figure 3 sweeps K. The linear model improves by 16.7 dB from $K = 1$ to 64, slightly less than $10 \log_{10} K = 18.1$ dB. With compression the compensated estimator gains only 2.9 dB and saturates beyond $K = 16$ (it changes by 0.3 dB from 16 to 64), while the conventional estimator stays within 0.15 dB of −2.7 dB because its error is bias-dominated. This confirms Remark 3 for the rank-one geometry; Section 6.6 extends it to scattering backward channels.

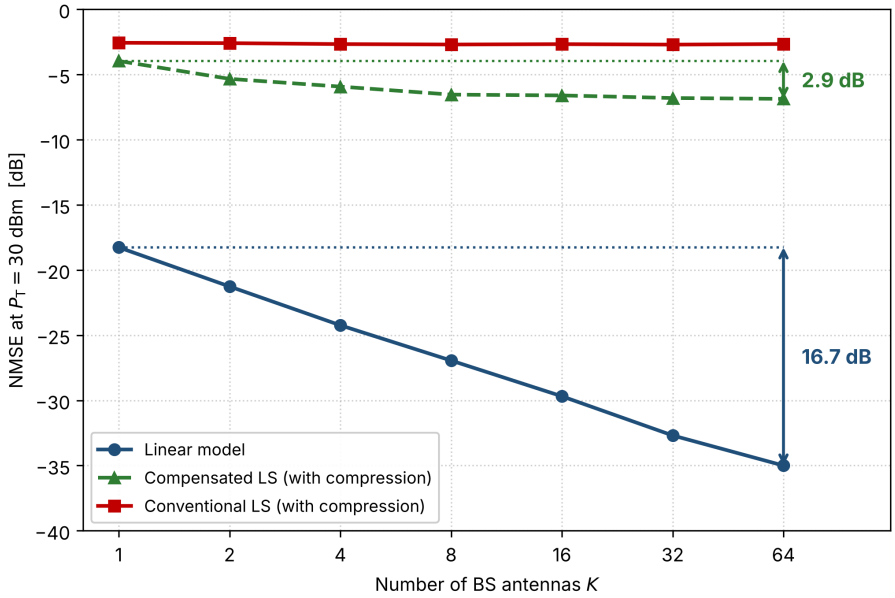


Figure 3. NMSE versus the number of BS antennas at $P_T = 30$ dBm.

Figure 4a sweeps N with $T = N + 1$. The compensated floor moves only from −6.15 to −6.74 dB as N grows eightfold, while the linear model improves by 7.3 dB. The corrected expression (8) gives −6.44 and −6.85 dB at $N = 8$ and 64, within 0.29 and 0.11 dB of simulation, and its static term $\gamma_s = -6.94$ dB does not depend on N, confirming Corollary 1.

Figure 4b sweeps . Softer compression is better despite yielding a smaller equivalent gain: at = 0.5 the compensated NMSE is −8.71 dB against −6.09 dB at = 8, because γ_s ($= \gamma_d$ here), not κ, sets the floor (Equation (8) gives −8.68 and −6.24 dB, within 0.15 dB at every tested). Harder clipping raises the floor and then saturates. When selecting a reflection amplifier, compression smoothness therefore matters more than peak gain if channel estimation is the binding constraint.

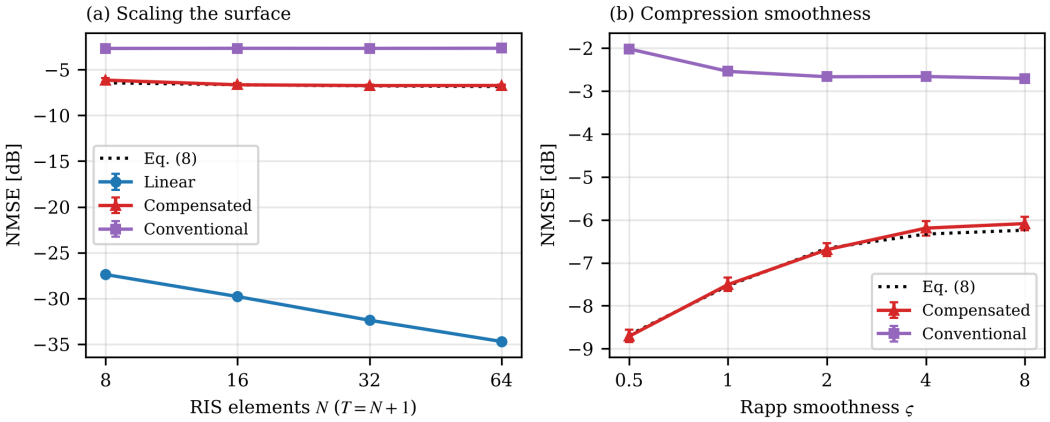


Figure 4. Scaling in (**a**) surface size and (**b**) compression smoothness.

6.4. Calibration Sensitivity

Figure 5 evaluates a compensated estimator calibrated at a nominal A_{sat} and deployed at a different true value. The failure is asymmetric. The mismatch is the true value minus the assumed one. A positive mismatch means that more compression was assumed than occurs: $\hat{\kappa}$ is too small, and the estimate, which is scaled by $\kappa/\hat{\kappa}$, is inflated and its error amplified. Beyond +3 dB the compensated estimator is worse than the uncompensated one, with a 14.3 dB penalty at +10 dB. A negative mismatch means that less compression was

assumed than occurs: $\hat{\kappa}$ is too large, the estimate is scaled down, which is harmless and can help, consistent with (13).

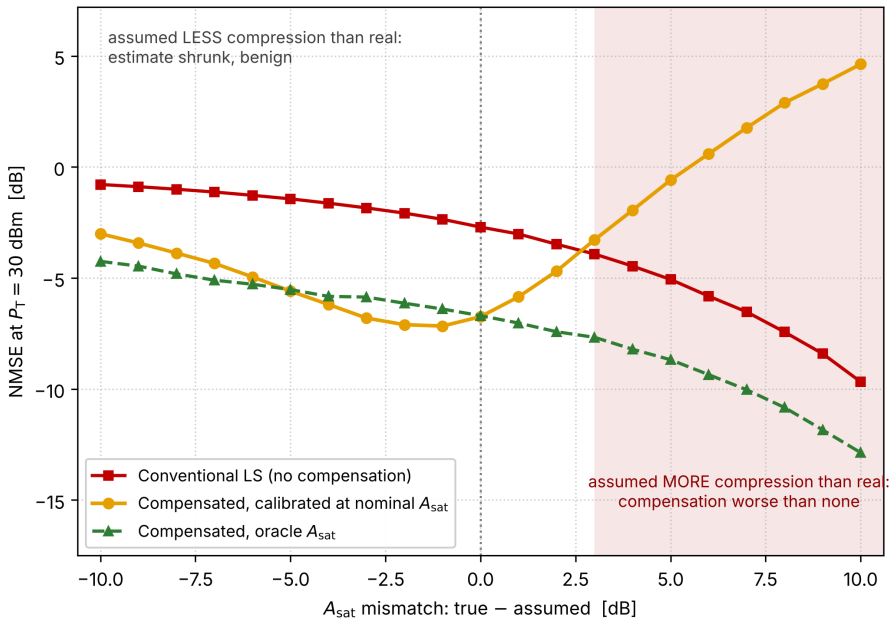


Figure 5. Sensitivity to A_{sat} calibration error at $P_T = 30$ dBm.

If A_{sat} is known only to within ±6 dB, calibrating at nominal gives a worst-case NMSE of +0.6 dB, worse than not compensating at all (−1.3 dB). Deliberately calibrating 3 dB *above* nominal is minimax-optimal at −3.3 dB, read from the curve of Figure 5. Where tracking is unavailable, a biased conservative calibration is preferable to an unbiased nominal one.

6.5. Gain Tracking

Figure 6 evaluates Algorithm 1 with the exact RMS-gain table and the guard window (300 realizations, 95% confidence intervals). Fixed calibration at the nominal A_{sat} degrades from −6.7 to +0.6 dB as the true A_{sat} rises by 6 dB, whereas tracking improves to −9.5 dB: a 10.1 dB gain at the largest drift, within 0.1 dB of the oracle for deviations of 0 dB and above. For deviations of −3 and −6 dB the true back-off (−13 and −16 dB) lies below the guard window, the estimate is clipped at −12 dB, and the tracked estimate is 1.0 and 0.8 dB better than the oracle: clipping over-estimates κ, and a κ biased high produces the same beneficial shrinkage as (13).

Figure 7 locates the probe optimum on a plateau at $\varrho = 10$ to 14 dB (−7.72 and −7.73 dB). The loss is about 1 dB at $\varrho = 4$ dB, where $\kappa(P_1) \approx \kappa(P_2)$ and the ratio is ill-conditioned, and 1.8 dB at 25 dB, where P_1 falls into the noise and the back-off estimate is biased by −1.3 dB; between 4 and 18 dB the bias stays within 0.5 dB. Excessive separation is thus the more damaging error. A single probe frame suffices.

Algorithm 1 Dual-power gain tracking

1: **Offline:** tabulate $f(\mathrm{IBO}) = \kappa/p$ and $h(\mathrm{IBO})$ from (4), (14)
2: **Input:** operating power P_2, probe ratio ϱ
3: Acquire $\hat{\mathbf{b}}(P_2)$ and $\hat{\mathbf{b}}(P_2/\varrho)$ with the conventional estimator
4: $r \leftarrow \|\hat{\mathbf{b}}(P_2)\|/\|\hat{\mathbf{b}}(P_2/\varrho)\|$
5: Solve (16) for $\widehat{\mathrm{IBO}}_2$ by scalar root finding ; clip the root to the guard window $[-12, +8]$ dB
6: $\hat{\kappa} \leftarrow p f(\widehat{\mathrm{IBO}}_2)$
7: **Return** compensated estimate $\hat{\mathbf{b}}(P_2)$ formed with $\hat{\kappa}$

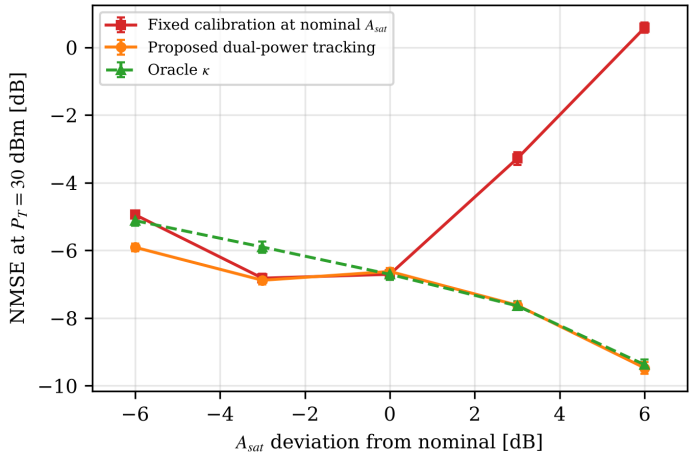


Figure 6. Tracking versus fixed calibration under saturation drift.

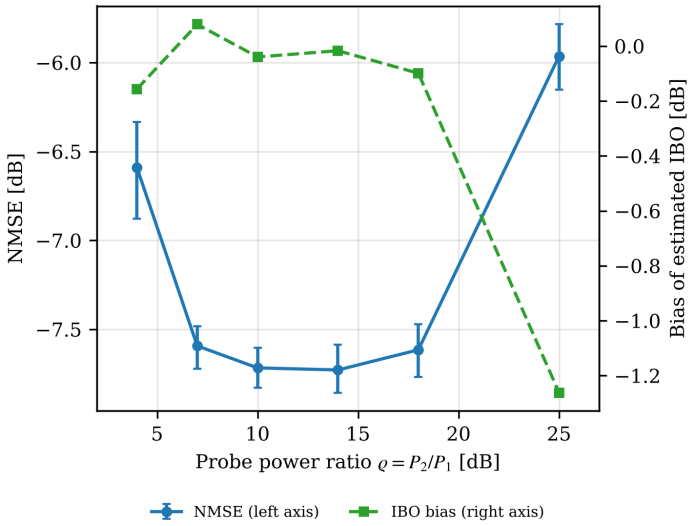


Figure 7. Probe power ratio: observability against probe noise.

6.6. Static Versus Fluctuating Distortion and Scattering Backward Channels

Figure 8a varies the training length T at $P_T = 30$ dBm. With the surface noise of Table 2 (-70 dBm) the compensated NMSE stays at -6.6 to -6.8 dB from $T = 17$ to 136, and equals the NMSE of the noiseless static mapping, -7.0 dB. The previous form $(N/T)\gamma_d$ predicts a fall to -15.9 dB and agrees with simulation only at $T = N + 1$, where $N/T \approx 1$. With surface noise raised to -30 dBm the fluctuating component dominates and the NMSE falls as $1/T$, from $+1.6$ to -7.1 dB, as Equation (8) predicts. Table 3 lists all twelve configurations; eleven fall inside the 95% confidence interval (one at its boundary) and the largest deviation is 0.24 dB.

Figure 8b sweeps K for four backward channels. For rank-one LoS, Rician (10 and 0 dB K-factor) and Rayleigh links the compensated NMSE moves only from between -5.1 and -4.0 dB at $K = 2$ to between -6.9 and -6.7 dB at $K = 128$, the common floor $\gamma_s = -6.9$ dB, while the linear model improves by 15.6 dB. The static floor is thus independent of the backward-channel geometry, as Remark 3 states, and the statement "additional antennas do not mitigate the distortion" is not restricted to the rank-one geometry in this regime. Equations (8) and (9) track the simulated rank-one and Rayleigh curves to within 0.2 dB for $K \geq 4$.

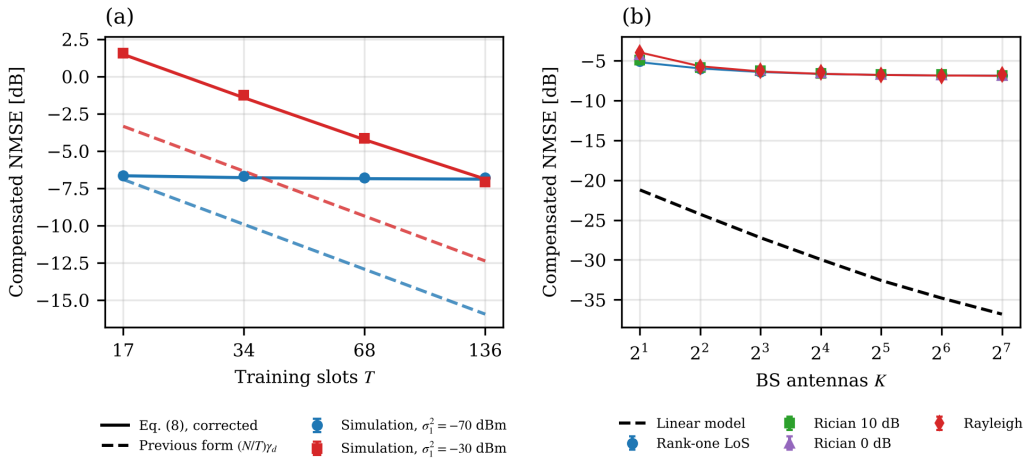


Figure 8. (**a**) Compensated NMSE versus training length for two surface-noise levels: simulation with 95% confidence intervals, corrected Equation (8) (solid) and previous form (dashed). (**b**) NMSE versus BS antennas for four backward channels, with Equations (8) and (9) (thin lines) and the linear model (dashed).

6.7. Benchmark with the Exact Amplifier Response

To measure what the Bussgang linearization costs, we compare the estimators with a numerical Bayes-MMSE estimator that uses the exact Rapp response, evaluated per element by prior sampling with the measured estimation-noise variance (a genie noise level). Figure 9 shows that BLMMSE is within 0.06 dB of it at $P_{\mathrm{T}} \leq 10$ dBm and at $P_{\mathrm{T}} \geq 40$ dBm, and at most 1.32 dB worse, at 20 dBm, where the knee of the amplifier lies. The linearization therefore costs little, and the floor is not an artifact of it: it reflects the information destroyed by saturation, which no antenna count restores.

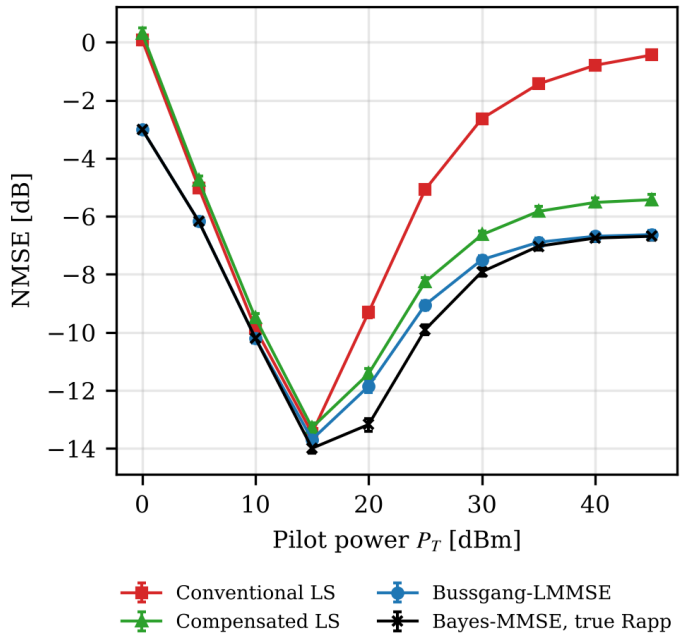


Figure 9. NMSE versus pilot power: conventional, compensated, BLMMSE and the Bayes-MMSE with the exact Rapp response (95% confidence intervals).

6.8. Where the Training Rule Reverses

Figure 10 maps the optimal training amplification $a^\star$ against three quantities, with the estimator assuming gain a as in Section 6. The direct-link strength has no effect: $a^\star$ stays at

3 to 3.5 and the loss at $a_{\max} = 10$ stays at 3.1 to 3.5 dB from $\varrho_d/\varrho_b = -30$ to +10 dB. Receiver noise and headroom decide the outcome. The maximum-gain rule is near-optimal ($a^\star = 9$, 0.03 dB loss) when the receiver noise reaches −70 dBm, and optimal at −60 dBm, whereas at −100 dBm it loses 16 dB. Raising A_{sat} by 14 dB moves $a^\star$ to 9 and the loss to 0.7 dB, while lowering it by 6 dB leaves $a^\star = 3$ with 1.9 dB loss. The rule reverses when the distortion floor, not the receiver noise, limits the estimate. We do not have a closed-form condition: the ratio of the receiver-noise term to the distortion term at $a_{\max}$ does not give one threshold across the headroom sweep.

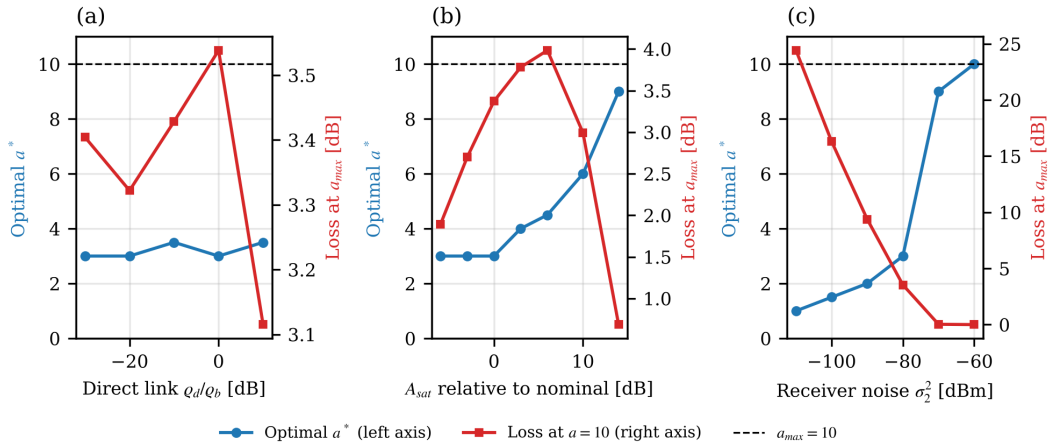


Figure 10. Optimal training amplification $a^\star$ (left axis) and loss of the maximum-gain rule (right axis) versus (**a**) direct-link strength, (**b**) amplifier headroom and (**c**) receiver noise. The dashed line marks $a_{\max} = 10$.

6.9. Tracker Well-Posedness and Robustness

Figure 11 summarizes the experiments of Section 4, with $N = K = 16$ unless stated. Figure 11a shows the monotone probe ratio for both tables. Figure 11b compares the two tables at $P_{\text{T}} = 30$ dBm: between IBO = −10 and 0 dB the exact RMS table gives a κ bias of 0.2% to 0.4% against 4% to 16% for the $|\kappa|/p$ table, and the loss against the oracle is at most 0.24 dB. Below −13 dB the true back-off is outside the guard window, the estimate sits at the window edge, and the loss stays within about ±1 dB because the edge value over-estimates κ and shrinks the estimate; the NMSE there is already high. Without the guard, rare estimates at the end of the table gave losses of 20 to 30 dB, so the guard is necessary. Negative losses in the figure are the shrinkage effect of a κ biased high.

Figure 11c: the standard deviation of $\widehat{\text{IBO}}$ at IBO = −10 dB falls from 1.7 dB at $N = 8$ to 0.44 dB at $N = 128$, and that of $\hat{\kappa}/\kappa$ from 20% to 4.8%, close to $N^{-1/2}$; the NMSE loss against the oracle is at most 0.15 dB. Figure 11d: decorrelating the channel between the probes, modeled as $\mathbf{b}_1 = \rho\mathbf{b}_2 + \sqrt{1-\rho^2}\,\mathbf{e}$, costs 0.5 dB at $\rho = 0.6$, because only norms enter the ratio. Figure 11e: what matters is a power change between the probes. A +1 dB channel-gain drift gives −10% bias in $\hat{\kappa}$ and 1.2 dB loss, +3 dB gives 2.5 dB; a +2 dB drift of A_{sat} between the probes gives 2.1 dB. Figure 11f: with the table built for = 2, a true of 3, 5 and 10 gives κ biases of −6%, −8% and −9% and losses of 0.9, 1.1 and 1.3 dB, and = 1 gives a +25% bias with a 0.85 dB gain from shrinkage. In this single operating point the tracker is therefore robust to aging and shape mismatch, and sensitive mainly to a power change between the probes.

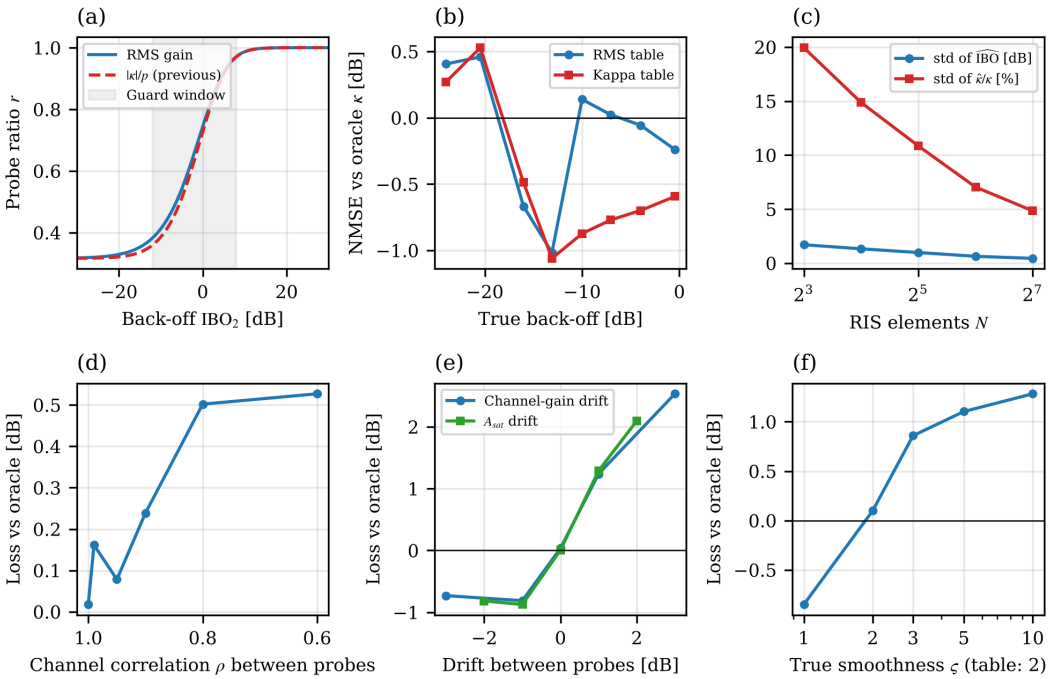


Figure 11. Tracker well-posedness and robustness at $P_T = 30$ dBm unless stated. (**a**) Probe ratio. (**b**) Loss against oracle κ for the two tables. (**c**) Finite-sample spread versus N. (**d**) Channel correlation between the probes. (**e**) Power drift between the probes. (**f**) Shape mismatch of .

6.10. AM/PM and Validity Limit

Figure 12 evaluates the Saleh model. The phase rotation saturates near 25°. Correcting the magnitude alone leaves −7.94 dB at IBO = 0 dB, against −11.72 dB with complex κ: phase correction is worth 3.8 dB and is not optional. Below IBO ≈ −15 dB, however, compensation becomes counterproductive—+1.68 dB against −0.27 dB for the conventional estimator—because the AM/AM response of (17) is non-monotone, the useful component vanishes, and debiasing amplifies an estimate that is predominantly distortion. Compensation is therefore beneficial within a defined operating region rather than universally.

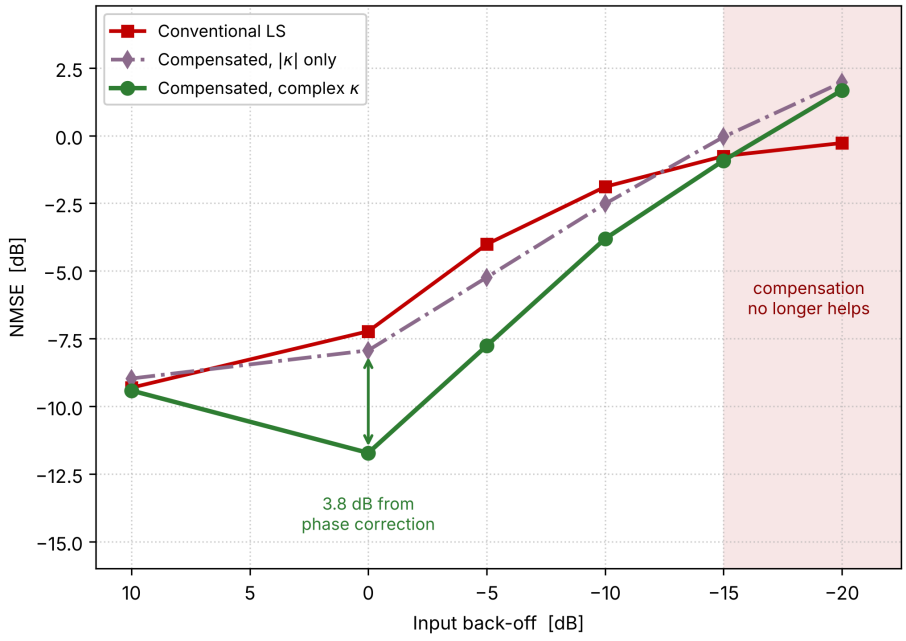


Figure 12. Saleh model: AM/PM and the validity limit of compensation.

6.11. Element Spread

Figure 13 sweeps σ_A; a fresh log-normal spread is drawn for every realization, and the per-element estimator is given κ_n evaluated at the same draw. That estimator is a genie, since the individual gains are not available in practice. Its NMSE is flat within 0.2 dB across the whole range (-6.65 to -6.77 dB), so the spread costs nothing physically and the loss is attributable to the single-gain approximation, which costs 3.2 dB at $\sigma_A = 4$ dB (nominal gain, -3.54 dB). Setting the single gain to the mean of $\kappa(A)$ over the spread rather than to the nominal value recovers 0.8 dB at $\sigma_A = 4$ dB. Compensation with a single gain remains better than none throughout, but the margin over the conventional estimator shrinks to 0.8 dB at $\sigma_A = 4$ dB (-3.54 against -2.77 dB), so a single gain is of limited value for large spreads.

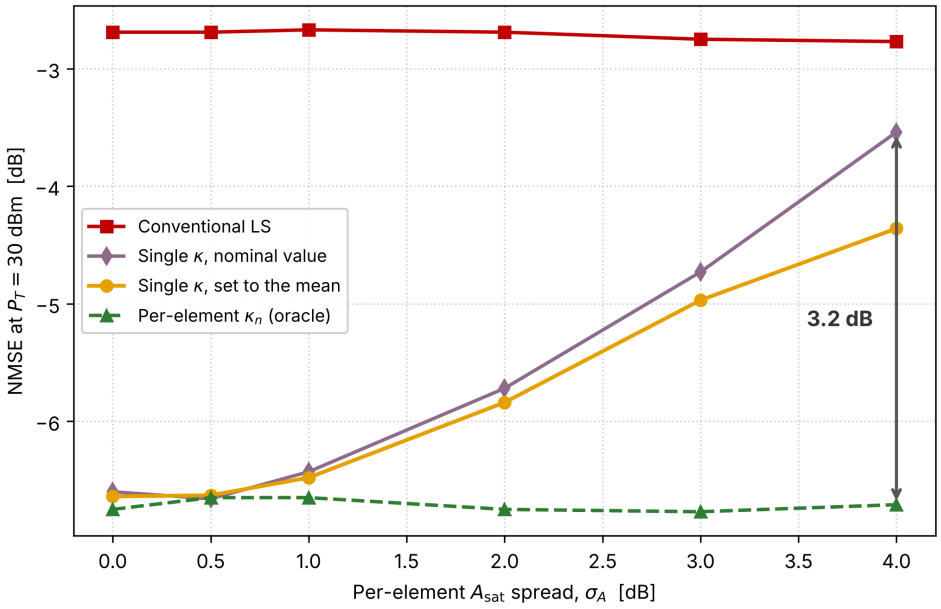


Figure 13. Per-element saturation spread.

6.12. Simulation Details

The MATLAB R2017b (MathWorks, Natick, MA, USA) figures use 250 to 800 realizations per point (the default is 600; Figures 2, 3, 5, 12 and 13 use 700, 250, 800, 300 and 600) with fixed seeds listed in the released code. The MATLAB reproductions of Figures 6 and 7 use fewer realizations (120 and 350) and agree with the Python 3.13 values to within 0.6 dB. Figures 4 and 6–11 use 120 to 300 realizations and report 95% bootstrap confidence intervals over realizations; the Python scripts that produce them are released with the MATLAB code, and they agree with the MATLAB values of Figure 1 to within 0.2 dB (conventional NMSE -0.44 dB at 45 dBm in both; compensated floor -5.45 dB and -5.43 dB). Independent reruns of the same code with different random-number generators differ by up to 0.2 dB. Where differences below 1 dB are discussed, the confidence interval half-width is 0.1 to 0.3 dB.

7. Conclusions

The linear amplifier assumption underlying active-RIS channel estimation was validated only in the small-signal regime. Removing it changes the problem qualitatively: the observation matrix becomes pilot-power dependent, the conventional estimator, which assumes a constant gain, becomes biased (LS itself remains valid when it uses the compression-aware observation matrix), the closed-form training rules of the linear model lose their optimality, and the NMSE exhibits both a floor and a non-monotone dependence on pilot

power. The dominant part of the distortion is static per element, so neither more antennas, nor more elements, nor longer training escapes that floor, whatever the backward-channel geometry in the simulated regime, and the Bussgang-based estimators lose at most 1.3 dB against a Bayes-MMSE benchmark with the exact amplifier response. The published training-amplification rule points the wrong way unless the receiver noise is high or the headroom is large. A dual-power probe identifies the operating back-off without baseband processing at the surface and recovers most of the loss caused by saturation drift. Compensation is beneficial over a defined operating region, is asymmetrically sensitive to calibration, and is limited by the single-gain approximation rather than by element spread itself. Experimental validation on a fabricated surface, and the coupling between distorted CSI and the predicted distortion directions of [7], remain open.

Author Contributions: Conceptualization, A.B.I. and I.K.E.; methodology, A.B.I.; software, A.B.I.; validation, A.B.I.; formal analysis, A.B.I.; investigation, A.B.I.; writing—original draft preparation, A.B.I.; writing—review and editing, A.B.I. and I.K.E.; supervision, I.K.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The MATLAB code that produces the numbers of Figures 1–3, 5, 12 and 13 (folder matlab/), the Python scripts that produce the numbers and plots of Figures 4 and 6–11, and the Python plotting scripts that draw Figures 1, 2, 3, 5, 12 and 13 from the saved numbers (folder revision/) are openly available at https://github.com/ApiEdu/Channel-Estimation-for-Active-RIS-under-Amplifier-Compression (accessed on 20 October 2026).

Acknowledgments: The authors used an artificial intelligence assistant (Claude Sonnet 5.5, Anthropic) during the preparation of this work for literature search, implementation of the simulation code, derivation checking, and language editing. All analytical results were independently verified by the authors, who reviewed every source cited and take full responsibility for the content of this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Derivation of the Error Variance

After LS with the DFT design, the per-antenna error is

$$\hat{b}_n^{(k)} - b_n = \frac{1}{\sqrt{P_\mathrm{T}} T \kappa g_{k,n}} \sum_{t=1}^{T} F_{t,n}^{*} n_{\mathrm{eff},k}(t), \tag{A1}$$

with $n_{\mathrm{eff},k}(t) = \sum_m g_{k,m} F_{t,m} w_m(t) + z_k(t)$ and $w_m(t) = \bar{d}_m + \tilde{w}_m(t)$, where $\tilde{w}_m(t) = \kappa v_m(t) + \tilde{d}_m(t)$ has variance σ_f^2 and is independent across m and t. *Static part.* Since $\bar{d}_m$ does not depend on t, $\sum_t F_{t,n}^{*} F_{t,m} \bar{d}_m = T \delta_{mn} \bar{d}_n$ by orthogonality of the DFT columns, so the static part contributes $\bar{d}_n / (\sqrt{P_\mathrm{T}} \kappa)$ exactly, with variance $\sigma_s^2 / (P_\mathrm{T} |\kappa|^2)$, for any $\mathbf{G}$ and any K (the factor $g_{k,n}$ cancels in the per-antenna and in the combined estimate). *Fluctuating part.* The inner sum does not collapse to δ_{mn}; its variance is $T\sigma_f^2$ per m, and summing over m with $|g_{k,m}|^2 = \varrho_g$ gives $N \varrho_g T \sigma_f^2$. Adding $T\sigma_2^2$ and dividing by $P_\mathrm{T} T^2 |\kappa|^2 \varrho_g$ gives the fluctuating and receiver terms of the per-antenna variance, whose sum with the static part is (8) for $K = 1$.

For the K-antenna average, write $g_{k,m} = \sqrt{\varrho_g}\, a_{r,k} a_{t,m}^{*}$ for the rank-one LoS link. Then $g_{k,m} / g_{k,n} = a_{t,m}^{*} / a_{t,n}^{*}$ is independent of k: the fluctuating contributions are identical for every antenna and survive averaging unchanged, whereas z_k is independent across k and averages down by K. This yields (8). For independent Rayleigh entries with maximal-ratio combining, conditioned on $\mathbf{g}_n$ the cross terms $\sum_k g_{k,n}^{*} g_{k,m}$, $m \neq n$, are $\mathcal{CN}(0, \varrho_g \sum_k |g_{k,n}|^2)$, and $\mathbb{E}\{1/\sum_k |g_{k,n}|^2\} = 1/((K-1)\varrho_g)$ for $K \geq 2$. The fluctuating variance is then $(1 + (N-1)/(K-1))\sigma_f^2/T$ and the receiver term is $\sigma_2^2/((K-1)T\varrho_g)$ before division by $P_\mathrm{T}|\kappa|^2$, which is

(9). The static part is unchanged. For Rician links no closed form is claimed; the simulations of Figure 8b show the same floor. □

References

1. Zhang, Z.; Dai, L.; Chen, X.; Liu, C.; Yang, F.; Schober, R.; Poor, H.V. Active RIS vs. Passive RIS: Which Will Prevail in 6G? *IEEE Trans. Commun.* **2023**, *71*, 1707–1725. [CrossRef]
2. Long, R.; Liang, Y.C.; Pei, Y.; Larsson, E.G. Active Reconfigurable Intelligent Surface-Aided Wireless Communications. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 4962–4975. [CrossRef]
3. Zhang, Z.; Dai, L.; Chen, X.; Liu, C.; Yang, F.; Schober, R.; Poor, H.V. Active RISs: Signal Modeling, Asymptotic Analysis, and Beamforming Design. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), Rio de Janeiro, Brazil, 4–8 December 2022; pp. 1618–1624. [CrossRef]
4. Bousquet, J.F.; Magierowski, S.; Messier, G.G. A 4-GHz Active Scatterer in 130-nm CMOS for Phase Sweep Amplify-and-Forward. *IEEE Trans. Circuits Syst. I* **2012**, *59*, 529–540. [CrossRef]
5. Jia, Y.; Tang, X.; Xu, Y.; Guo, Y.X. Assessing the Third-Order Intermodulation of Reconfigurable Intelligent Surface with Embedded Varactor Diode. *IEEE Trans. Microw. Theory Tech.* **2025**, *73*, 10152–10163. [CrossRef]
6. Tang, X.; Jia, Y.; Xu, Y.; Zhang, X.; Lu, W.B. Radiation Pattern Prediction for Third-Order Intermodulation of the Reconfigurable Reflectarray Antenna with Embedded Varactor. *Int. J. Numer. Model. Electron. Netw. Devices Fields* **2026**, *39*, e70151. [CrossRef]
7. Kolomvakis, N.; Kosasih, A.; Björnson, E. Nonlinear Distortion Radiated From Large Arrays and Active Reconfigurable Intelligent Surfaces. *IEEE Trans. Wirel. Commun.* **2025**, *24*, 5037–5049. [CrossRef]
8. Kolomvakis, N.; Björnson, E. Nonlinear Distortion Issues Created by Active Reconfigurable Intelligent Surfaces. In *Proceedings of the 18th European Conference on Antennas and Propagation (EuCAP)*; IEEE: New York, NY, USA, 2024. [CrossRef]
9. Khan, B.; Aikio, J.P.; Rahkonen, T.; Pärssinen, A.; Juntti, M.; Tervo, N. On the Behavior of Radiated Nonlinear Distortion with Digital Predistortion in Wideband Large Arrays with Near-Field Beamfocusing. *IEEE Trans. Veh. Technol.* **2025**, *75*, 6240–6253. [CrossRef]
10. Chen, H.; Li, N.; Long, R.; Liang, Y.C. Channel Estimation and Training Design for Active RIS Aided Wireless Communications. *arXiv* **2023**, arXiv:2311.02935. [CrossRef]
11. Wang, Y.; Shu, F.; Dong, R.; Gao, W.; Zhang, Q.; Liu, J. Power Optimization and Deep Learning for Channel Estimation of Active IRS-Aided IoT. *arXiv* **2024**, arXiv:2407.09254. [CrossRef]
12. Lawal, A.; Zerguine, A. Deep Learning-Based Channel Estimation for RIS-Assisted Full-Duplex MIMO System with Hardware Impairments. *IEEE Signal Process. Lett.* **2026**, *33*, 2146–2150. [CrossRef]
13. Ahmad, M.; Sarwar, M.S.; Shin, S.Y. Deep-Learning-Assisted Channel Estimation for Adaptive Parameter Selection in mMIMO-SEFDM. *IEEE Internet Things J.* **2025**, *12*, 24174–24184. [CrossRef]
14. Pan, Z.; Wang, W.; Nie, G.; Zheng, A.; Ni, W. Channel Estimation and Beamforming Design for MF-RIS-Aided Communication Systems. *IEEE Signal Process. Lett.* **2025**, *32*, 916–920. [CrossRef]
15. Sharini, D.L.; Dilli, R.; Kanthi, M.; Simha, G.D.G. Nonlinear Distortion-Aware Channel Estimation Strategies for IRS-Aided MIMO Systems. *IEEE Access* **2025**, *13*, 180847–180861. [CrossRef]
16. Elmaroud, B. Spectral Efficiency Analysis and Optimization in Hybrid RIS Systems Under Nonlinear Power Amplifier Effects. *Wirel. Pers. Commun.* **2025**, *144*, 553–575. [CrossRef]
17. Li, Z.; Hu, J.; Xing, J. Joint Beamforming Design for Active RIS-Assisted ISAC Systems with Transmitter Hardware Impairments. *Sensors* **2026**, *26*, 4682. [CrossRef] [PubMed]
18. Arjas, A.; Atzeni, I. Nonlinear Sparse Bayesian Learning Methods with Application to Massive MIMO Channel Estimation with Hardware Impairments. *IEEE Trans. Signal Process.* **2026**, *74*, 830–845. [CrossRef]
19. Radpour, H.; Hofer, M.; Mayer, L.W.; Hofmann, A.; Schiefer, M.; Zemen, T. Active Reconfigurable Intelligent Surface for the Millimeter-Wave Frequency Band: Design and Measurement Results. *arXiv* **2023**, arXiv:2306.04515. [CrossRef]
20. Gavriilidis, P.; Mishra, D.; Smida, B.; Basar, E.; Yuen, C.; Alexandropoulos, G.C. Active Reconfigurable Intelligent Surfaces: Circuit Modeling and Reflection Amplification Optimization. *arXiv* **2025**, arXiv:2503.24093. [CrossRef]
21. Rapp, C. Effects of HPA-Nonlinearity on a 4-DPSK/OFDM Signal for a Digital Sound Broadcasting System. In Proceedings of the European Conference on Satellite Communications (ECSC-2), Liège, Belgium, 22–24 October 1991; pp. 179–184.
22. Mollén, C.; Gustavsson, U.; Eriksson, T.; Larsson, E.G. Spatial Characteristics of Distortion Radiated From Antenna Arrays with Transceiver Nonlinearities. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 6663–6679. [CrossRef]
23. Bussgang, J.J. *Crosscorrelation Functions of Amplitude-Distorted Gaussian Signals*; Technical Report 216; Research Laboratory of Electronics (RLE) at the Massachusetts Institute of Technology (MIT): Cambridge, MA, USA, 1952.

24. Demir, O.T.; Björnson, E. The Bussgang Decomposition of Nonlinear Systems: Basic Theory and MIMO Extensions [Lecture Notes]. *IEEE Signal Process. Mag.* **2021**, *38*, 131–136. [CrossRef]
25. Kay, S.M. *Fundamentals of Statistical Signal Processing: Estimation Theory*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1993.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.