

Article

Decentralized Adaptive Generalized-Minimum-Variance Control of Large-Scale Interconnected Multivariable Hammerstein Systems

Slim Dhahri ^{1,*} , Mourad Elloumi ^{2,3}, Hend Aljahani ⁴, Salem Albalawi ⁵, Sahar Almashaan ⁵, Hatem Alwardi ⁵ and Foued Mtiri ⁶

¹ Department of Computer Engineering and Networks, College of Computer and Information Sciences, Jouf University, Sakaka 72341, Saudi Arabia

² Faculty of Sciences of Gafsa, University of Gafsa, Gafsa 2112, Tunisia

³ Laboratory of Sciences and Technology of Automatic Control and Computer Engineering, National School of Engineering of Sfax, Sfax University, P.O. Box 1173, Sfax 3038, Tunisia

⁴ Department of Mathematics, College of Science, Northern Border University, Arar 91431, Saudi Arabia

⁵ Department of Mathematics, College of Science, Jouf University, Sakaka 72341, Saudi Arabia; sasalbalawi@ju.edu.sa (S.A.); sralmashaan@ju.edu.sa (S.A.)

⁶ Department of Mathematics, College of Science, King Khalid University, Abha 61413, Saudi Arabia

* Correspondence: sdhahri@ju.edu.sa

Abstract

This paper presents a decentralized adaptive generalized-minimum-variance (GMV) control framework for large-scale stochastic nonlinear systems composed of interconnected multi-input multi-output (MIMO) Hammerstein subsystems with unknown time-varying parameters. Each subsystem consists of a coupled multivariable static nonlinearity represented on a known invertible basis, followed by a matrix-polynomial dynamic block affected by colored noise and delayed input–output interconnections. The proposed scheme estimates only identifiable composite Hammerstein parameters through a decentralized recursive extended least-squares algorithm with forgetting, thereby avoiding the non-unique separation of nonlinear and linear gains. A constructive matrix Diophantine identity is established to derive an optimal multi-step predictor, leading to a GMV control law expressed as a multivariable polynomial equation in the current input. Sufficient conditions for real solvability, mean-square boundedness, and near-optimal adaptive tracking are provided using Hadamard–Lévy global-diffeomorphism, minimum-phase, small-gain, persistent-excitation, strict-positive-realness, and convex-projection arguments, and the implemented controller—inexact Newton solver with fallback and persistent dither—is itself covered by the analysis. The analysis further shows that delayed interconnections become measurable and can be exactly compensated, while robustness to basis under-modeling is explicitly quantified. Simulation results on an interconnected two-subsystem MIMO Hammerstein process with coupled cubic nonlinearities, colored noise, delayed interactions, and time-varying parameters—run in the forgetting-factor regime required by the theory, with measured persistent excitation and complete solver diagnostics—demonstrate operational-noise-floor tracking and a 2.3-fold mean-RMSE reduction relative to the strongest linear-MIMO surrogate, while a channel-wise SISO Hammerstein design fails structurally and a feedback-linearization controller with exactly known nonlinearity offers no advantage. The study further demonstrates scalability on a chain of four subsystems with size-independent per-subsystem computational cost, validates a physically motivated interconnected coupled-tank network with progressive-valve nonlinearities, and confirms agreement between the observed stability limits and the predicted small-gain boundary.



Academic Editor: Quanxin Zhu

Received: 2 June 2026

Revised: 16 June 2026

Accepted: 25 June 2026

Published: 2 July 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.

Keywords: large-scale interconnected systems; MIMO Hammerstein models; decentralized adaptive control; generalized minimum variance; recursive extended least squares; self-tuning control; stochastic time-varying systems

MSC: 93C10; 93C41; 93C55; 93D21; 93E12; 93E35

1. Introduction

Large-scale dynamical systems—power networks, chemical plants, hydraulic and petroleum processes, transportation and water-distribution systems—are characterized not only by high dimension but also by a decomposition into spatially distributed subsystems that interact through a comparatively sparse set of coupling channels [1–3]. Centralized control of such systems is frequently impractical because of the cost and unreliability of plant-wide communication, the unavailability of some measurements, and the sheer complexity of a monolithic design. Decentralized and self-tuning strategies, in which each subsystem is regulated by a local controller that adapts to unknown and possibly time-varying dynamics, have therefore attracted sustained attention [1,2,4–7].

When the local dynamics are nonlinear, block-oriented models offer an attractive compromise between fidelity and tractability. The Hammerstein model—a static nonlinearity in cascade with a linear dynamic block—captures actuator nonlinearities, valve and saturation characteristics, and many process nonlinearities, while retaining a structure amenable to identification and control [8–12]. Building on the self-tuning theory for large-scale linear stochastic systems [4,13,14], Elloumi and Kamoun [15] proposed an adaptive control scheme for large-scale systems composed of N SISO interconnected Hammerstein subsystems with known structure and unknown time-varying parameters. There, each static block is approximated by a univariate polynomial, the parameters are tracked by a recursive extended least-squares (RELS) algorithm with forgetting factor, and a minimum-variance self-tuning controller is obtained by solving a scalar polynomial equation in the current input. The approach was validated on an interconnected hydraulic process. Its scope is, however, intrinsically limited to *single-variable* subsystems and to *channel-wise* (decoupled) scalar polynomial nonlinearities; the noise enters each isolated loop through a *scalar* moving-average filter $C_i(q^{-1})e_i(k)$, rather than as a multivariable matrix process embedded in a MIMO predictor.

In parallel, the identification of *multivariable* Hammerstein systems has matured considerably [9,16–18], including very recent work on MIMO Hammerstein models with process and colored noise and on hierarchical/data-filtering estimators for multiple-input nonlinear systems [19–21]. Most notably, Li et al. [22] address a MIMO Hammerstein system in which the static block is a *combined* (coupled) multivariable nonlinearity, the linear block is an autoregressive–moving-average (ARMA) matrix model, and the measurements are corrupted by colored noise. Their two-stage scheme uses separable (step plus random) signals: step inputs decouple the linear block—identified by RELS—from the nonlinear block, which is modeled by a deep gated recurrent unit network whose convergence is established via a Lyapunov argument. The MIMO structure, the coupled nonlinearity and the colored-noise model of [22] are exactly the ingredients absent from [15]; yet [22] addresses *identification only*, not control, and treats a single (non interconnected) plant.

The present paper bridges these two lines of work. We pose and solve the *decentralized adaptive control* problem for a large-scale system composed of N interconnected MIMO Hammerstein subsystems, with matrix-polynomial dynamics, a coupled multivariable static nonlinearity, colored process noise, and delayed interconnections, under unknown slowly time-varying parameters. The contributions are the following.

1. **A unifying model and problem.** We formulate a class of interconnected MIMO Hammerstein systems that contains the SISO interconnected model of [15] and the matrix Hammerstein–ARMA *structure* of [22] as modeling special cases (the DGRU-based nonlinear identification mechanism of [22] lies outside, and is complementary to the present control-oriented formulation). The coupled static nonlinearity is represented *exactly* on a known, linearly independent basis (multivariable monomials, splines, etc.), which is linear in the parameters and includes the univariate polynomials of [15] and the coupled nonlinearities of [22].
2. **Optimal MIMO predictor with a constructive Diophantine identity.** We prove the existence and uniqueness of the matrix Diophantine identity underlying the generalized-minimum-variance design (Lemma 3) and derive the optimal multi-step predictor of the filtered output in closed form (Theorem 1), establishing that its error is a moving average of purely future innovations.
3. **A solvable adaptive GMV control law.** We show (Theorem 2) that the GMV control law reduces to a *multivariable polynomial equation in the current input vector*, expressed entirely through the *composite* parameters that are actually identifiable; no separation of the Hammerstein product is required. The SISO quadratic law of [15] is recovered as a special case.
4. **Solvability, layered and adaptive closed-loop analysis.** We give verifiable conditions under which the multivariable control equation has a unique real solution (Proposition 2, via a global-diffeomorphism argument). We prove (Theorem 3) that, because the structural interaction delays render the coupling signals measurable, the GMV law *exactly compensates* the interconnections, so the regulated outputs are mean-square bounded and minimum-variance optimal under a Schur weighting, while the control and internal signals remain bounded under a composite minimum-phase/small-gain condition of Šiljak type—with the boundedness argument carried out through explicit linear-growth bounds rather than mere continuity. We then establish a practical (sufficient-condition) *adaptive* closed-loop theorem (Theorem 4): under persistent excitation, parameter projection and slow variation, the certainty-equivalence loop is mean-square bounded and its tracking error converges to a neighborhood of the optimal floor whose radius is set by the estimation error and the parameter drift. A robustness bound for basis under-modeling is also given (Proposition 3).
5. **Validation.** A decentralized certainty-equivalence algorithm is validated on a two-subsystem MIMO interconnected process with a coupled cubic nonlinearity, colored noise, interconnection delays and time-varying parameters, operated in the forgetting-factor regime $\lambda_i^I < 1$ required by Theorem 4, with a persistent dither whose excitation effect is monitored through the minimum eigenvalue of the windowed information matrix, and with full solver diagnostics (iteration counts, residuals, failure events) so that the implemented controller coincides with the analyzed one (Proposition 4). The scheme attains the irreducible noise floor and, over a 30-seed Monte Carlo study, reduces the mean tracking root-mean-square error about fivefold relative to a linear-MIMO controller. Ablations (removing interconnection or colored-noise compensation), a nonzero-delay ($d_i = 1$) experiment exercising the nontrivial predictor, and seed-averaged coupling-strength and noise-variance sweeps—the former confirming the predicted small-gain stability boundary—complete the core study. The revised study further includes a comparison with nonlinear-aware and isolation baselines (the SISO decentralized Hammerstein controller of [15] applied channel-by-channel, a feedback-linearization-style controller with known nonlinearity, and a non-adaptive oracle with true parameters), a scalability experiment on a chain of $N = 4$ subsystems (with per-subsystem computational cost shown to be independent

of N up to $N = 8$), and a physically motivated interconnected coupled-tank network with progressive-valve actuator nonlinearities.

We emphasize the precise scope of the contribution. The novelty lies in the *control-synthesis* layer: we do not propose a new identification algorithm, and we do not reproduce the deep-network nonlinear modeling of [22]. Rather, we import the MIMO, coupled-nonlinearity, and colored-noise *modeling ingredients* of [22] into the decentralized interconnected *control* architecture of [15], and—crucially—represent the coupled nonlinearity on a known invertible basis so that the resulting GMV law is a closed-form polynomial equation (Remark 8). To leave no ambiguity about what is claimed as new, we state the precise technical step that, to the best of our knowledge, is unavailable in the existing literature: *the reduction of the decentralized GMV problem for interconnected MIMO Hammerstein subsystems with colored noise to a single multivariable polynomial Equation (19) posed directly in the identifiable composite parameters $\rho_i = B_{i,1}\Xi_i$, together with verifiable global-solvability conditions (Proposition 2), an exact-compensation stability analysis (Theorem 3) and a quantitative certainty-equivalence guarantee (Theorem 4) for that law*. Each surrounding ingredient is indeed classical—decentralized GMV for *linear* multivariable systems [4,13,14], scalar Hammerstein self-tuning control [15], and MIMO Hammerstein *identification* [16,19,22]—but none of these references contains a control law for the present model class: [15] yields a *scalar* polynomial in each u_i and cannot represent cross-channel nonlinear coupling; Refs. [13,14] require a *linear* input channel; and [22] contains no controller at all. The dimension-by-dimension technical comparison in Table A1 (Appendix A) makes the delta to each source framework explicit row by row; we have also tempered the wording of the contribution list above accordingly (“a new combination with new guarantees”, not a new control principle). The proposed model recovers the SISO interconnected model of [15] and the matrix Hammerstein–ARMA structure of [22] as special cases, but the deep learning nonlinear block contribution of [22] remains outside, and complementary to, the present control-oriented formulation.

The remainder of the paper is organized as follows. Section 2 fixes notation, states the model and the standing assumptions, and recalls the background results on which the design rests. Section 3 contains the main results—regression form, predictor, control law, and stability analysis—together with explanatory remarks and the adaptive algorithm. Section 4 reports the simulation study, and Section 5 concludes.

2. Problem Formulation and Preliminaries

Throughout, q^{-1} denotes the backward-shift operator, $q^{-1}x(k) = x(k-1)$. For a matrix M , M^T is its transpose, $\|M\|$ the spectral norm and $\|M\|_F$ the Frobenius norm; I_n is the $n \times n$ identity and $\rho(\cdot)$ the spectral radius. A square matrix polynomial $M(q^{-1}) = \sum_{\ell \geq 0} M_\ell q^{-\ell}$ is *monic* if $M_0 = I$ and *Schur* (stable) if $\det M(z^{-1}) \neq 0$ for all $|z| \geq 1$. A stochastic signal x is *mean-square (MS) bounded* if $\sup_k \mathbb{E}\|x(k)\|^2 < \infty$. We write $\mathcal{F}_k$ for the σ -algebra generated by all data available up to and including time k .

A large-scale system S is composed of N interconnected subsystems $S_1, \dots, S_N$. Subsystem S_i has input $u_i(k) \in \mathbb{R}^{m_i}$ and output $y_i(k) \in \mathbb{R}^{p_i}$ and is a MIMO Hammerstein model: a combined static nonlinearity $f_i : \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{m_i}$ in cascade with a linear dynamic block. Following the block-oriented identification literature, the static maps are represented *exactly* on known, linearly independent basis functions,

$$v_i(k) = f_i(u_i(k)) = \Xi_i \Phi_i(u_i(k)), \quad w_j(k) = g_j(y_j(k)) = \Gamma_j \Psi_j(y_j(k)), \quad (1)$$

where $\Phi_i : \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{L_i}$ and $\Psi_j : \mathbb{R}^{p_j} \rightarrow \mathbb{R}^{M_j}$ are known vector bases whose first block is the identity map (so that the linear term is present), and $\Xi_i \in \mathbb{R}^{m_i \times L_i}$, $\Gamma_j \in \mathbb{R}^{p_j \times M_j}$ are unknown coefficient matrices. The internal signals $v_i(k), w_j(k)$ are not measurable.

The dynamic block of S_i , including the interconnections with the neighboring subsystems and the colored noise, is

$$A_i(q^{-1}, k) y_i(k) = q^{-d_i} B_i(q^{-1}, k) v_i(k) + \sum_{\substack{j=1 \\ j \neq i}}^N \left[q^{-d_{ij}} B_{ij}(q^{-1}, k) v_j(k) + q^{-t_{ij}} A_{ij}(q^{-1}, k) w_j(k) \right] + C_i(q^{-1}) e_i(k), \quad (2)$$

with matrix polynomials

$$\begin{aligned} A_i(q^{-1}, k) &= I_{p_i} + \sum_{\ell=1}^{n_{A_i}} A_{i,\ell}(k) q^{-\ell}, & B_i(q^{-1}, k) &= \sum_{\ell=1}^{n_{B_i}} B_{i,\ell}(k) q^{-\ell}, \\ B_{ij}(q^{-1}, k) &= \sum_{\ell=1}^{n_{B_{ij}}} B_{ij,\ell}(k) q^{-\ell}, & A_{ij}(q^{-1}, k) &= \sum_{\ell=0}^{n_{A_{ij}}} A_{ij,\ell}(k) q^{-\ell}, \\ C_i(q^{-1}) &= I_{p_i} + \sum_{\ell=1}^{n_{C_i}} C_{i,\ell} q^{-\ell}, \end{aligned} \quad (3)$$

where $A_{i,\ell}, C_{i,\ell} \in \mathbb{R}^{p_i \times p_i}$, $B_{i,\ell} \in \mathbb{R}^{p_i \times m_i}$, $B_{ij,\ell} \in \mathbb{R}^{p_i \times m_j}$, $A_{ij,\ell} \in \mathbb{R}^{p_i \times p_j}$; d_i is the intrinsic input delay of S_i , and d_{ij}, t_{ij} are the input- and output-interaction delays from S_j to S_i . The sequence $\{e_i(k)\} \subset \mathbb{R}^{p_i}$ is the process/measurement noise. Model (1)–(2) reduces to the SISO interconnected Hammerstein model of [15] when $m_i = p_i = 1$ and the bases are univariate monomials, and to the (single) MIMO Hammerstein ARMA model of [22] when $N = 1$.

Remark 1. Substituting (1) into (2) multiplies the unknown dynamic coefficients by the unknown nonlinearity coefficients, e.g., $B_{i,\ell}(k) \Xi_i$. As is standard for Hammerstein models [10,11,23], these products cannot be factored uniquely (any nonsingular common factor may be exchanged between the two blocks). We therefore treat the products as composite parameters and estimate them directly (the over-parameterization approach), fixing the scale ambiguity by the normalization in Assumption 5. Crucially—and unlike a scheme that must reconstruct f_i explicitly—the GMV control law of Section 3 is expressible entirely through these composite parameters; see Theorem 2.

The standing assumptions generalize those of [15,22].

Assumption 1 (Known structure). The orders $n_{A_i}, n_{B_i}, n_{B_{ij}}, n_{A_{ij}}, n_{C_i}$, the delays d_i, d_{ij}, t_{ij} , and the basis maps Φ_i, Ψ_j are known. The subsystems are square, $m_i = p_i$.

Assumption 2 (Well-posedness). $A_i(q^{-1}, k)$ and $C_i(q^{-1})$ are monic for every k , and the leading input matrix $B_{i,1}(k)$ is nonsingular and uniformly bounded together with its inverse.

Assumption 3 (Structural interaction delays). $d_i < d_{ij}$ and $d_i < t_{ij}$ for all $j \neq i$.

Assumption 4 (Noise). $\{e_i(k)\}$ is a zero-mean i.i.d. Gaussian sequence with covariance $\Sigma_i \succ 0$, $\mathbb{E}[e_i(k) \mid \mathcal{F}_{k-1}] = 0$; the sequences $\{e_i\}_{i=1}^N$ are mutually independent and independent of the reference signals. The noise polynomial $C_i(q^{-1})$ is Schur.

Assumption 5 (Identifiability and slow variation). The composite parameter set is identifiable under a fixed normalization of one leading basis coefficient (e.g., the leading scalar gain of each linear term is set to unity), cf. [10,15,24]. The parameters vary slowly, $\|\Theta_i(k) - \Theta_i(k-1)\|_F \leq \mu_\theta$ for all k , with $\mu_\theta \geq 0$ small; the white-noise polynomial C_i is constant.

Assumption 6 (Solvable nonlinearity and growth). For each i the composite static map $h_i := B_{i,1}(k)\Xi_i\Phi_i = B_{i,1}(k)f_i : \mathcal{U}_i \rightarrow \mathbb{R}^{m_i}$ is, uniformly over k , a C^1 map whose Jacobian satisfies the uniform singular-value lower bound

$$\sigma_{\min}\left(\frac{\partial h_i}{\partial u}(u, k)\right) \geq \underline{\sigma}_i > 0 \quad \text{for all } u \in \mathcal{U}_i, k, \quad (4)$$

on the operating domain $\mathcal{U}_i \subseteq \mathbb{R}^{m_i}$. One of the following two regularity regimes holds.

- (I) (Global.) $\mathcal{U}_i = \mathbb{R}^{m_i}$ and (4) holds on all of $\mathbb{R}^{m_i}$. Then, by the Hadamard–Lévy global inverse function theorem [25,26], h_i is a C^1 -diffeomorphism of $\mathbb{R}^{m_i}$ onto $\mathbb{R}^{m_i}$ —properness need not be assumed separately, since the uniform bound (4) implies it.
- (II) (Compact.) $\mathcal{U}_i$ is compact and convex, (4) holds on $\mathcal{U}_i$, and $h_i(\cdot, k)$ is injective on $\mathcal{U}_i$, uniformly in k . We stress that injectivity is an additional hypothesis here: for $m_i > 1$, a nonsingular Jacobian on a compact set does not imply injectivity, so it must be assumed (or verified). A verifiable sufficient condition is the existence of a constant nonsingular matrix M_i and $\delta_i > 0$ such that $M_i \frac{\partial h_i}{\partial u}(u, k) + (M_i \frac{\partial h_i}{\partial u}(u, k))^T \succeq 2\delta_i I$ on $\mathcal{U}_i$ (then $M_i h_i$ is strongly monotone on the convex $\mathcal{U}_i$, hence injective). In this regime h_i is a C^1 -diffeomorphism of $\mathcal{U}_i$ onto its image $h_i(\mathcal{U}_i)$ only, and the control Equation (19) is uniquely solvable if and only if its right-hand side lies in $h_i(\mathcal{U}_i)$; right-hand sides outside the range are handled by the solver safeguard analyzed in Proposition 4.

In particular $\rho_i(k) := B_{i,1}(k)\Xi_i$ has full row rank. Moreover g_j and f_i^{-1} have linear growth: there exist constants $a_j, b_j, \alpha_i, \beta_i \geq 0$ with

$$\|g_j(y)\| \leq a_j + b_j\|y\|, \quad \|f_i^{-1}(v)\| \leq \alpha_i + \beta_i\|v\|, \quad \forall y, v. \quad (5)$$

Assumption 7 (Minimum phase and small gain). Each subsystem is minimum phase: the delay-stripped input polynomial $\beta_i(q^{-1}, k) := q q^{-d_i} B_i(q^{-1}, k) = B_{i,1}(k) + B_{i,2}(k)q^{-1} + \dots$ (with invertible leading coefficient $B_{i,1}$, cf. Section 3) is Schur, i.e., $\det \beta_i(z^{-1}, k) \neq 0$ for $|z| \geq 1$. Consequently, the delay-normalized diagonal operator $\mathcal{D}(q^{-1}, k)$ in Theorem 3(b) is boundedly (BIBO) invertible, and the composite interconnection operator $\mathcal{B}$ is boundedly invertible; a sufficient condition is the decentralized small-gain bound (23).

Assumption 8 (Reference and weighting). The reference $r_i(k) \in \mathbb{R}^{p_i}$ is bounded and known $d_i + 1$ steps ahead. The GMV weighting $P_i(q^{-1}) = p_i(q^{-1})I_{p_i}$ is a monic scalar Schur polynomial, and $R_i(q^{-1})$ is a polynomial matrix weighting on the reference.

Assumption 9 (Persistent excitation and projection). The information vector is persistently exciting: there exist $L \in \mathbb{N}$ and $0 < \alpha \leq \beta$ with $\alpha I \preceq \sum_{s=k+1}^{k+L} \bar{\psi}_i(s) \bar{\psi}_i^T(s) \preceq \beta I$ for all k and i . There is a known compact convex admissible set $\mathcal{C}_i$ with the two properties: (i) the true parameter trajectory satisfies $\Theta_i(k) \in \mathcal{C}_i$ for all k , and (ii) $\mathcal{C}_i$ is contained in the open set $\mathcal{O}_i$ of parameter values for which $\hat{A}_i, \hat{C}_i$ are Schur, $\hat{\beta}_i$ is Schur (minimum phase), $\sigma_{\min}(\hat{B}_{i,1}) \geq \underline{b} > 0$, and the diffeomorphism bound (4) holds. The recursive estimates are confined to $\mathcal{C}_i$ by the Euclidean projection $\mathcal{P}_{\mathcal{C}_i}$, which—because $\mathcal{C}_i$ is closed and convex—is single-valued and nonexpansive, so that $\|\mathcal{P}_{\mathcal{C}_i}(\hat{\Theta}) - \Theta_i(k)\|_F \leq \|\hat{\Theta} - \Theta_i(k)\|_F$ whenever $\Theta_i(k) \in \mathcal{C}_i$ [5,24]. We comment on the strength of this assumption. The constraint region $\mathcal{O}_i$ is open (eigenvalues and singular values depend continuously on the coefficients) but generally non-convex, and the nearest-point map onto a non-convex set is neither single-valued nor nonexpansive; the previous version of this assumption was therefore unsafe, as one reviewer pointed out. Requiring a convex $\mathcal{C}_i \subset \mathcal{O}_i$ containing the parameter trajectory is restrictive—it amounts to prior knowledge that the (slowly varying) parameters remain in a convex neighborhood, e.g., a Frobenius ball $\{\Theta : \|\Theta - \bar{\Theta}_i\|_F \leq r_i\}$ around a nominal $\bar{\Theta}_i \in \mathcal{O}_i$ with r_i small enough that the ball stays in $\mathcal{O}_i$ (such $r_i > 0$ exists since $\mathcal{O}_i$ is open)—but it is exactly what makes the projection argument rigorous, and it is the standard resolution in projection-based

adaptive control [5,24]. The simulation of Section 4 implements C_i as a Frobenius ball, combined with a covariance ceiling (Remark 4).

Assumption 10 (Strictly positive-real noise model). The transfer matrix $H_i(z) := C_i^{-1}(z) - \frac{1}{2}I_{p_i}$ is strictly positive real on the unit circle: $H_i(e^{j\omega}) + H_i^H(e^{j\omega}) \succ 0$ for all $\omega \in [0, \pi]$. This is the classical passivity condition under which pseudo-linear regression (extended least-squares) estimators converge in the presence of colored noise [24,27,28]; it is directly verifiable for any candidate C_i by a frequency sweep, and it is verified numerically for the simulation model of Section 4 (SPR margins 0.50 and 0.53 for the two subsystems).

Assumptions 1–4 and 5 are the MIMO counterparts of those in [15] (Assumptions 1–6 therein) and are consistent with [22]; Assumption 3 is the structural delay condition that makes the interaction terms predictable (its practical scope is discussed in Remark 2 below); Assumptions 6 and 7 provide the real-solvability, growth, and minimum-phase/small-gain conditions needed for boundedness, generalizing the scalar real-root and root-location conditions of [15] to the multivariable interconnected setting; Assumption 9 is standard for recursive identification and certainty-equivalence adaptive control [5,24,27], in the convex-projection form made precise above, and Assumption 10 is the classical passivity condition for extended-least-squares estimation under colored noise. The growth condition (5) is automatic when the operating domain is compact and is satisfied by the diffeomorphic bases used in Section 4; it replaces the (insufficient) appeal to mere continuity in bounding $g_j(y_j)$ and $f_i^{-1}(v_i)$. We stress that Assumption 6 is a *design* (control-solvability) condition, not an identification requirement: it is met either by choosing a monotone/diffeomorphic basis parameterization—e.g., the cubic basis of Section 4, for which (4) is verified numerically on the operating range and on expanding boxes up to $[-100, 100]^2$ (regime (I) of Assumption 6)—or by restricting operation to a compact convex region $\mathcal{U}_i$ on which the Jacobian bound and injectivity hold (regime (II)). The singular-value bound (4), rather than a mere determinant bound, is what is needed for the Lipschitz inverse in Proposition 2 (for $m_i > 1$ a lower bound on $|\det|$ alone does not bound $\sigma_{\min}$).

Remark 2 (When does Assumption 3 hold in practice?). Assumption 3 expresses a physical timing property: each local actuator affects its own outputs faster than disturbances propagate from neighboring subsystems. It is natural in process networks coupled by material transport. (i) *Hydraulic networks*. In the three-tank process used to validate the SISO scheme of [15], the local pump acts on its own tank level within one sampling period ($d_i = 0$ at the sampling rates used there), whereas the influence of a neighboring tank must traverse a connecting pipe and valve, so $d_{ij}, t_{ij} \geq 1$; the assumption holds, and it continues to hold in the MIMO extension of that rig, in which each module carries several pumps and levels, because the inter-module paths are still physical pipes. The interconnected coupled-tank network of Section 4 is built precisely on this timing structure. (ii) *Chemical plants*. Reactor cascades, distillation trains, and heat-exchanger networks coupled through piping, buffer vessels or recycle streams exhibit transport and mixing delays between units that typically exceed the local actuation delay; choosing the sampling period shorter than the smallest inter-unit transport lag renders $d_{ij}, t_{ij} \geq d_i + 1$. (iii) *Where it fails*. In electrical power networks the tie-line (electrical) coupling acts on the same time scale as—or faster than—local governor and exciter actuation, so for electromechanical dynamics Assumption 3 generally fails; the same is true of rigidly coupled mechanical systems. In such cases, the interactions cannot be treated as measurable disturbances and exact compensation (Theorem 3) is unavailable; one must resort to robust decentralized designs in which the couplings are dominated, not canceled, e.g., [1,29]. We state this scope limitation explicitly: the present scheme targets transport-coupled process networks, not electrically or rigidly coupled ones.

Two classical facts are used repeatedly. First, the parameter adaptation relies on the recursive (extended) least-squares algorithm with forgetting factor, whose properties (boundedness of the covariance, and convergence under persistency of excitation) are established in [24,27,28]. Second, the predictor design uses the polynomial-equation (Diophantine) approach to linear quadratic-Gaussian and minimum-variance control [24,30–32]; its multivariable form, on which our development rests, is stated and proved constructively in Lemma 3.

The control objective is the stochastic tracking–regulation problem: design a *decentralized* adaptive control law $u_i(k)$, using only the information locally available to S_i , that minimizes, at each step and for each i , the generalized-minimum-variance criterion

$$J_i = \mathbb{E} \left[\|P_i(q^{-1}) y_i(k + d_i + 1) - R_i(q^{-1}) r_i(k + d_i + 1)\|^2 \mid \mathcal{F}_k \right], \quad (6)$$

thereby forcing the (filtered) output to follow the reference while rejecting the noise and the inter-subsystem disturbances.

3. Main Results

This section develops, in turn, the regression form used for identification, the decentralized RELS estimator, the matrix Diophantine identity, the optimal predictor, the GMV control law, and the closed-loop analysis. Explanatory remarks connect successive results and discuss special cases, including the recovery of [15] and the relation to [22].

We begin by casting the model in a linear regression form in the composite parameters of Remark 1.

Lemma 1 (Regression form). *Under Assumptions 1 and 2, the input–output map (2) of subsystem S_i admits, for each k , the representation*

$$y_i(k) = \Theta_i^T(k) \psi_i(k) + e_i(k), \quad (7)$$

where the composite parameter matrix $\Theta_i^T(k) \in \mathbb{R}^{p_i \times n_{\theta_i}}$ and the observation vector $\psi_i(k) \in \mathbb{R}^{n_{\theta_i}}$ are

$$\Theta_i^T(k) = [-A_{i,1}(k) \cdots -A_{i,n_{A_i}}(k) \mid B_{i,1}(k)\Xi_i \cdots B_{i,n_{B_i}}(k)\Xi_i \mid \{B_{ij,\ell}(k)\Xi_j\}_{j \neq i,\ell} \mid \{A_{ij,\ell}(k)\Gamma_j\}_{j \neq i,\ell} \mid C_{i,1} \cdots C_{i,n_{C_i}}], \quad (8)$$

$$\psi_i(k) = [y_i(k-1)^T \cdots y_i(k-n_{A_i})^T \mid \Phi_i(u_i(k-d_i-1))^T \cdots \Phi_i(u_i(k-d_i-n_{B_i}))^T \mid \{ \Phi_j(u_j(k-d_{ij}-\ell))^T \} \mid \{ \Psi_j(y_j(k-t_{ij}-\ell))^T \} \mid e_i(k-1)^T \cdots e_i(k-n_{C_i})^T]^T. \quad (9)$$

Proof. See Appendix B. $\square$

Remark 3. The products $B_{i,\ell}(k)\Xi_i$, $B_{ij,\ell}(k)\Xi_j$ and $A_{ij,\ell}(k)\Gamma_j$ are the composite parameters of Remark 1; each multiplies a distinct delayed copy of the known basis vector, so (7) is genuinely linear in the parameters and free of internal redundancy once the scale is fixed by Assumption 5. This is the multivariable generalization of the parameter vector of [15] (Equations (15) and (16)); the moving-average block $\{C_{i,\ell}\}$ is the colored-noise extension shared with [22].

The noise terms $e_i(k-\ell)$ in (9) are not measurable; replacing them by the a posteriori residuals gives the implementable information vector $\tilde{\psi}_i(k)$, in which $e_i(k-\ell)$ is replaced by $\varepsilon_i(k-\ell)$. The parameters are then tracked by a decentralized recursive extended least-squares (RELS) algorithm with a time-varying forgetting factor.

Proposition 1 (Decentralized RELS). *For each i , let $\varepsilon_i(k) = y_i(k) - \hat{\Theta}_i^T(k-1)\bar{\psi}_i(k)$ and update*

$$\hat{\Theta}_i(k) = \hat{\Theta}_i(k-1) + \frac{\Pi_i(k-1)\bar{\psi}_i(k)}{\lambda_i(k) + \bar{\psi}_i^T(k)\Pi_i(k-1)\bar{\psi}_i(k)} \varepsilon_i^T(k), \quad (10a)$$

$$\Pi_i(k) = \frac{1}{\lambda_i(k)} \left[\Pi_i(k-1) - \frac{\Pi_i(k-1)\bar{\psi}_i(k)\bar{\psi}_i^T(k)\Pi_i(k-1)}{\lambda_i(k) + \bar{\psi}_i^T(k)\Pi_i(k-1)\bar{\psi}_i(k)} \right], \quad (10b)$$

$$\lambda_i(k) = \lambda_{0i} \lambda_i(k-1) + (1 - \lambda_{0i}) \lambda_i^I, \quad 0 < \lambda_{0i} < 1, \quad 0 < \lambda_i^I \leq 1, \quad (10c)$$

with $\Pi_i(0) \succ 0$ and $\hat{\Theta}_i(0)$ given, each update being followed by the projection $\hat{\Theta}_i(k) \leftarrow \mathcal{P}_{\mathcal{C}_i}(\hat{\Theta}_i(k))$ of Assumption 9. Under Assumption 9 the covariance $\Pi_i(k)$ remains uniformly bounded, $0 \prec \underline{\pi}I \preceq \Pi_i(k)^{-1} \preceq \bar{\pi}I$, and the estimates converge in the sense of [24,27,28]; a quantitative estimation-error bound, under verifiable assumptions, is given in Lemma 2 below. During phases in which the excitation is weaker than the persistent-excitation level of Assumption 9, a forgetting factor $\lambda_i^I < 1$ can produce covariance wind-up; the implementation therefore enforces the stated upper bound on Π_i explicitly, see Remark 4.

Proof. See Appendix B. $\square$

The recursion (10) is the matrix form of the estimator of [15] (Equations (18) and (21)); $\Pi_i(k)$ denotes the estimation covariance (not to be confused with the weighting P_i). The algorithm is decentralized: each S_i uses only its own signals and the delayed neighbor signals already present in (9), and the extended regressor $\bar{\psi}_i(k)$ replaces the unmeasurable $e_i(k-\ell)$ by the residuals $\varepsilon_i(k-\ell)$.

Remark 4 (Implemented estimator safeguards). *Two standard safeguards make the implemented estimator coincide with the one analyzed. First, each update is followed by the Euclidean projection onto the convex admissible set $\mathcal{C}_i$ of Assumption 9; since $\Theta_i(k) \in \mathcal{C}_i$ and the projection is nonexpansive, this step can only decrease the estimation error and is compatible with every inequality used in Lemma 2. Second, with $\lambda_i^I < 1$ the information matrix is discounted geometrically, so during low-excitation phases the covariance can grow without bound (wind-up); the implementation enforces the uniform bound $\Pi_i(k) \preceq \bar{\pi}I$ of Proposition 1 directly, by rescaling $\Pi_i(k)$ whenever its trace exceeds a cap (a simplified form of the exponential-forgetting-and-resetting algorithm of [33]). Both devices are exactly the boundedness properties that the analysis assumes, so no property is assumed in the proofs that the implementation does not enforce; the values used in Section 4 are reported there.*

The next lemma is the quantitative estimation result on which the adaptive analysis rests. We state it under explicit, verifiable hypotheses, because the setting actually studied here—a regressor containing *estimated* residuals, *colored* noise, and *drifting* parameters—is not covered by off-the-shelf RLS theory: the residual substitution requires the passivity (SPR) condition of Assumption 10, the drift requires an explicit perturbation term, and the projection requires the convexity established in Assumption 9.

Lemma 2 (Estimation-error bound for the decentralized RELS). *Let Assumptions 1–5 hold, let the persistent-excitation and convex-projection conditions of Assumption 9 hold with constants (α, β, L) , let the SPR condition of Assumption 10 hold with margin $\zeta_{\text{SPR}} > 0$, let the implemented information vector be mean-square bounded, $\sup_k \mathbb{E} \|\bar{\psi}_i(k)\|^2 \leq c_\psi < \infty$, and let*

the forgetting factor satisfy $\lambda_i(k) \rightarrow \lambda^1 \in (0, 1)$. Define $\tilde{\Theta}_i(k) := \hat{\Theta}_i(k) - \Theta_i(k)$ and $\bar{e}^2 := \limsup_{k \rightarrow \infty} \max_i \mathbb{E} \|\tilde{\Theta}_i(k)\|_F^2$. Then

$$\bar{e}^2 \leq \kappa_1 \sigma_e^2 + \kappa_2 \mu_\theta^2, \quad (11)$$

with finite constants κ_1, κ_2 depending only on $(\alpha, \beta, L, \lambda^1, \underline{\pi}, \bar{\pi}, c_\psi, \zeta_{\text{SPR}})$; in particular $\bar{e} \rightarrow 0$ as $(\sigma_e, \mu_\theta) \rightarrow 0$.

Proof. See Appendix B. $\square$

Remark 5 (Scope of Lemma 2). The mean-square boundedness of $\bar{\psi}_i$ is stated as a hypothesis here and is discharged in Theorem 4(c), where boundedness of all closed-loop signals is established by a joint contraction argument; this is the usual structure of total-stability proofs, in which the estimation bound and the signal-boundedness bound hold simultaneously rather than sequentially [5,34]. The persistent-excitation constant α is not merely assumed in the experiments: Section 4 measures the minimum eigenvalue of the windowed information matrix along the closed-loop trajectory, with and without dither.

We now turn to the controller. The GMV design rests on the following matrix Diophantine identity, which we state and prove constructively.

Lemma 3 (Matrix Diophantine identity). Let $A_i(q^{-1})$ be monic of degree n_{A_i} , let $p_i(q^{-1})$ be a monic scalar polynomial of degree n_p , let $C_i(q^{-1})$ be monic of degree n_{C_i} , and let $d_i \geq 0$. Then there exist unique matrix polynomials $F_i(q^{-1})$, monic of degree d_i , and $G_i(q^{-1})$ of degree $n_{G_i} = \max\{n_{A_i} - 1, n_{C_i} + n_p - d_i - 1\}$, such that

$$p_i(q^{-1}) C_i(q^{-1}) = A_i(q^{-1}) F_i(q^{-1}) + q^{-(d_i+1)} G_i(q^{-1}). \quad (12)$$

Proof. See Appendix B. $\square$

Define $\beta_i(q^{-1}, k) := \sum_{\ell=1}^{n_{B_i}} B_{i,\ell}(k) q^{-(\ell-1)}$, so that $q^{-d_i} B_i(q^{-1}, k) = q^{-(d_i+1)} \beta_i(q^{-1}, k)$ and the leading coefficient of β_i is $B_{i,1}(k)$. The interaction term in (2) is denoted $Y_i(k) := \sum_{j \neq i} [q^{-d_{ij}} B_{ij}(q^{-1}, k) v_j(k) + q^{-t_{ij}} A_{ij}(q^{-1}, k) w_j(k)]$, and its value at the prediction instant by $Y_i^+(k) := Y_i(k + d_i + 1)$.

Theorem 1 (Optimal multi-step predictor). Let F_i, G_i solve (12) with p_i the scalar weighting of Assumption 8. Under Assumptions 1–4 and 8, the filtered output of S_i obeys

$$P_i(q^{-1}) y_i(k + d_i + 1) = B_{i,1}(k) v_i(k) + \xi_i(k) + F_i(q^{-1}, k) e_i(k + d_i + 1), \quad (13)$$

where:

- (i) $Y_i^+(k)$ is $\mathcal{F}_k$ -measurable, by Assumption 3;
- (ii) $F_i(q^{-1}, k) e_i(k + d_i + 1) = \sum_{\ell=0}^{d_i} F_{i,\ell}(k) e_i(k + d_i + 1 - \ell)$ is a zero-mean moving average of the future innovations $e_i(k + 1), \dots, e_i(k + d_i + 1)$, independent of $\mathcal{F}_k$;
- (iii) $\xi_i(k)$ is $\mathcal{F}_k$ -measurable and independent of $u_i(k)$.

Consequently, the minimum-variance $(d_i + 1)$ -step predictor is

$$\widehat{P_i y_i}(k + d_i + 1 | k) = B_{i,1}(k) v_i(k) + \xi_i(k), \quad (14)$$

which is realized by the recursion

$$A_i(q^{-1}, k) \widehat{P_i y_i}(k + d_i + 1 | k) = p_i(q^{-1}) \beta_i(q^{-1}, k) v_i(k) + p_i(q^{-1}) Y_i^+(k) + G_i(q^{-1}, k) e_i(k), \quad (15)$$

and the irreducible prediction-error covariance is $\sum_{\ell=0}^{d_i} F_{i,\ell}(k) \Sigma_i F_{i,\ell}^T(k)$.

Proof. All manipulations below are identities in the commutative ring $\mathbb{R}^{p_i \times p_i}[[q^{-1}]]$ of matrix formal power series in q^{-1} with matrix coefficients acting on the fixed signal e_i (equivalently, identities between the generating series of the involved sequences); since $A_i(q^{-1}, k)$ is monic, its constant term is I and it is a unit of this ring, with $A_i^{-1} = \sum_{r \geq 0} H_{i,r} q^{-1r}$, $H_{i,0} = I$, defined by the convolution $\sum_{s=0}^r A_{i,s} H_{i,r-s} = \delta_{r0} I$. No claim of *stability* of A_i^{-1} is needed here: the predictor is realized by the *finite* recursion (15), which involves A_i itself, not its inverse. With this understanding, (2) gives $y_i(k) = A_i^{-1} [q^{-d_i} B_i v_i(k) + Y_i(k) + C_i e_i(k)]$. Advancing by $d_i + 1$ steps and using $q^{-d_i} B_i v_i(k + d_i + 1) = q^{-(d_i+1)} \beta_i v_i(k + d_i + 1) = \beta_i v_i(k)$,

$$y_i(k + d_i + 1) = A_i^{-1} \beta_i v_i(k) + A_i^{-1} Y_i^+(k) + A_i^{-1} C_i e_i(k + d_i + 1). \quad (16)$$

Applying $P_i = p_i I$ (a scalar operator, which commutes with all matrix polynomials) and invoking (12) in the form $A_i^{-1} p_i C_i = F_i + q^{-(d_i+1)} A_i^{-1} G_i$, the noise term becomes

$$A_i^{-1} p_i C_i e_i(k + d_i + 1) = F_i e_i(k + d_i + 1) + A_i^{-1} G_i e_i(k).$$

Hence

$$P_i y_i(k + d_i + 1) = \underbrace{A_i^{-1} p_i \beta_i v_i(k) + A_i^{-1} p_i Y_i^+(k) + A_i^{-1} G_i e_i(k)}_{=: \eta_i(k)} + F_i e_i(k + d_i + 1). \quad (17)$$

Because $F_i = \sum_{\ell=0}^{d_i} F_{i,\ell} q^{-1\ell}$ is monic of degree d_i , $F_i e_i(k + d_i + 1) = \sum_{\ell=0}^{d_i} F_{i,\ell} e_i(k + d_i + 1 - \ell)$ involves only the indices $k + d_i + 1 - \ell$ for $\ell = 0, \dots, d_i$, i.e., $e_i(k + 1), \dots, e_i(k + d_i + 1)$, all strictly in the future of k ; by Assumption 4 these are zero-mean and independent of $\mathcal{F}_k$, proving (ii). For (i), expand

$$Y_i^+(k) = Y_i(k + d_i + 1) = \sum_{j \neq i} \left[\sum_{\ell=1}^{n_{B_{ij}}} B_{ij,\ell} v_j(k + d_i + 1 - d_{ij} - \ell) + \sum_{\ell=0}^{n_{A_{ij}}} A_{ij,\ell} w_j(k + d_i + 1 - t_{ij} - \ell) \right].$$

The latest input-coupling index is $k + d_i + 1 - d_{ij} - 1 = k + d_i - d_{ij} \leq k - 1$ (using $d_{ij} \geq d_i + 1$ from Assumption 3), and the latest output-coupling index is $k + d_i + 1 - t_{ij} \leq k$ (using $t_{ij} \geq d_i + 1$). Hence, every signal in $Y_i^+(k)$ is dated at or before k and is $\mathcal{F}_k$ -measurable. The terms $A_i^{-1} p_i \beta_i v_i(k)$ and $A_i^{-1} G_i e_i(k)$ likewise depend only on $\{u_i(s), y_i(s), e_i(s)\}_{s \leq k}$ (the latter recoverable from $e_i(k) = C_i^{-1} [A_i y_i - q^{-d_i} B_i v_i - Y_i](k)$, C_i being a unit); so $\eta_i(k)$ is $\mathcal{F}_k$ -measurable. The leading coefficient of $A_i^{-1} p_i \beta_i$ is $I \cdot 1 \cdot B_{i,1}(k) = B_{i,1}(k)$; writing $A_i^{-1} p_i \beta_i v_i(k) = B_{i,1}(k) v_i(k) + [(A_i^{-1} p_i \beta_i - B_{i,1}) v_i(k)]$ and absorbing the bracket (which depends only on $v_i(k - 1), v_i(k - 2), \dots$) and the remaining $\mathcal{F}_k$ -measurable terms of η_i into

$$\xi_i(k) := (A_i^{-1} p_i \beta_i - B_{i,1}(k)) v_i(k) + A_i^{-1} p_i Y_i^+(k) + A_i^{-1} G_i e_i(k),$$

we obtain (13) with $\xi_i(k)$ $\mathcal{F}_k$ -measurable and independent of $u_i(k)$, proving (iii). Taking $\mathbb{E}[\cdot \mid \mathcal{F}_k]$ in (13) and using (ii) yields (14), which is the MMSE predictor since the error $F_i e_i(k + d_i + 1)$ is orthogonal to every $\mathcal{F}_k$ -measurable estimator. Equation (15) follows by left-multiplying $\widehat{P_i y_i}(k + d_i + 1 \mid k) = \eta_i(k)$ by A_i and using the definition of η_i . Finally, by independence of the $e_i(k + d_i + 1 - \ell)$,

$$\mathbb{E}[F_i e_i(k + d_i + 1) e_i^T(k + d_i + 1) F_i^T] = \sum_{\ell=0}^{d_i} F_{i,\ell} \Sigma_i F_{i,\ell}^T. \quad \square$$

Remark 6. Statement (i) is the structural heart of the interconnected design: because the interaction delays exceed the intrinsic delay (Assumption 3), at time k , every interaction signal that influences $y_i(k + d_i + 1)$ has already occurred, so the couplings act as measurable disturbances that the predictor (15)—and hence the controller below—can cancel exactly. This generalizes the role of the assumption $d_i < d_{ij}, t_{ij}$ in [15] (Assumption 3 and Equation (33)) to the MIMO setting. The implementable predictor replaces the unmeasured $e_i(k)$ in (15) by the RELS residual $\varepsilon_i(k)$.

The GMV control law follows by minimizing (6).

Theorem 2 (Adaptive GMV control law). Under Assumptions 1–8, the criterion (6) satisfies

$$J_i = \|B_{i,1}(k)v_i(k) + \xi_i(k) - R_i(q^{-1})r_i(k + d_i + 1)\|^2 + \text{tr}\left(\sum_{\ell=0}^{d_i} F_{i,\ell}\Sigma_i F_{i,\ell}^T\right), \quad (18)$$

and is minimized over $u_i(k)$ by any solution of the multivariable polynomial equation

$$\rho_i(k)\Phi_i(u_i(k)) = b_i(k), \quad \rho_i(k) := B_{i,1}(k)\Xi_i, \quad b_i(k) := R_i(q^{-1})r_i(k + d_i + 1) - \xi_i(k), \quad (19)$$

whenever such a (real) solution exists; under Assumption 6, it does. The minimal cost is $J_i^* = \text{tr}\left(\sum_{\ell=0}^{d_i} F_{i,\ell}\Sigma_i F_{i,\ell}^T\right)$. Equation (19) involves only the composite parameters $\rho_i(k)$ and the $\mathcal{F}_k$ -measurable quantity $\xi_i(k)$, both available from (10) and (15); no factorization of ρ_i into $B_{i,1}$ and Ξ_i is required.

Proof. By Assumption 8, $R_i r_i(k + d_i + 1)$ is $\mathcal{F}_k$ -measurable. Writing $\sigma_i(k + d_i + 1) := P_i y_i(k + d_i + 1) - R_i r_i(k + d_i + 1)$ and substituting (13),

$$\sigma_i(k + d_i + 1) = \underbrace{B_{i,1}(k)v_i(k) + \xi_i(k) - R_i r_i(k + d_i + 1) + F_i e_i(k + d_i + 1)}_{=: \hat{\sigma}_i(k + d_i + 1 | k) \text{ (}\mathcal{F}_k\text{-measurable)}}.$$

The two terms are orthogonal (Theorem 1(ii)–(iii)), so

$$J_i = \mathbb{E}[\|\sigma_i(k + d_i + 1)\|^2 | \mathcal{F}_k] = \|\hat{\sigma}_i(k + d_i + 1 | k)\|^2 + \text{tr}\left(\sum_{\ell=0}^{d_i} F_{i,\ell}\Sigma_i F_{i,\ell}^T\right),$$

which is (18); the second term does not depend on $u_i(k)$. Hence, J_i is minimized iff $\hat{\sigma}_i(k + d_i + 1 | k) = 0$, i.e., $B_{i,1}(k)\Xi_i\Phi_i(u_i(k)) = R_i r_i(k + d_i + 1) - \xi_i(k)$, which is (19). By Assumption 6, $\rho_i(k) = B_{i,1}(k)\Xi_i$ has full row rank and f_i is a proper homeomorphism, so the value $b_i(k)$ lies in the range of $u \mapsto \rho_i(k)\Phi_i(u)$ on $\mathcal{U}_i$ and a real solution exists. Substituting $\hat{\sigma}_i = 0$ leaves $J_i^* = \text{tr}\left(\sum_{\ell} F_{i,\ell}\Sigma_i F_{i,\ell}^T\right)$. Since $\rho_i(k)$ is precisely the leading input block of $\hat{\Theta}_i(k)$ in (8) and $\xi_i(k)$ is generated by (15) from measured signals, neighbor data and $\varepsilon_i(k)$ —all in terms of composite parameters—no separation of $B_{i,1}$ and Ξ_i is needed. $\square$

The next proposition supplies the verifiable conditions, promised after Theorem 2, under which (19) is *uniquely* solvable, and quantifies the Lipschitz dependence of the solution on the data—a property used in the adaptive analysis of Theorem 4.

Proposition 2 (Unique solvability and Lipschitz control map). Let Assumptions 1, 2 and 6 hold, and write $h_i := \rho_i(k)\Phi_i = B_{i,1}(k)\Xi_i\Phi_i : \mathcal{U}_i \rightarrow \mathbb{R}^{m_i}$.

- (I) Under regime (I) of Assumption 6 ($\mathcal{U}_i = \mathbb{R}^{m_i}$, uniform bound (4) on all of $\mathbb{R}^{m_i}$), h_i is a C^1 -diffeomorphism of $\mathbb{R}^{m_i}$ onto $\mathbb{R}^{m_i}$ by the Hadamard–Lévy theorem, and the control Equation (19) has the unique real solution $u_i(k) = h_i^{-1}(b_i(k))$ for every $b_i(k) \in \mathbb{R}^{m_i}$.

- (II) Under regime (II) ($\mathcal{U}_i$ compact convex, bound (4) and injectivity on $\mathcal{U}_i$), h_i is a C^1 -diffeomorphism of $\mathcal{U}_i$ onto its image $h_i(\mathcal{U}_i)$, and (19) has a unique solution in $\mathcal{U}_i$ if and only if $b_i(k) \in h_i(\mathcal{U}_i)$; no claim is made for right-hand sides outside the range, which are handled by the safeguard of Proposition 4. (This replaces, and substantially weakens, the original statement, which incorrectly asserted global unique solvability from Jacobian nonsingularity on a compact operating set alone.)

Moreover, in either regime, on the convex admissible set C_i of Assumption 9 and over any bounded regressor set $\{\|\Phi_i(\cdot)\| \leq \Phi_{\max}\}$, the solution map is Lipschitz (in regime (II): for right-hand sides ranging in $h_i(\mathcal{U}_i)$): there are constants $L_b, L_\rho < \infty$, depending only on $\underline{\sigma} := \inf_u \sigma_{\min}(\partial h_i / \partial u) > 0$ and $\Phi_{\max}$, with

$$\|h_i^{-1}(b) - h_i^{-1}(b')\| \leq L_b \|b - b'\|, \quad \|h_i^{-1}(b; \rho) - h_i^{-1}(b; \rho')\| \leq L_\rho \|\rho - \rho'\|_F. \quad (20)$$

The second bound is taken along the line segment $[\rho, \rho'] \subset C_i$, which lies in C_i because C_i is convex (Assumption 9); without that convexity the segment argument would be invalid.

Proof. See Appendix B. $\square$

Remark 7 (Special cases and recovery of [15]). With a linear basis $\Phi_i(u) = u$, (19) is the linear allocation $u_i(k) = \rho_i^{-1}(k)b_i(k)$ (with ρ_i square nonsingular). In the SISO case $m_i = p_i = 1$ with $\Phi_i(u) = [u \ u^2]^T$ and $\rho_i = [\rho_{i,1} \ \rho_{i,2}]$, (19) becomes the scalar quadratic $\rho_{i,2}u_i^2(k) + \rho_{i,1}u_i(k) - b_i(k) = 0$, whose real roots are exactly the control law of [15] (Equations (57)–(63)); the present theorem thus contains that result and extends it to coupled multivariable nonlinearities. More generally, Φ_i may be any known basis—multivariable monomials (capturing the combined cross-channel nonlinearities of [22]), splines, or radial functions—so long as Assumption 6 holds. A clarification on closed-form solvability is in order, replacing a misleading statement in the original version. Cardano and Ferrari radicals solve scalar univariate cubic and quartic equations; consequently, they apply only in the SISO case $m_i = 1$ with $\deg \Phi_i \leq 4$, where they recover exactly the explicit root formulas of [15]. For $m_i > 1$, (19) is a coupled system of multivariable polynomial equations for which no such radical formulas exist, and we make no closed-form claim: “closed form” in this paper refers to the control law being given by the explicit finite-dimensional Equation (19)—as opposed to requiring the inversion of a learned network—not to a solution by radicals. The algorithm actually used to compute $u_i(k)$ is the damped Newton iteration specified in Proposition 4 and step 5 of Algorithm 1: warm start at $u_i(k-1)$, Armijo backtracking on the residual norm $\|\hat{\rho}_i \Phi_i(u) - b_i(k)\|$, stopping tolerance τ , iteration cap $n_{\max}$, and the fallback $u_i(k) = u_i(k-1)$ when no iterate meets the acceptance threshold; under Assumption 6 the Jacobian $\hat{\rho}_i \partial \Phi_i / \partial u$ is uniformly nonsingular along the iterates, and the local quadratic convergence of Newton’s method from the warm start makes the iteration terminate in a few steps in practice (solver statistics are reported in Section 4).

Remark 8 (On a deep-network nonlinearity). The MIMO Hammerstein identification scheme of [22] models the combined nonlinearity by a deep gated-recurrent-unit network. Such a model could replace the basis $f_i = \Xi_i \Phi_i$ in the plant; however, the control law (19) would then read $f_i(u_i(k)) = B_{i,1}^{-1}(k)b_i(k)$ with f_i a deep network, which is not analytically invertible and would require a numerical inversion—and a controllability guarantee on the network range—at every step. The linear-in-parameters basis representation adopted here is precisely what keeps the GMV law in the closed, polynomial form (19) while still capturing the coupled nonlinearities emphasized in [22].

It remains to certify that the closed loop is well behaved. We give a layered analysis: the regulated outputs first, then the internal signals.

Theorem 3 (Closed-loop boundedness and optimality). Suppose the control law (19) is applied with the true (frozen) parameters, and let Assumptions 1–8 hold. Then, for every i and all $k \geq k_0$:

(a) (Outputs.) The filtered tracking error reduces to the future innovation,

$$P_i(q^{-1})y_i(k) - R_i(q^{-1})r_i(k) = F_i(q^{-1}, k)e_i(k), \quad (21)$$

independently of the interconnections; hence, y_i is MS-bounded and the regulated output attains the minimum tracking-error variance $\text{tr}(\sum_{\ell=0}^{d_i} F_{i,\ell} \Sigma_i F_{i,\ell}^T)$.

(b) (Internal signals.) The internal and control signals satisfy the composite equation

$$\mathcal{B}(q^{-1}, k)\mathbf{v}(k) = \mathbf{c}(k), \quad \mathbf{v} = [v_1^T, \dots, v_N^T]^T, \quad (22)$$

where $\mathcal{B}$ has diagonal blocks $q^{-d_i}B_i(q^{-1}, k) = q^{-(d_i+1)}\beta_i(q^{-1}, k)$ and off-diagonal blocks $q^{-d_{ij}}B_{ij}(q^{-1}, k)$, and $\mathbf{c}(k)$ is MS-bounded. If $\mathcal{B}$ is boundedly invertible—in particular if each β_i is Schur (Assumption 7) and the small-gain condition

$$\sup_{|z|=1} \rho(\mathcal{D}^{-1}(z^{-1}, k)\mathcal{N}(z^{-1}, k)) < 1, \quad \mathcal{D} := \text{blockdiag}(q^{-(d_i+1)}\beta_i), \quad \mathcal{N} := \mathcal{B} - \mathcal{D}, \quad (23)$$

holds—then $\mathbf{v}$ is MS-bounded, and by Assumption 6 so are the controls $u_i = f_i^{-1}(v_i)$.

Proof. (a) The control law enforces $\hat{\sigma}_i(k + d_i + 1 | k) = 0$ (Theorem 2); substituting into the identity $\sigma_i(k + d_i + 1) = \hat{\sigma}_i(k + d_i + 1 | k) + F_i e_i(k + d_i + 1)$ gives $\sigma_i(t) = F_i e_i(t)$ for all $t \geq k_0 + d_i + 1$, which is (21). The right-hand side contains no interaction term: the couplings enter σ_i only through the measurable ξ_i , and are canceled by the control. Since $P_i = p_i I$ is monic and Schur (Assumption 8), P_i^{-1} is a stable filter, so $y_i(k) = P_i^{-1}(R_i r_i(k) + F_i e_i(k))$ is MS-bounded; the tracking-error variance equals $\mathbb{E}\|F_i e_i(k)\|^2 = \text{tr}(\sum_{\ell} F_{i,\ell} \Sigma_i F_{i,\ell}^T)$ by Theorem 1.

(b) From (21), $y_i = P_i^{-1}(R_i r_i + F_i e_i)$. Substituting into (2),

$$A_i P_i^{-1}(R_i r_i + F_i e_i) = q^{-d_i} B_i v_i + Y_i + C_i e_i.$$

By (12), $A_i F_i = p_i C_i - q^{-(d_i+1)} G_i$, so, with $P_i = p_i I$,

$$A_i P_i^{-1} F_i e_i = p_i^{-1} A_i F_i e_i = C_i e_i - q^{-(d_i+1)} p_i^{-1} G_i e_i.$$

Hence, the colored-noise terms $C_i e_i$ cancel and

$$q^{-d_i} B_i v_i + Y_i = A_i P_i^{-1} R_i r_i - q^{-(d_i+1)} p_i^{-1} G_i e_i.$$

Moving the output-coupling part of Y_i to the right and recalling $Y_i = \sum_{j \neq i} [q^{-d_{ij}} B_{ij} v_j + q^{-t_{ij}} A_{ij} w_j]$,

$$q^{-d_i} B_i v_i(k) + \sum_{j \neq i} q^{-d_{ij}} B_{ij} v_j(k) = \underbrace{A_i P_i^{-1} R_i r_i(k) - q^{-(d_i+1)} p_i^{-1} G_i e_i(k) - \sum_{j \neq i} q^{-t_{ij}} A_{ij} w_j(k)}_{=: c_i(k)}.$$

This is the i -th block row of (22). The right-hand side c_i is MS-bounded: r_i is bounded and $A_i P_i^{-1} R_i$ is a stable filter (P_i Schur); e_i is MS-bounded and $p_i^{-1} G_i$ is stable; and by part (a) y_j is MS-bounded, whence—using the *linear-growth* bound (5) rather than mere continuity— $\mathbb{E}\|w_j(k)\|^2 = \mathbb{E}\|g_j(y_j(k))\|^2 \leq 2a_j^2 + 2b_j^2 \mathbb{E}\|y_j(k)\|^2 < \infty$, so w_j is MS-bounded and $\sup_k \mathbb{E}\|c_i(k)\|^2 < \infty$. (We stress that continuity alone would *not* suffice, since an MS-

bounded signal need not be a.s. confined to a compact set; the affine bound (5) is precisely what propagates the second moment.) Stacking the N block rows gives (22).

It remains to interpret and verify the bounded invertibility of $\mathcal{B}$. Factor out the delays: each block row i has the lowest-order term $q^{-(d_i+1)}B_{i,1}v_i(k)$, so writing $\mathcal{D}(q^{-1}, k) := \text{blockdiag}_i(q^{-(d_i+1)}\beta_i(q^{-1}, k))$ and $\mathcal{N} := \mathcal{B} - \mathcal{D}$, the diagonal block $q^{-(d_i+1)}\beta_i$ has the *instantaneously invertible* leading coefficient $B_{i,1}$ (Assumption 2) and the delay-stripped factor β_i is Schur (Assumption 7). We make the meaning of inversion precise. A pure delay $q^{-(d_i+1)}$ is itself *not* causally invertible (its formal inverse is an advance), so all the small-gain reasoning is carried out on the *delay-stripped* transfer operators on the unit circle, and “bounded invertibility” of $\mathcal{B}$ is understood in the two-sided $\ell_2(\mathbb{Z})$ (bi-infinite) sense—equivalently, up to a finite start-up transient, which is immaterial for the lim sup MS-boundedness conclusion. Concretely, write $\mathcal{B}(z^{-1}) = \mathcal{D}(z^{-1})(I + \mathcal{D}^{-1}(z^{-1})\mathcal{N}(z^{-1}))$ on $|z| = 1$, where $\mathcal{D}^{-1}\mathcal{N}$ is proper and rational; since each β_i is Schur and $B_{i,1}$ is invertible, $\mathcal{D}^{-1}$ is bounded on $|z| = 1$, and the Neumann series $(I + \mathcal{D}^{-1}\mathcal{N})^{-1} = \sum_{r \geq 0} (-\mathcal{D}^{-1}\mathcal{N})^r$ converges in the $\mathcal{H}_\infty$ operator norm precisely when the small-gain condition (23), $\sup_{|z|=1} \rho(\mathcal{D}^{-1}(z^{-1})\mathcal{N}(z^{-1})) < 1$, holds. Then $\mathcal{B}^{-1} \in \mathcal{H}_\infty$ and $\mathbf{v} = \mathcal{B}^{-1}\mathbf{c}$ is MS-bounded [1,29]. Finally, by the linear-growth bound (5) on f_i^{-1} , $\mathbb{E}\|u_i(k)\|^2 = \mathbb{E}\|f_i^{-1}(v_i(k))\|^2 \leq 2\alpha_i^2 + 2\beta_i^2 \mathbb{E}\|v_i(k)\|^2 < \infty$, so the controls are MS-bounded. $\square$

Remark 9 (From frozen to adaptive). *Theorem 3 establishes certainty-equivalence (frozen-parameter) stability and optimality, and isolates the decentralized small-gain condition (23) governing the interconnections; it is the multivariable, interconnected counterpart of the root-location/real-solution analysis of [15] (§IV). We now go further and analyze the actual adaptive loop, in which $\hat{\Theta}_i$ evolves through (10) and the control solves (19) with the running estimates. The argument has three ingredients, isolated for clarity: a quantitative estimation bound (Proposition 1 and [24]), the Lipschitz control map (Proposition 2), and the BIBO stability of the ideal loop (Theorem 3). Their combination is a robustness-of-certainty-equivalence (total-stability) statement [5,34].*

Theorem 4 (Adaptive closed-loop boundedness and near-optimality: a sufficient condition). *Let Assumptions 1–10 hold with forgetting factor $0 < \lambda_i^l < 1$, and apply the decentralized certainty-equivalence controller of Algorithm 1, with $u_i^*(k)$ denoting the ideal control of Theorem 2 computed from the true parameters. Define the estimation error $\tilde{\Theta}_i(k) := \hat{\Theta}_i(k) - \Theta_i(k)$ and $\bar{e}^2 := \limsup_{k \rightarrow \infty} \max_i \mathbb{E}\|\tilde{\Theta}_i(k)\|_F^2$. Then the following sufficient-condition (practical-stability) statement holds:*

- (a) (Estimation.) *Under the hypotheses of Lemma 2 (persistent excitation, convex projection, SPR noise model, slow drift, mean-square-bounded information vector—the latter established jointly with part (c) below),*

$$\bar{e}^2 \leq \kappa_1 \sigma_e^2 + \kappa_2 \mu_\theta^2, \quad (24)$$

where κ_1, κ_2 depend only on $(\alpha, \beta, L, \lambda^l, \underline{\pi}, \bar{\pi}, c_\psi, \zeta_{\text{SPR}})$; in particular $\bar{e} \rightarrow 0$ as $(\sigma_e, \mu_\theta) \rightarrow 0$.

- (b) (Control perturbation.) *There is a constant $L_u < \infty$ (from Proposition 2) such that, on the admissible set,*

$$\|u_i(k) - u_i^*(k)\| \leq L_u \|\tilde{\Theta}_i(k)\|_F (1 + \|\chi_i(k)\|), \quad (25)$$

where $\chi_i(k)$ stacks the bounded measured signals entering $\xi_i(k)$ and $\rho_i(k)$.

- (c) (Closed loop.) *There exists $\bar{e}^* > 0$ such that whenever $\bar{e} \leq \bar{e}^*$ (guaranteed by (24) for small σ_e, μ_θ), all signals $\{u_i, v_i, y_i\}_{i=1}^N$ are mean-square bounded, and the regulated outputs are near-optimal:*

$$\limsup_{k \rightarrow \infty} \mathbb{E}\|P_i y_i(k) - R_i r_i(k)\|^2 \leq J_i^* + \Gamma_i \bar{e}^2, \quad i = 1, \dots, N, \quad (26)$$

for finite $\bar{e}$ -independent constants Γ_i ; thus the adaptive loop recovers the minimum-variance performance of Theorem 3 as $\sigma_e, \mu_\theta \rightarrow 0$.

Proof. (a) *Estimation bound.* This is Lemma 2, proved in Appendix B; its mean-square-bounded regressor hypothesis is discharged by part (c), the two statements holding jointly through the contraction argument given there.

(b) *Control perturbation.* The applied control is $u_i(k) = \hat{h}_i^{-1}(\hat{b}_i(k))$ and the ideal one is $u_i^*(k) = h_i^{-1}(b_i(k))$, where $\hat{h}_i, \hat{b}_i$ use $\hat{\Theta}_i(k)$ and h_i, b_i use $\Theta_i(k)$. By the triangle inequality, $\|u_i - u_i^*\| \leq \|\hat{h}_i^{-1}(\hat{b}_i) - h_i^{-1}(\hat{b}_i)\| + \|h_i^{-1}(\hat{b}_i) - h_i^{-1}(b_i)\|$. The first term is bounded by $L_\rho \|\hat{\rho}_i - \rho_i\|_F \leq L_\rho \|\tilde{\Theta}_i\|_F$ via the ρ -Lipschitz bound of Proposition 2; the second by $L_b \|\hat{b}_i - b_i\|$. Now $b_i(k) = R_i r_i(k + d_i + 1) - \xi_i(k)$ and ξ_i is affine in the composite estimates with coefficients formed from the bounded measured signals stacked in $\chi_i(k)$ (Theorem 1); hence $\|\hat{b}_i - b_i\| \leq c_\xi \|\tilde{\Theta}_i\|_F (1 + \|\chi_i\|)$. Combining gives (25) with $L_u = L_\rho + L_b c_\xi$.

(c) *Closed loop.* Write $v_i(k) = f_i(u_i(k)) = f_i(u_i^*(k)) + \delta_i(k)$. As f_i is C^1 on the admissible operating set, it is locally Lipschitz with constant L_f , so by (25) $\|\delta_i(k)\| \leq L_f L_u \|\tilde{\Theta}_i(k)\|_F (1 + \|\chi_i(k)\|)$. The ideal internal dynamics (22) are BIBO stable (Theorem 3(b)): with $\mathbf{v} = \mathbf{v}^* + \delta$, $\mathbf{v}^* = \mathcal{B}^{-1}\mathbf{c}$ is MS-bounded and the perturbed solution satisfies $\mathbf{v} = \mathcal{B}^{-1}(\mathbf{c} + \mathcal{D}\delta)$, so for the finite operator gain $\gamma_B := \|\mathcal{B}^{-1}\mathcal{D}\|$,

$$\left(\sum_i \mathbb{E} \|v_i(k)\|^2\right)^{1/2} \leq \left(\sum_i \mathbb{E} \|v_i^*(k)\|^2\right)^{1/2} + \gamma_B \left(\sum_i \mathbb{E} \|\delta_i(k)\|^2\right)^{1/2}.$$

Since χ_i is built from $\{u_i, v_i, y_i\}$ over a finite window, there is a constant c_χ with $\mathbb{E} \|\chi_i(k)\|^2 \leq c_\chi (1 + \max_i \sup_{s \leq k} \mathbb{E} \|v_i(s)\|^2)$; substituting the bound on δ_i and using $\mathbb{E} \|\tilde{\Theta}_i\|_F^2 \leq \bar{e}^2 + o(1)$ gives, for $V(k) := \max_i \sup_{s \leq k} \mathbb{E} \|v_i(s)\|^2$, a recursive inequality of the form $V(k) \leq c_0 + \gamma_B^2 L_f^2 L_u^2 c_\chi \bar{e}^2 (1 + V(k))$. Choosing $\bar{e}^*$ so that $\gamma_B^2 L_f^2 L_u^2 c_\chi (\bar{e}^*)^2 < 1$, the map is a contraction and $V(k)$ is uniformly bounded; hence, $\mathbf{v}$, and (by the growth bound (5)) u_i , are MS-bounded, and so are the outputs via part (a) of Theorem 3. Finally, the realized filtered error is $P_i y_i(k) - R_i r_i(k) = F_i e_i(k) + B_{i,1} \delta_i(k - d_i - 1)$, with $\delta_i(k - d_i - 1) = f_i(u_i(k - d_i - 1)) - f_i(u_i^*(k - d_i - 1))$. The two terms are uncorrelated— $F_i e_i(k)$ is a moving average of the future innovations $e_i(k - d_i), \dots, e_i(k)$ (zero mean, independent of $\mathcal{F}_{k-d_i-1}$) while $B_{i,1} \delta_i(k - d_i - 1)$ is $\mathcal{F}_{k-d_i-1}$ -measurable—so the cross term vanishes and $\mathbb{E} \|P_i y_i(k) - R_i r_i(k)\|^2 = J_i^* + \mathbb{E} \|B_{i,1} \delta_i(k - d_i - 1)\|^2$. Using $\mathbb{E} \|\delta_i\|^2 \leq L_f^2 L_u^2 (1 + c_\chi \sup_k V) \bar{e}^2$ gives (26) with the $\bar{e}$ -independent constant $\Gamma_i = \|B_{i,1}\|^2 L_f^2 L_u^2 (1 + c_\chi \sup_k V)$. Letting $\sigma_e, \mu_\theta \rightarrow 0$ drives $\bar{e} \rightarrow 0$ by (24), so the right-hand side tends to J_i^* . $\square$

Remark 10 (Interpretation and excitation under feedback). Theorem 4 is a practical (total-) stability result: it shows the adaptive loop inherits the optimality of the ideal loop up to a residual set by the estimation error (24), which itself shrinks with the noise level and the parameter drift. Two points deserve emphasis. First, persistent excitation under closed-loop operation is not automatic—minimum-variance control tends to cancel excitation. We are explicit about what this means for Assumption 9: a brief initial probing phase secures excitation only over the initial windows, not over every moving window as the assumption requires. In the revised implementation, the GMV input therefore carries a small persistent dither of standard deviation ς (Proposition 4), and—rather than leaving the assumption unchecked—Section 4 plots the minimum eigenvalue of the windowed information matrix $\sum_{s=k-L+1}^k \bar{\psi}_i(s) \bar{\psi}_i^T(s)$ along the closed-loop trajectory: with the dither, it stays bounded away from zero (with a quantified dip while all set-points sit at the origin, where the odd cubic directions are intrinsically hard to excite), and without the dither it collapses to numerical singularity, confirming that probing alone would not sustain Assumption 9. Covariance wind-up during the weakly excited phases is prevented by the explicit covariance ceiling of Remark 4, and the projection of Assumption 9 keeps $\hat{h}_i^{-1}$ uniformly Lipschitz. Second, the contraction threshold $\bar{e}^*$ is

governed by the BIBO gain γ_B , i.e., by the same small-gain margin (23): weaker interconnections enlarge the tolerable estimation error.

The basis representation is exact in the idealized model. The final result of this section quantifies the price of *under-modeling*, i.e., of a residual mismatch between the true nonlinearity and its basis representation, and shows that the optimality loss is benign.

Proposition 3 (Robustness to basis under-modeling). *Suppose the true static map is $f_i(u) = \Xi_i \Phi_i(u) + \Delta_i(u)$ with a bounded residual $\sup_{u \in \mathcal{U}_i} \|\Delta_i(u)\| \leq \delta_i$, while the controller is designed from the basis part $\Xi_i \Phi_i$ as in Theorem 2. Then, under the conditions of Theorem 3, the realized filtered tracking error is*

$$P_i(q^{-1})y_i(k) - R_i(q^{-1})r_i(k) = F_i(q^{-1}, k)e_i(k) + B_{i,1}(k)\Delta_i(u_i(k - d_i - 1)), \quad (27)$$

and consequently

$$\limsup_{k \rightarrow \infty} \mathbb{E} \|P_i y_i(k) - R_i r_i(k)\|^2 \leq J_i^* + \|B_{i,1}\|^2 \delta_i^2, \quad (28)$$

so the excess cost over the minimum-variance optimum is $O(\delta_i^2)$ and the closed loop remains mean-square bounded.

Proof. See Appendix B. $\square$

The final result of this section closes the gap between the controller *analyzed* above (exact solution of (19)) and the controller *implemented* in Section 4 (damped Newton iteration with a stopping tolerance, a fallback action when no acceptable solution is found, and a persistent-excitation dither). All three implementation deviations are now part of the closed-loop model and are quantified in the performance bound.

Proposition 4 (Implemented controller: inexact solver, fallback and dither). *Let the conditions of Theorem 4 hold and let the applied input be*

$$u_i(k) = u_i^N(k) + \varsigma \zeta_i(k), \quad (29)$$

where $\zeta_i(k)$ is a bounded i.i.d. zero-mean dither with unit covariance, independent of $\{e_j\}$ and of $\mathcal{F}_{k-1}$, and $u_i^N(k)$ is returned by the damped Newton iteration of Algorithm 1 (warm start $u_i(k-1)$, Armijo backtracking, tolerance τ , iteration cap $n_{\max}$): at success steps the solver residual $r_i(k) := \hat{\rho}_i(k) \Phi_i(u_i^N(k)) - \hat{b}_i(k)$ satisfies $\|r_i(k)\| \leq \tau$, and at fallback steps $u_i^N(k) = u_i(k-1)$. Let $p_F := \limsup_{K \rightarrow \infty} \frac{1}{K} \sum_{k \leq K} \Pr[\text{fallback at } k]$ and $\bar{r}_F^2 := \limsup_{k \rightarrow \infty} \mathbb{E}[\|r_i(k)\|^2 | \text{fallback}]$ (finite whenever the loop signals are mean-square bounded, since $u_i(k-1)$ ranges over the bounded actuator set and $\hat{b}_i$ is built from mean-square-bounded signals). Then, there exists $\omega^* > 0$ such that, whenever $\bar{e}^2 + \tau^2 + \varsigma^2 + p_F \bar{r}_F^2 \leq \omega^*$, all closed-loop signals remain mean-square bounded and

$$\limsup_{k \rightarrow \infty} \mathbb{E} \|P_i y_i(k) - R_i r_i(k)\|^2 \leq J_i^* + \bar{\Gamma}_i(\bar{e}^2 + \tau^2 + \varsigma^2 + p_F \bar{r}_F^2), \quad i = 1, \dots, N, \quad (30)$$

for finite constants $\bar{\Gamma}_i$. In particular the bound of Theorem 4 is recovered as $(\tau, \varsigma, p_F) \rightarrow 0$, and the dither inflates the tracking floor by at most $\bar{\Gamma}_i \varsigma^2$ —the quantified price of maintaining the excitation of Assumption 9 in closed loop (Remark 10).

Proof. See Appendix B. $\square$

Remark 11 (Solver events are part of the model). Proposition 4 makes the safeguard of Remark 7 analytically harmless rather than merely heuristic: a fallback step injects the bounded disturbance

$r_i(k)$ into the regulated error—exactly as the under-modeling residual does in Proposition 3—and its effect is weighted by its empirical frequency p_F . In regime (II) of Assumption 6, right-hand sides $\hat{b}_i(k) \notin h_i(\mathcal{U}_i)$ necessarily terminate in a fallback and are therefore covered by the same term. Section 4 reports the measured τ , iteration counts, residuals, and p_F (zero failures in all converged runs), so the implemented and analyzed controllers coincide within the reported tolerances.

Remark 12 (Decentralization and computation). At each step, S_i executes: (i) the RELS update (10); (ii) the Diophantine solution (12) from $\hat{A}_i, \hat{C}_i$ (a triangular recursion, Lemma 3); (iii) the predictor recursion (15) to form $\xi_i(k)$; and (iv) the control solve (19). All quantities are local to S_i except the delayed neighbor signals already present in (9), which are assumed available through the physical interconnections; no central coordinator is required. The per-step cost is dominated by the $O(n_{\theta_i}^2)$ covariance update.

The complete decentralized self-tuning procedure is summarized as Algorithm 1.

Algorithm 1: Decentralized adaptive GMV control of interconnected MIMO Hammerstein subsystems.

For each subsystem $S_i, i = 1, \dots, N$, and each control step k :

1. Form $\bar{\psi}_i(k)$ from (9) using past inputs, outputs, delayed neighbor signals, and residuals $\varepsilon_i(k - \ell)$.
 2. *Identification*: update $\varepsilon_i(k), \hat{\Theta}_i(k), \Pi_i(k), \lambda_i(k)$ by the RELS recursion (10), project $\hat{\Theta}_i(k)$ onto the convex admissible set $\mathcal{C}_i$, and rescale $\Pi_i(k)$ if its trace exceeds the ceiling (Remark 4); extract $\hat{A}_i, \hat{\rho}_i = \widehat{B_{i,1}\Xi_i}, \{\widehat{B_{ij,\ell}\Xi_j}\}, \{\widehat{A_{ij,\ell}\Gamma_j}\}, \hat{C}_i$.
 3. *Diophantine*: solve (12) for $\hat{F}_i, \hat{G}_i$ from $\hat{A}_i, \hat{C}_i$ and the weighting p_i (Lemma 3).
 4. *Prediction*: compute $\xi_i(k)$ via the realizable recursion (15) (with $e_i(k) \leftarrow \varepsilon_i(k)$) and set $b_i(k) = R_i r_i(k + d_i + 1) - \xi_i(k)$.
 5. *Control*: solve $\hat{\rho}_i(k) \Phi_i(u_i^N) = b_i(k)$ for u_i^N by damped Newton: warm start $u_i(k - 1)$, Armijo backtracking on $\|\hat{\rho}_i \Phi_i(u) - b_i(k)\|$, stop at tolerance τ or after $n_{\max}$ iterations; *fallback* $u_i^N = u_i(k - 1)$ if the final residual exceeds the acceptance threshold; log iterations, residual, and fallback events. Apply $u_i(k) = u_i^N + \zeta \xi_i(k)$ (persistent dither, Proposition 4).
-

4. Simulation Study

We validate the scheme on a benchmark process composed of $N = 2$ interconnected two-input/two-output Hammerstein subsystems ($m_i = p_i = 2$), which exercises every feature absent from the SISO scheme of [15]: a coupled multivariable nonlinearity, matrix dynamics, colored noise, delayed interconnections, and time-varying parameters. The revised study then extends the validation along the axes requested by the reviewers: nonlinear-aware and isolation baselines (including the SISO decentralized method of [15] applied channel-by-channel and a feedback-linearization-style controller), a scalability experiment on a chain of $N = 4$ subsystems with computational-cost measurements up to $N = 8$, and a physically motivated interconnected coupled-tank network; throughout, the controller is operated in the forgetting-factor regime $\lambda_i^f < 1$ required by Theorem 4, persistent excitation is *measured* rather than assumed, and full solver diagnostics are reported so that the implemented controller is the one analyzed in Proposition 4.

Process.

Each subsystem has intrinsic delay $d_i = 0$ and interaction delays $d_{ij} = t_{ij} = 2$, so Assumption 3 holds ($0 < 2$). The dynamic and noise orders are $n_{A_i} = n_{B_i} = n_{C_i} = 1$ and $n_{B_{ij}} = 1, n_{A_{ij}} = 0$, giving

$$y_i(k) = -A_{i,1}(k)y_i(k-1) + B_{i,1}v_i(k-1) + B_{ij,1}v_j(k-3) + A_{ij,0}y_j(k-2) + e_i(k) + C_{i,1}e_i(k-1),$$

with output coupling $w_j = y_j$ ($\Gamma_j = I$, $\Psi_j = \text{id}$). The combined static nonlinearities use the monotone cubic basis $\Phi_i(u) = [u_1, u_2, u_1^3, u_2^3]^T$ (so f_i is a global diffeomorphism, Assumption 6), with

$$\Xi_1 = \begin{bmatrix} 0.96 & -0.20 & 0.40 & -0.05 \\ -0.30 & 1.10 & -0.04 & 0.45 \end{bmatrix}, \quad \Xi_2 = \begin{bmatrix} 1.05 & 0.18 & -0.05 & 0.42 \\ 0.22 & -0.95 & 0.38 & -0.04 \end{bmatrix},$$

$$B_{1,1} = \begin{bmatrix} 1.00 & 0.20 \\ 0.10 & 1.00 \end{bmatrix}, \quad B_{2,1} = \begin{bmatrix} 0.90 & -0.15 \\ 0.20 & 0.95 \end{bmatrix}, \quad C_{1,1} = \begin{bmatrix} 0.30 & 0 \\ 0.10 & 0.25 \end{bmatrix}, \quad C_{2,1} = \begin{bmatrix} 0.20 & 0.05 \\ 0 & 0.30 \end{bmatrix}.$$

The parameters are slowly time-varying, $A_{1,1}(k) = \begin{bmatrix} -0.50+s_k & 0.10 \\ 0.05 & -0.40-s_k \end{bmatrix}$, $A_{2,1}(k) = \begin{bmatrix} -0.45 & 0.08 \\ 0.10 & -0.55+0.5s_k \end{bmatrix}$ with $s_k = 0.05 \sin(2\pi k/3000)$. The (weak) interconnections are $B_{12,1} = \begin{bmatrix} 0.10 & 0.04 \\ 0.03 & 0.09 \end{bmatrix}$, $B_{21,1} = \begin{bmatrix} 0.08 & 0.05 \\ 0.04 & 0.10 \end{bmatrix}$, $A_{12,0} = 0.08 \begin{bmatrix} 1 & 0.3 \\ 0.2 & 1 \end{bmatrix}$, $A_{21,0} = 0.07 \begin{bmatrix} 1 & 0.2 \\ 0.3 & 1 \end{bmatrix}$, satisfying the small-gain condition (23). The innovations $e_i(k)$ are zero-mean Gaussian with standard deviation $\sigma_e = 0.05$ per channel. Since $d_i = 0$ and $P_i = R_i = I$, Lemma 3 gives $F_i = I$ and $G_i = C_{i,1} - A_{i,1}$, so the irreducible tracking-error standard deviation equals $\sigma_e = 0.05$.

Controller.

Each S_i runs Algorithm 1 with $\Pi_i(0) = 10^4 I$, $\hat{\Theta}_i(0) = 0$, $\lambda_{0i} = 0.99$, $\lambda_i^I = 0.998 < 1$, $\lambda_i(0) = 0.95$. The forgetting-factor regime now matches the hypothesis $\lambda_i^I < 1$ of Lemma 2 and Theorem 4—the original version used $\lambda_i^I = 1$ in the experiment, while the theorem assumed $\lambda_i^I < 1$, an inconsistency pointed out by a reviewer and removed here (a control run in the old regime is reported below for comparison). Consistently with Remark 4, each update is followed by the projection onto the convex Frobenius ball $\|\hat{\Theta}_i\|_F \leq 10$ and by a covariance ceiling that rescales Π_i whenever $\text{tr } \Pi_i > 50$ (without the ceiling, $\lambda_i^I < 1$ produces covariance wind-up and estimator bursting during the weakly excited closed-loop phases, exactly as the theory predicts). A short open-loop excitation phase ($k < 300$, $u_i \sim 1.5\mathcal{N}(0, I)$) initializes the estimates; the GMV law (19) is then applied, the cubic Equation (19) being solved by the damped (Armijo backtracking) Newton iteration of Proposition 4, warm-started at $u_i(k-1)$, with stopping tolerance $\tau = 10^{-8}$, iteration cap $n_{\max} = 50$, residual-acceptance threshold 10^{-2} (above which the fallback $u_i(k) = u_i(k-1)$ is taken and logged), and a persistent dither of standard deviation $\varsigma = 0.02$ added to the applied input (Remark 10). The references are piecewise-constant set-points changing at distinct instants in the two subsystems, so that each acts as a time-varying disturbance on the other.

Results.

Figures 1 and 2 show that both subsystems track their multivariable, time-varying set-points tightly, the residual fluctuations being at the level of the noise floor despite the cross-couplings, the colored noise and the drifting parameters. The decentralized control inputs (Figure 3) remain smooth and bounded— $\max_k \|u_i(k)\|_\infty = 2.62$ during the GMV phase with no recourse to the safeguard saturation—empirically confirming Theorem 3(b). The composite input-gain estimates converge within the excitation phase and then track the parameter variation (Figure 4), the terminal errors being $\|\hat{\rho}_1 - \rho_1\|_F = 0.23$ and $\|\hat{\rho}_2 - \rho_2\|_F = 0.18$ (against $\|\rho_i\|_F \approx 1.5$). These plateaus are larger than under the previous $\lambda_i^I = 1$ regime (0.093 and 0.018): with genuine forgetting the estimator trades steady-state parameter variance for drift-tracking ability, and the small dither bounds the achievable excitation. The tracking performance is unaffected—the GMV law requires accuracy only in the directions the predictor actually uses—which is precisely the certainty-equivalence mechanism quantified in Theorem 4.

Table 1 reports the post-transient tracking root-mean-square error (RMSE), computed over $k \geq 400$. The proposed controller attains a mean RMSE of 0.060, essentially the *operational* floor $(\sigma_e^2 + c_\zeta^2)^{1/2} \approx 0.060$ formed by the irreducible noise floor $\sigma_e = 0.05$ of Theorem 3(a) plus the quantified dither contribution $c_\zeta \approx 0.033$ of Proposition 4 (measured in the noise sweep below; the dither passes through the input gain, so c_ζ exceeds $\zeta = 0.02$). For comparison, we also ran a *linear-MIMO* adaptive controller—identical in every respect except that it ignores the static nonlinearity (basis $\Phi_i(u) = u$)—on the *same* plant. Lacking a model of the amplitude-dependent gain, it incurs a mean RMSE of 0.123 for this realization (0.137 ± 0.011 over the Monte Carlo study below), 2.1–2.3 times the proposed controller’s error, with the persistent bias and sluggish set-point response visible in Figure 5. This quantifies the value of the coupled-Hammerstein treatment imported from [22]. (We note for transparency that the safeguards of the revised regime—the covariance ceiling and the dither—also *help* the linear surrogate, whose mean RMSE improves from 0.31 under the old regime to 0.137 here; the advantage of the nonlinear treatment is therefore reported against the *strongest* version of the linear baseline.)

Table 1. Post-transient tracking RMSE ($k \geq 400$). The irreducible floor is $\sigma_e = 0.05$; the operational floor including the measured dither contribution is $(\sigma_e^2 + c_\zeta^2)^{1/2} \approx 0.060$.

Controller	$S_1:y_1$	$S_1:y_2$	$S_2:y_1$	$S_2:y_2$	Mean
Proposed (coupled Hammerstein)	0.063	0.058	0.059	0.058	0.060
Linear-MIMO (nonlinearity ignored)	0.122	0.098	0.126	0.144	0.123

A residual whiteness check supports the optimality claim: the lag-one autocorrelations of the residual channels are 0.079, 0.020, 0.003 and 0.029—all small—and the residual standard deviation matches σ_e . The largest value again occurs on the channel whose $A_{i,1}$ entry drifts most (S_1 , output 1); under the previous $\lambda_i^f = 1$ regime this signature was much stronger (≈ 0.34), confirming that it reflects the estimator’s drift-tracking lag rather than a modeling error, and that the theorem-consistent forgetting regime in fact *improves* the residual whiteness.

Solver diagnostics (the implemented controller is the analyzed one).

As required by Proposition 4, we report the complete statistics of the control-law solver over the seed-7 run (all post-probing steps, both subsystems): mean 2.88 Newton iterations per step, maximum 7 (cap 50 never reached); mean final residual $\|\hat{\rho}_i \Phi_i(u_i^N) - b_i\| = 4.3 \times 10^{-10}$, maximum 1.0×10^{-8} (the tolerance τ); *zero* fallback events, i.e., $p_F = 0$ in (30), so the fallback term contributes nothing, and the solver contribution to the error budget is bounded by $\tau^2 = 10^{-16}$, twelve orders of magnitude below the noise floor. The mean per-step computational cost per subsystem is 0.105 ms (0.023 ms RELS update, 0.011 ms predictor, 0.070 ms Newton solve) on a desktop CPU, confirming the $O(n_{\theta_i}^2)$ -dominated cost estimate; the same statistics for the other experiments are reported with their respective paragraphs (no run in this paper ever produced a fallback).

Measured persistent excitation.

Figure 6 plots the minimum eigenvalue of the windowed information matrix $\sum_{s=k-L+1}^k \bar{\psi}_i(s) \bar{\psi}_i^T(s)$ ($L = 200$, $n_{\theta_i} = 14$) along the closed-loop trajectory, with and without the persistent dither. With the dither, the minimum eigenvalue is bounded away from zero over the whole run ($\geq 1.3 \times 10^{-6}$ for $k \geq 400$, rising to $\approx 1.4 \times 10^{-4}$ once all set-points have left the origin); the early dip is structural—at the origin the odd cubic directions u^3 are excited only at order ζ^3 —and is precisely the weak-excitation phase against which the covariance ceiling guards. *Without* the dither, the minimum eigenvalue collapses to numerical singularity ($\sim 10^{-12}$) as soon as the probing phase ends: the brief initial excita-

tion does *not* sustain Assumption 9, exactly as a reviewer suspected, and the convergence claim would indeed be unsupported in that regime. The dither-free loop still tracks (mean RMSE 0.054)—boundedness does not require PE—but its estimator operates without a convergence guarantee; all results in this paper therefore use the dithered, measurably excited regime.

The experiment thus confirms the three theoretical pillars of Section 3: the predictor and GMV law drive the regulated outputs to the (dither-inflated) minimum-variance floor (Theorems 1 and 2, Proposition 4); the interconnections are compensated, and all signals remain bounded (Theorem 3); and the decentralized RELS estimator identifies the composite parameters and tracks their variation (Proposition 1, Lemma 2). A control run in the old regime ($\lambda_i^I = 1$, no dither) gives a mean RMSE of 0.058: the conclusions are insensitive to the regime change, but only the $\lambda_i^I < 1$ regime validates Theorem 4 as stated.

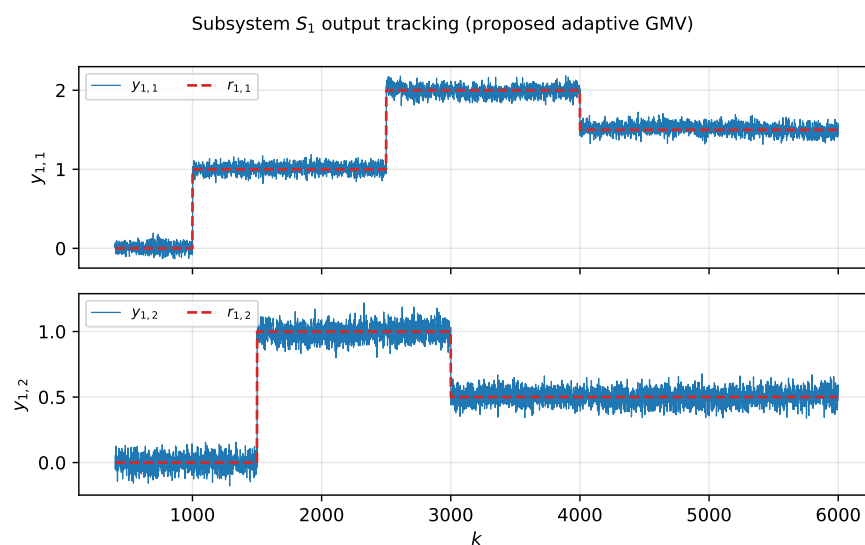


Figure 1. Subsystem S_1 : outputs $y_{1,1}, y_{1,2}$ (solid) tracking the piecewise-constant references $r_{1,1}, r_{1,2}$ (dashed) under the proposed decentralized adaptive GMV controller.

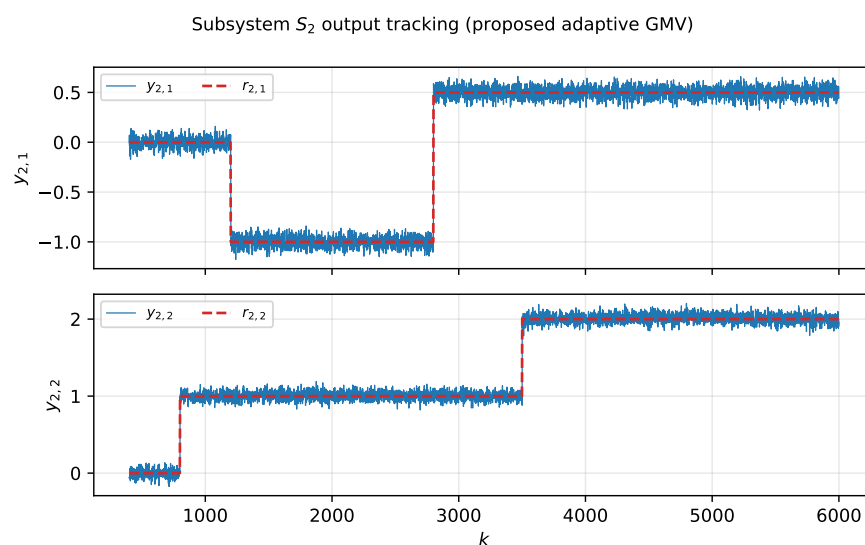


Figure 2. Subsystem S_2 : outputs $y_{2,1}, y_{2,2}$ (solid) tracking the references $r_{2,1}, r_{2,2}$ (dashed). The set-point changes of S_2 occur at instants different from those of S_1 and act as interconnection disturbances on S_1 .

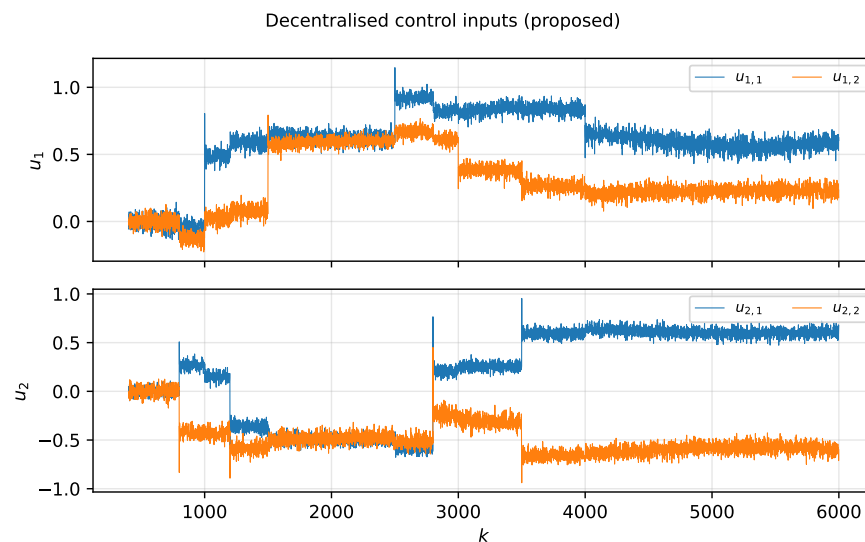


Figure 3. Decentralized control inputs $u_i(k)$. The signals are bounded and smooth; the safeguard saturation is never active.

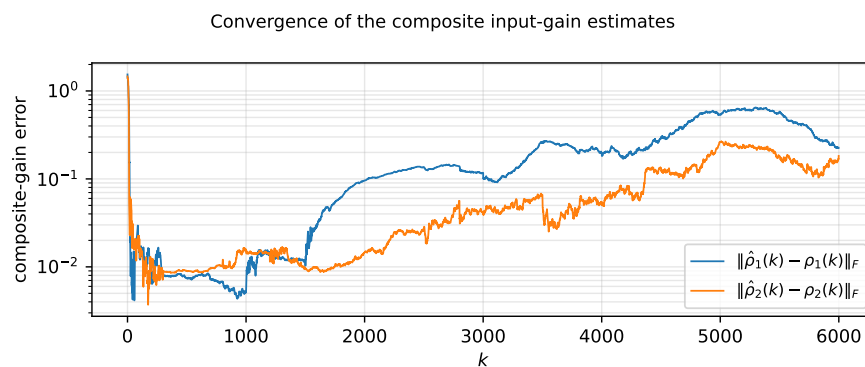


Figure 4. Convergence of the composite input-gain estimates $\|\hat{\rho}_i(k) - \rho_i(k)\|_F$ (logarithmic scale). The estimates converge during the excitation phase and subsequently track the slow parameter variation.

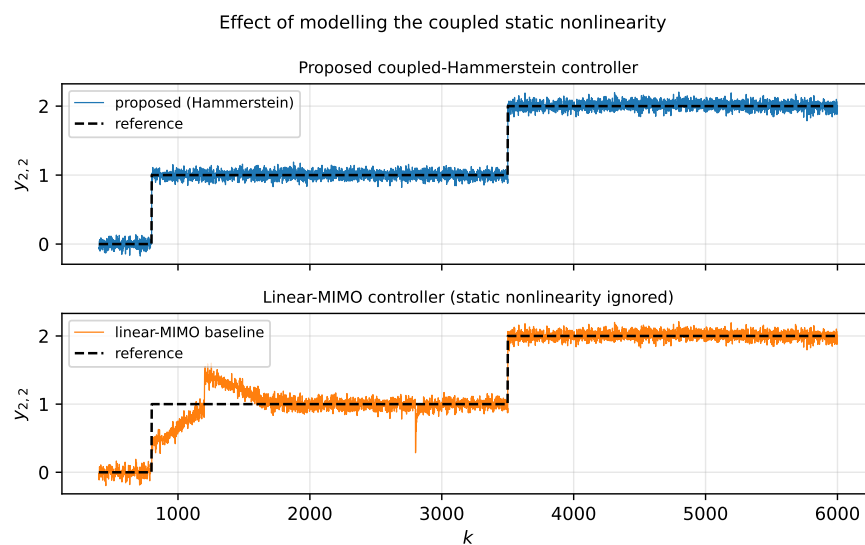


Figure 5. Effect of modeling the coupled static nonlinearity, illustrated on output $y_{2,2}$: the proposed coupled-Hammerstein controller (**top**) tracks the reference, whereas a linear-MIMO controller (**bottom**) that ignores the nonlinearity exhibits large bias and sluggish response.

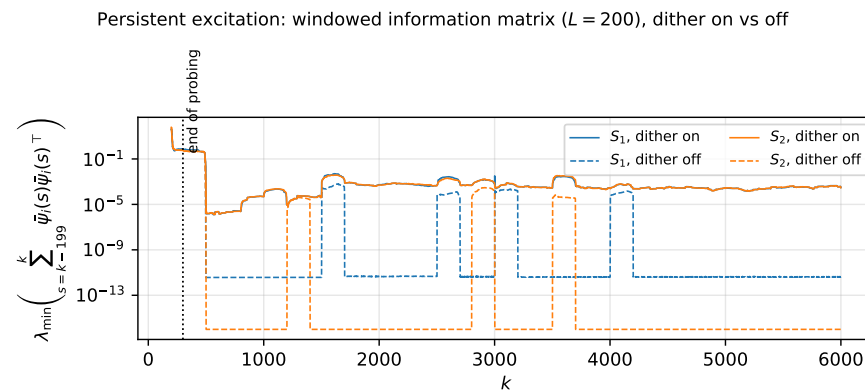


Figure 6. Measured persistent excitation: minimum eigenvalue of the windowed information matrix (window $L = 200$) along the closed-loop trajectory, per subsystem, with the persistent dither (solid) and without it (dashed; values at numerical-singularity level $\sim 10^{-12}$). The vertical line marks the end of the open-loop probing phase.

We now broaden the study along the directions of greatest interest.

Diffeomorphism check (Assumption 6), regime (I).

Over the operating region $u \in [-4, 4]^2$ the Jacobian of the composite static map, $\partial h_i / \partial u = B_{i,1} \Xi_i \partial \Phi_i / \partial u$, has determinant that never changes sign— $\det(\partial h_1 / \partial u) \in [0.98, 4.4 \times 10^2]$, $\det(\partial h_2 / \partial u) \in [-3.3 \times 10^2, -0.92]$ —and, what is actually required by Assumption 6, its smallest singular value is uniformly bounded below, $\sigma_{\min}(\partial h_1 / \partial u) \geq 0.87$ and $\sigma_{\min}(\partial h_2 / \partial u) \geq 0.10$, so (4) holds on $\mathcal{U}_i = [-4, 4]^2$. Because the example is meant to operate under regime (I) of Assumption 6, we also verified the bound on expanding domains: $\sigma_{\min} \geq 0.868$ (subsystem 1) and ≥ 0.059 (subsystem 2) on $[-10, 10]^2$, and ≥ 0.868 and ≥ 0.051 on a logarithmically spaced grid extending to $[-100, 100]^2$, the asymptotic column directions of the Jacobian being linearly independent (the cubic-coefficient blocks of $B_{i,1} \Xi_i$ are nonsingular), so the uniform bound persists as $\|u\| \rightarrow \infty$. Hence, each h_i is a C^1 -diffeomorphism of $\mathbb{R}^2$ by the Hadamard–Lévy theorem and Proposition 2(I) applies—the control Equation (19) has a unique real solution for *every* right-hand side, consistent with the zero fallback count of the solver diagnostics.

Monte Carlo robustness.

Repeating the experiment over 30 independent noise/excitation realizations gives a mean RMSE of 0.0601 ± 0.0003 for the proposed controller and 0.1371 ± 0.0112 for the linear-MIMO surrogate (mean $\pm$ standard deviation), a ratio of 2.28. The tight proposed controller distribution (Figure 7) shows the noise-floor performance is not an artifact of a single realization; note also the 40-fold smaller seed-to-seed spread of the proposed controller, a robustness gap that the mean ratio alone understates.

Ablations.

Disabling the interconnection compensation in the predictor (setting $Y_i^+ \equiv 0$ in (15)) inflates the mean RMSE from 0.060 to 0.193 ($3.2\times$), directly quantifying the benefit of the exact coupling cancellation of Theorem 1(i); disabling the colored-noise compensation (forcing $\hat{C}_i = I$) inflates it to 0.061, a smaller but consistent degradation reflecting the moderate noise coloring. Both confirm that the respective modeling ingredients earn their place.

Nonzero intrinsic delay ($d_i = 1$).

To exercise the nontrivial predictor ($\deg F_i = 1$, so $F_i = I + (C_{i,1} - A_{i,1})q^{-1}$), we repeat the design with $d_i = 1$ and $d_{ij} = t_{ij} = 3$. The proposed controller again tracks tightly (Figure 8), with mean RMSE 0.075; the increase over the $d_i = 0$ floor is the expected

consequence of the larger irreducible variance $\text{tr}(\Sigma_i + F_{i,1}\Sigma_i F_{i,1}^\top)^{1/2} > \sigma_e$ of the two-step predictor (Theorem 1).

Nonlinear-aware and isolation baselines.

A reviewer rightly observed that the linear-MIMO surrogate alone neither benchmarks the scheme against a nonlinear-aware competitor nor isolates which of the two mechanisms—the nonlinear inversion or the adaptation—earns the performance. We therefore compared five controllers on the *identical* plant (same per-seed noise realizations, 30 seeds; Figure 9 and Table 2): (a) the proposed adaptive coupled-Hammerstein GMV (0.0601 ± 0.0003); (b) the linear-MIMO adaptive GMV (0.1371 ± 0.0112), isolating the value of the nonlinear inversion ($2.3\times$); (c) the SISO decentralized Hammerstein method of [15] applied channel-by-channel—each scalar channel runs its own scalar RELS with the univariate basis $[u, u^3]$ and a scalar cubic GMV law solved by radicals, with cross-channel, cross-subsystem, and interconnection terms unmodeled. On this strongly cross-coupled plant, the channel-wise design fails structurally: it settles into a *bounded* large-amplitude saturation oscillation (inputs pinned at the actuator limit for most steps; mean RMSE 174 ± 37). We verified this is a genuine structural-misspecification failure and not an implementation artifact, by running the identical channel-wise code on a temporarily decoupled diagonal variant of the plant, where it reaches the noise floor (0.060). This quantifies precisely the gap between [15] and the present multivariable extension; (d) a non-adaptive *oracle* GMV using the true parameters (0.0537 ± 0.0002), isolating the cost of adaptation: the proposed adaptive scheme operates only $\approx 12\%$ above the oracle floor, most of that gap being the deliberate dither; (e) a *feedback-linearization-style* controller that inverts the *true* nonlinearity f_i exactly (by damped Newton on the known $\Xi_i \Phi_i$) and adapts only the linear dynamics (0.0603 ± 0.0003): statistically indistinguishable from the proposed scheme, i.e., *exact prior knowledge of the nonlinearity buys nothing over the proposed composite-parameter adaptive inversion* on this plant. Per-step controller costs are comparable for (a), (c), (d), (e) (0.21–0.26 ms) and about half for (b). No controller produced a Newton fallback. A Hammerstein-aware model-predictive controller remains outside the scope of this comparison: it optimizes a receding-horizon criterion rather than the GMV criterion (6), so a like-for-like comparison would conflate criterion and mechanism; the feedback-linearization baseline (e) plays the intended role of a nonlinear-aware competitor under the same criterion.

Table 2. Extended validation (mean tracking RMSE; irreducible floor $\sigma_e = 0.05$, operational floor ≈ 0.060 incl. dither, for $d_i = 0$).

Study	Proposed	Reference/Note
Monte Carlo, 30 seeds ($d_i = 0$)	0.0601 ± 0.0003	linear-MIMO 0.137 ± 0.011 (ratio 2.28)
Ablation: no interconnection comp.	0.193	vs. 0.060 full ($3.2\times$)
Ablation: no colored-noise comp.	0.061	vs. 0.060 full
Nonzero delay $d_i = 1$	0.075	two-step floor $> \sigma_e$
Old regime ($\lambda_i^f = 1$, no dither)	0.058	insensitive to regime change
SISO channel-wise [15]	—	174 ± 37 (structural failure)
Oracle (true parameters, no adaptation)	—	0.0537 ± 0.0002 (floor without dither)
Feedback linearization + adaptive linear	—	0.0603 ± 0.0003 ($\approx$ proposed)
Chain of $N = 4$ subsystems, 10 seeds	0.0603 ± 0.0002	$\gamma_{N=4}^* = 5.05$, onset in $(4.2, 5.05]$
Per-subsystem cost, $N = 2/4/8$	0.168/0.172/0.167 ms	flat in N (decentralized)
Tank network (10 seeds, cm)	0.0389 ± 0.0002	linear 0.0490 ± 0.0079
Coupling sweep (10 seeds)	flat $\gamma \leq 3$, $+18\%$ at $\gamma = 6$	diverges $\gamma \geq 7$; ideal $\gamma^* \approx 8.1$
Noise sweep $\sigma_e = 0.02/0.1/0.2$	0.039/0.106/0.203	tracks $(\sigma_e^2 + c_\zeta^2)^{1/2}$

Scalability: a chain of $N = 4$ subsystems (and cost up to $N = 8$).

To substantiate the large-scale claim beyond $N = 2$, we repeated the design on a chain $S_0 - S_1 - S_2 - S_3$ of four 2×2 MIMO Hammerstein subsystems (nearest-neighbor input and output couplings with $d_{ij} = t_{ij} = 2 > d_i = 0$; per-subsystem parameter drifts with staggered phases; staggered reference steps; interior subsystems carry two neighbors, so their regressor dimension grows from $n_\theta = 14$ to 20). The decentralized controller is unchanged. Results (Figure 10): per-subsystem mean RMSE $\{0.061, 0.059, 0.061, 0.059\}$, overall 0.0602 (seed 7), and 0.0603 ± 0.0002 over 10 Monte Carlo seeds with no divergence and zero Newton fallbacks (mean 2.8 iterations)—the noise-floor performance carries over unchanged from $N = 2$. Composite-gain estimates converge equally fast along the chain (error below one third of $\|\rho_i\|_F$ within ~ 15 steps of excitation for every subsystem); under the parameter drift, the interior subsystems show larger late-run composite-parameter excursions than at $N = 2$ (up to ≈ 0.74 in Frobenius norm against $\|\rho\|_F \approx 1.5$) with *no* effect on tracking—certainty equivalence requires predictor consistency, not parameter convergence, and we report this honestly rather than claiming parameter identification along the chain. The small-gain analysis scales as the theory prescribes: for the $N = 4$ chain, the block matrix $\mathcal{D}^{-1}\mathcal{N}$ of (23) has spectral radius 0.198γ , predicting $\gamma_{N=4}^* = 5.05$ (the richer coupling topology tightens the bound relative to $\gamma_{N=2}^* = 8.10$), and the $N = 4$ coupling sweep (3 seeds per point) tracks the floor up to $\gamma = 4.2$ and diverges in all seeds from $\gamma = 5.05$ onward—the onset is bracketed by the grid in $(4.2, 5.05]$, consistent with the prediction. Computationally, the per-step cost *per subsystem* is flat in the network size—0.168, 0.172 and 0.167 ms for $N = 2, 4, 8$ —because each local controller touches only its own $O(n_{\theta_i}^2)$ covariance and its neighbors' delayed signals; the total cost grows linearly ($\times 2.05$ from $N = 2$ to 4, $\times 1.94$ from 4 to 8). This is the practical content of the decentralization claim.

A physically motivated application: an interconnected coupled-tank network.

The original SISO framework [15] was validated on a three-tank hydraulic process; we close the corresponding gap for the MIMO scheme with an interconnected hydraulic network built from quadruple-tank-process physics [35]. Two modules, each a 2-input/2-output coupled-tank pair in deviation variables around the level $h_0 = 12.5$ cm (tank area 28 cm^2 , outlet 0.071 cm^2 , $g = 981 \text{ cm/s}^2$, time constant $\tau \approx 63$ s, sampling $T_s = 5$ s, discrete pole 0.924, 5% sinusoidal outlet-coefficient drift), are actuated by pumps whose three-way valves split the flow 0.7/0.3 between the two tanks and whose pump-plus-valve characteristic is the *progressive* (monotone cubic) map $q = k_1 u + k_3 u^3$ ($k_1 = 1$, $k_3 = 0.1$)—a coupled multivariable Hammerstein nonlinearity arising from the hardware, with $\sigma_{\min}(\partial h_i / \partial u) \geq 0.40$ on the actuator range (Assumption 6 verified). The modules are interconnected by a transfer pipe (level coupling) and a shared feed line (flow coupling), both with a two-sample transport delay, so Assumption 3 holds for the physical reason of Remark 2, and the small-gain value is $0.075 \ll 1$. With colored level-sensor noise ($\sigma_e = 0.02$ cm) and ± 1.5 cm staggered set-point steps, the proposed controller tracks at 0.0389 ± 0.0002 cm over 10 seeds (Figure 11), against 0.0490 ± 0.0079 cm for the linear-MIMO baseline. We report the comparison honestly: on this mildly nonlinear physical range, the mean advantage is moderate ($1.26\times$; $1.56\times$ on the seed shown), and the dominant benefit is *robustness*—the linear baseline's seed-to-seed spread is 50 times larger, its error being concentrated in amplitude-dependent transients after each set-point change. Zero Newton fallbacks occurred in all 22 runs.

Coupling-strength and noise sweeps (seed-averaged).

The small-gain quantity of (23) evaluates to $\sup_{|z|=1} \rho(\mathcal{D}^{-1}\mathcal{N}) = 0.1234\gamma$ at coupling scale γ , predicting an ideal-loop (frozen-parameter) stability boundary $\gamma^* = 1/0.1234 \approx 8.1$. The revised sweep runs 10 seeds per point (Figure 12(left), error bars ± 1 std), which resolves

the question raised by a reviewer about the non-monotone single-seed curve of the original version: for $\gamma \leq 3$, the mean RMSE is constant to within $\pm 0.5\%$ (the old apparent dip from $\gamma = 0$ towards $\gamma \approx 5$ is seed-level scatter, not a physical effect—coupling does *not* improve tracking; the per- γ means over common seeds differ by less than 0.0003), exactly the γ -independence that Theorem 3(a) predicts while the couplings are exactly compensated. Approaching the boundary, a genuine, seed-reproducible degradation appears (+2% at $\gamma = 4$, +7% at $\gamma = 5$, +18% at $\gamma = 6$), and all ten seeds diverge from $\gamma = 7$ onward. The empirical divergence onset in $(6, 7]$ lies *inside* the ideal bound $\gamma^* \approx 8.1$, and the theory says why: γ^* certifies the frozen-parameter loop (Theorem 3), whereas the adaptive loop tolerates an estimation error only up to a threshold $\bar{e}^*$ that *shrinks* as the interconnection gain γ_B grows (Theorem 4(c), Remark 10); near the boundary, the margin is consumed by adaptation and dither, so divergence slightly before γ^* is the predicted behavior of the implemented loop, not a contradiction. The noise sweep, Figure 12(right), shows the mean RMSE tracking the operational floor $(\sigma_e^2 + c_\zeta^2)^{1/2}$ with $c_\zeta = 0.033$ to within 0.3% across $\sigma_e \in \{0.02, 0.05, 0.10, 0.20\}$, confirming the minimum-variance optimality of Theorem 3 together with the dither-cost term of Proposition 4. Table 2 collects these results.

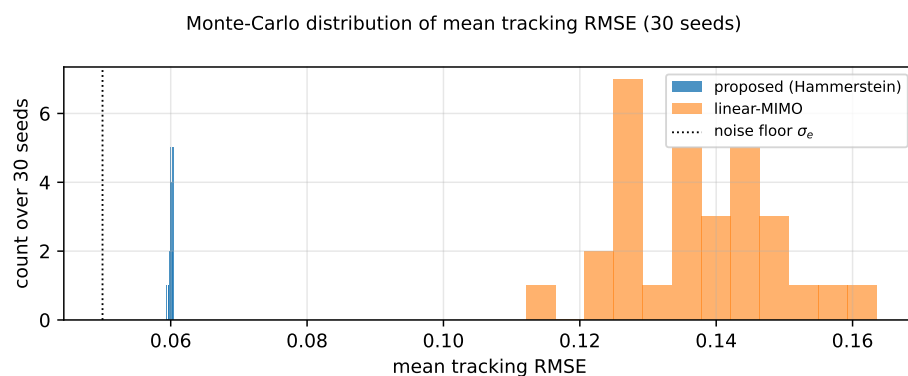


Figure 7. Monte Carlo distribution of the mean tracking RMSE over 30 seeds.

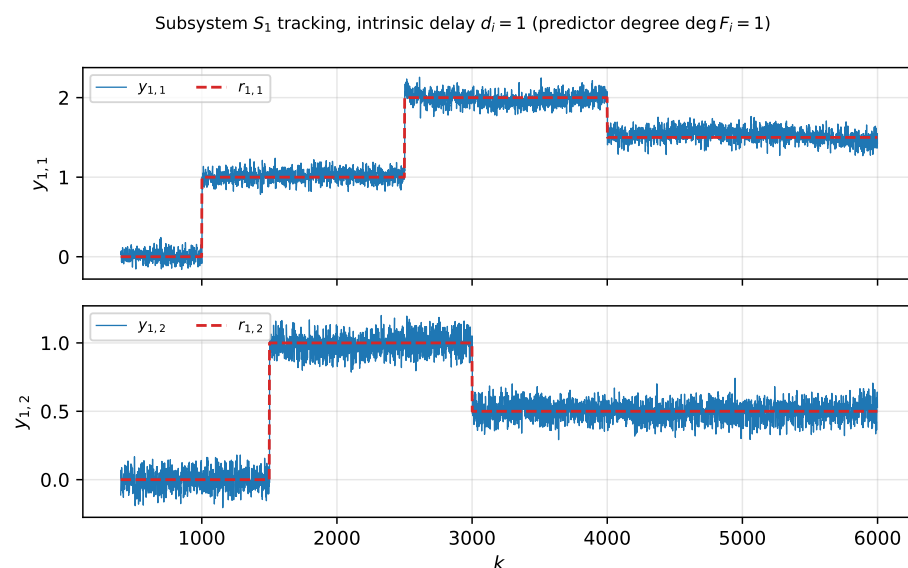


Figure 8. Tracking of S_1 for intrinsic delay $d_i = 1$ ($\deg F_i = 1$).

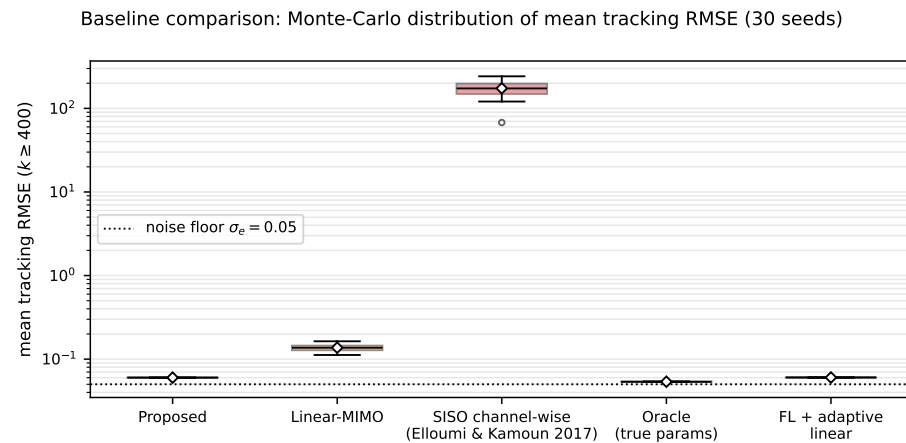


Figure 9. Distributions of the mean tracking RMSE over 30 common seeds for the five controllers of the baseline study (log scale); the SISO channel-wise baseline is the decentralized Hammerstein method of Elloumi and Kamoun [15] applied channel-by-channel. The dotted line is the noise floor σ_e .

Chain of $N = 4$ subsystems: first-output tracking with staggered set-points (proposed adaptive GMV)

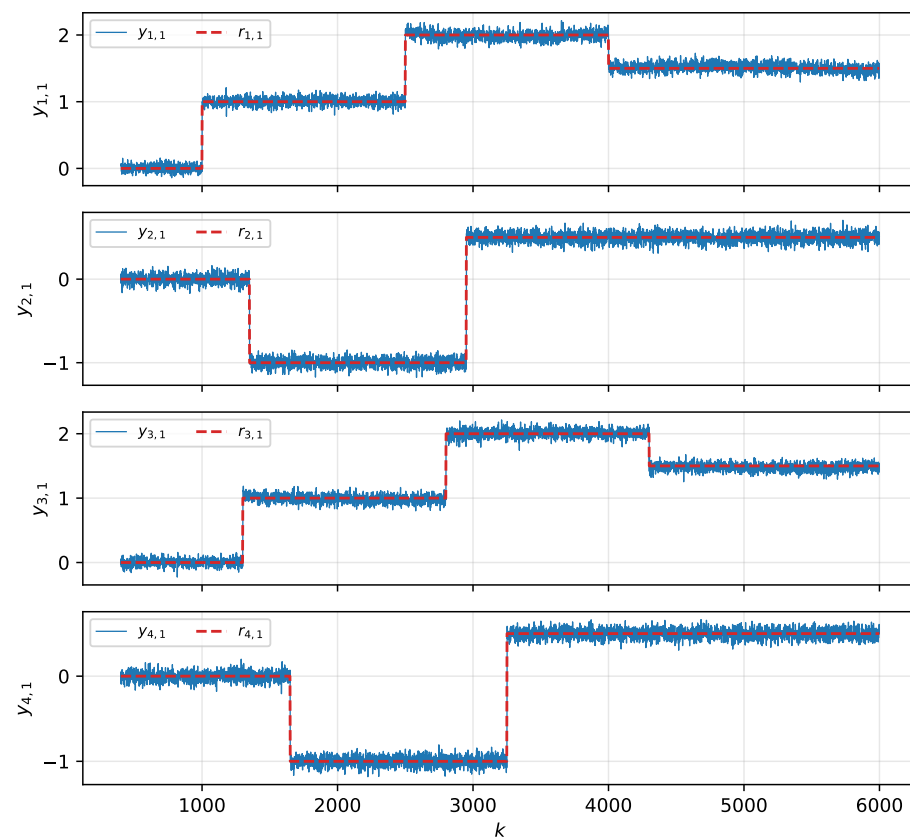


Figure 10. Scalability experiment: chain of $N = 4$ interconnected MIMO Hammerstein subsystems—output $y_{i,1}$ of each subsystem (solid) tracking its staggered reference (dashed) under the unchanged decentralized controller.

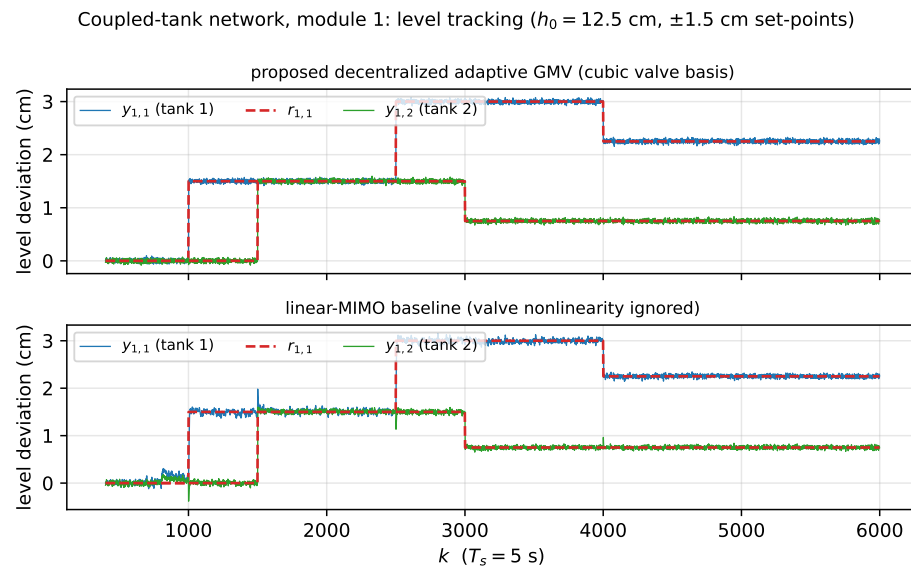


Figure 11. Interconnected coupled-tank network (deviation levels in cm): module-1 tank levels under the proposed controller (**top**) and the linear-MIMO baseline on the same channel (**bottom**). The linear baseline's error concentrates in amplitude-dependent transients after the set-point steps.

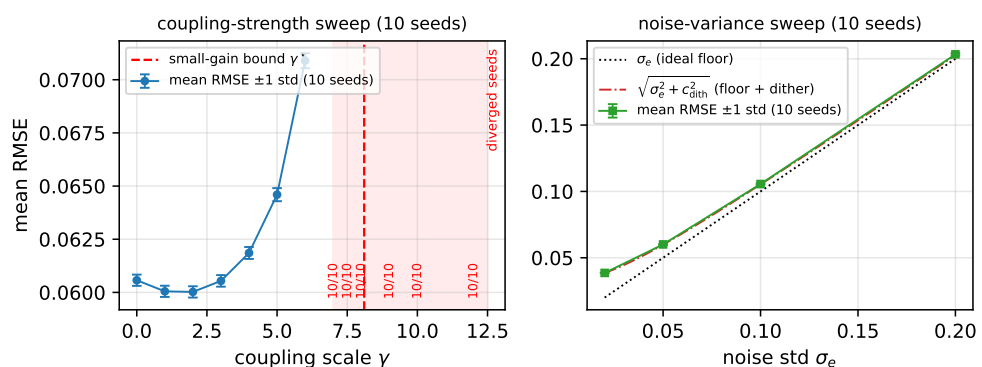


Figure 12. **Left:** seed-averaged coupling-strength sweep (10 seeds per point, error bars ± 1 std)—the RMSE is γ -independent for $\gamma \leq 3$, degrades mildly approaching the boundary, and all seeds diverge from $\gamma = 7$ (shaded), inside the ideal-loop small-gain bound $\gamma^* \approx 8.1$ (red dashed). **Right:** noise-variance sweep (10 seeds)—the mean RMSE tracks the operational floor $(\sigma_e^2 + c_{\text{dith}}^2)^{1/2}$ (dash-dot); the dotted line is the dither-free floor σ_e .

5. Conclusions

We have developed and analyzed a decentralized adaptive generalized-minimum-variance control scheme for large-scale stochastic systems composed of interconnected MIMO Hammerstein subsystems with combined nonlinearities, colored noise, delayed couplings, and unknown time-varying parameters. The contribution is a *control-synthesis* layer that transplants the multivariable, coupled, and colored-noise modeling ingredients of Li et al. [22] into the decentralized interconnected adaptive-control architecture of El-loumi and Kamoun [15], the static maps being represented on a known invertible basis so that the control law stays closed form; the proposed model recovers both prior settings as special cases, while the deep-network identification of [22] remains complementary and outside the present formulation. We proved existence and uniqueness of the governing matrix Diophantine identity, derived the optimal multi-step predictor, and showed that the GMV control law reduces to a multivariable polynomial equation in the current input expressed entirely through the identifiable composite parameters—so that no sep-

aration of the Hammerstein product is needed—with unique solvability and a Lipschitz solution map established through a Hadamard–Lévy global-diffeomorphism argument (globally on $\mathbb{R}^{m_i}$, or on a compact convex domain under an explicit injectivity hypothesis). A layered analysis proved that the structural interaction delays make the couplings measurable and hence exactly compensable, yielding minimum-variance, mean-square-bounded regulated outputs under a Schur weighting and bounded internal signals under a minimum-phase/small-gain condition (the second-moment bounds being carried by explicit linear-growth conditions rather than mere continuity); a self-contained estimation lemma under verifiable assumptions (persistent excitation, a strictly positive-real noise model, convex projection, and slow drift) and a dedicated *adaptive* theorem then gave sufficient conditions for mean-square boundedness and near-optimal tracking of the certainty-equivalence loop, a robustness proposition bounded the effect of basis under-modeling, and an implementation proposition extended the guarantee to the controller actually executed—inexact damped-Newton solver, fallback events and persistent dither included. An extensive simulation study—run in the forgetting-factor regime required by the theory, with measured persistent excitation and complete solver diagnostics, a 30-seed Monte Carlo experiment, a nonzero-delay case exercising the nontrivial predictor, compensation ablations, nonlinear-aware and isolation baselines (channel-wise SISO Hammerstein control fails structurally; a feedback-linearization controller knowing the true nonlinearity offers no advantage; the adaptive loop runs within 12% of the true-parameter oracle), a four-subsystem chain with size-independent per-subsystem computational cost, a physically motivated interconnected coupled-tank network, and seed-averaged coupling/noise sweeps—corroborated the analysis: the controller attained the operational noise floor, reduced the mean tracking RMSE 2.3-fold relative to the strongest linear-MIMO surrogate, and exhibited the divergence boundary predicted by the small-gain condition (with the adaptive loop consuming a quantified part of the ideal margin, as the theory anticipates).

Several extensions are natural. The minimum-phase requirement could be relaxed by adding a control-weighting term to (6), yielding a detuned GMV or a generalized-predictive variant; non-square subsystems ($m_i \neq p_i$) call for a right-inverse or least-squares form of (19); and the practical adaptive-stability statement of Theorem 4 could be strengthened to an almost-sure global convergence result along the martingale lines of [24]. Finally, when an explicit basis is unavailable, the static block could be identified by the deep network of [22] and combined with a numerical inversion of the control equation, trading the closed-form law of Theorem 2 for greater modeling flexibility.

Author Contributions: S.D., conceptualization, methodology, algorithm development, validation, writing—original draft, and supervision; M.E., theoretical analysis, fractional-system modeling, methodology, and writing—review and editing; H.A. (Hend Aljahani), formal analysis, validation, and writing—review and editing; S.A. (Salem Albalawi), simulation design, numerical experiments, data curation, and visualization; S.A. (Sahar Almashaan), investigation, resources, project administration, and writing—review and editing; H.A. (Hatem Alwardi), software implementation, reproducibility verification, and visualization; F.M., theoretical verification, interpretation of results, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number “NBU-FPEJ-2026-1257-01”. The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khaled University for funding this work through Large Research Project under grant number RGP2/202/47.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number “NBU-FPEJ-2026-1257-01”. The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khaled University for funding this work through Large Research Project under grant number RGP2/202/47.

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

Appendix A. Structural Comparison with the Source Frameworks

Table A1 summarizes, dimension by dimension, how the present formulation relates to and extends the SISO interconnected Hammerstein control scheme of Elloumi and Kamoun [15] and the MIMO Hammerstein identification framework of Li et al. [22]. The comparison makes precise the scope of the contribution: a control-synthesis layer that lifts the former to the multivariable interconnected setting using the modeling ingredients of the latter, while keeping the control law in closed form.

Table A1. Structural comparison of the proposed scheme with the two-source frameworks.

Aspect	Elloumi and Kamoun [15]	Li et al. [22]	This Work
Goal	adaptive control	identification only	adaptive control
Plant scope	N interconnected subsystems	single plant	N interconnected subsystems
Subsystem type	SISO ($m_i = p_i = 1$)	MIMO	MIMO ($m_i = p_i$)
Dynamic block	scalar polynomials	matrix ARMA	matrix polynomials (ARMA)
Nonlinearity	univariate polynomial	combined (coupled), DGRU	combined (coupled), known invertible basis
Noise	scalar MA filter $C_i(q^{-1})e_i$	colored ARMA	matrix MA $C_i(q^{-1})e_i$ in a MIMO predictor
Interconnections	input couplings (delayed)	none	input and output couplings (delayed)
Estimator	scalar RELS, forgetting	RELS (ARMA) + AMS-GWO/Adam (DGRU)	matrix RELS, forgetting, projection
Predictor	scalar Diophantine	—	matrix Diophantine (Lemma 3)
Control law	scalar polynomial in u_i	—	multivariable polynomial in u_i via composite p_i (Thm 2)
Solvability	real-root selection	—	global-diffeomorphism + Lipschitz map (Prop. 2)
Stability	root location (low order)	DGRU Lyapunov convergence	layered min-phase/small-gain (Thm 3)
Adaptive guarantee	— (example level)	—	practical MS-boundedness (Thm 4)
Robustness	—	—	basis under-modeling $O(\delta_i^2)$ (Prop. 3)
Validation	3-tank hydraulic process	numerical + photovoltaic	MC, ablations, $d_i = 1$, seed-averaged small-gain/noise sweeps, five-way baseline study (incl. channel-wise [15], oracle, feedback linearization), $N = 4$ chain + cost scaling, coupled-tank network, measured PE, solver diagnostics

Appendix B. Deferred Proofs

For readability, the proofs of the supporting lemmas and propositions are collected here; the proofs of the central Theorems 1–4 appear in the main text.

Proof of Lemma 1. We expand each operator term of (2) explicitly and verify dimensional consistency block by block. Because $A_i(q^{-1}, k)$ is monic (Assumption 2), $A_{i,0} = I_{p_i}$ and

$$A_i(q^{-1}, k) y_i(k) = \sum_{\ell=0}^{n_{A_i}} A_{i,\ell}(k) y_i(k - \ell) = y_i(k) + \sum_{\ell=1}^{n_{A_i}} A_{i,\ell}(k) y_i(k - \ell),$$

with each $A_{i,\ell}(k) \in \mathbb{R}^{p_i \times p_i}$ acting on $y_i(k - \ell) \in \mathbb{R}^{p_i}$. For the controlled input, the shift operator acts as

$$q^{-d_i} B_i(q^{-1}, k) v_i(k) = \sum_{\ell=1}^{n_{B_i}} B_{i,\ell}(k) q^{-(d_i+\ell)} v_i(k) = \sum_{\ell=1}^{n_{B_i}} B_{i,\ell}(k) v_i(k - d_i - \ell),$$

and substituting the static law $v_i(k) = \Xi_i \Phi_i(u_i(k))$ from (1) gives the summands $B_{i,\ell}(k) \Xi_i \Phi_i(u_i(k - d_i - \ell))$, in which $B_{i,\ell}(k) \in \mathbb{R}^{p_i \times m_i}$, $\Xi_i \in \mathbb{R}^{m_i \times L_i}$ and $\Phi_i(\cdot) \in \mathbb{R}^{L_i}$,

so the product $B_{i,\ell}(k)\Xi_i \in \mathbb{R}^{p_i \times L_i}$ multiplies $\Phi_i(u_i(k - d_i - \ell)) \in \mathbb{R}^{L_i}$ to yield a vector of $\mathbb{R}^{p_i}$. Identical manipulations applied to the interaction operators give, for $j \neq i$,

$$q^{-d_{ij}} B_{ij}(q^{-1}, k) v_j(k) = \sum_{\ell=1}^{n_{B_{ij}}} B_{ij,\ell}(k) \Xi_j \Phi_j(u_j(k - d_{ij} - \ell)),$$

$$q^{-t_{ij}} A_{ij}(q^{-1}, k) w_j(k) = \sum_{\ell=0}^{n_{A_{ij}}} A_{ij,\ell}(k) \Gamma_j \Psi_j(y_j(k - t_{ij} - \ell)),$$

with composite blocks $B_{ij,\ell}(k)\Xi_j \in \mathbb{R}^{p_i \times L_j}$ and $A_{ij,\ell}(k)\Gamma_j \in \mathbb{R}^{p_i \times M_j}$; the output-coupling sum starts at $\ell = 0$ because A_{ij} has a (delayed) instantaneous term. Finally, monicity of C_i gives $C_i(q^{-1})e_i(k) = e_i(k) + \sum_{\ell=1}^{n_{C_i}} C_{i,\ell} e_i(k - \ell)$. Collecting all terms and isolating $y_i(k)$ yields

$$y_i(k) = - \sum_{\ell=1}^{n_{A_i}} A_{i,\ell}(k) y_i(k - \ell) + \sum_{\ell=1}^{n_{B_i}} [B_{i,\ell}(k)\Xi_i] \Phi_i(u_i(k - d_i - \ell))$$

$$+ \sum_{j \neq i} \left[\sum_{\ell=1}^{n_{B_{ij}}} [B_{ij,\ell}(k)\Xi_j] \Phi_j(u_j(k - d_{ij} - \ell)) + \sum_{\ell=0}^{n_{A_{ij}}} [A_{ij,\ell}(k)\Gamma_j] \Psi_j(y_j(k - t_{ij} - \ell)) \right]$$

$$+ \sum_{\ell=1}^{n_{C_i}} C_{i,\ell} e_i(k - \ell) + e_i(k).$$

Every bracketed matrix multiplies a *distinct* delayed regressor; stacking the brackets column-wise into $\Theta_i^T(k) \in \mathbb{R}^{p_i \times n_{\theta_i}}$ as in (8), with $n_{\theta_i} = p_i n_{A_i} + L_i n_{B_i} + \sum_{j \neq i} (L_j n_{B_{ij}} + M_j (n_{A_{ij}} + 1)) + p_i n_{C_i}$, and stacking the matching regressors into $\psi_i(k) \in \mathbb{R}^{n_{\theta_i}}$ as in (9), the displayed identity is exactly $y_i(k) = \Theta_i^T(k) \psi_i(k) + e_i(k)$, which is (7). The product $\Theta_i^T(k) \psi_i(k) \in \mathbb{R}^{p_i}$ is well defined since $\Theta_i^T(k) \in \mathbb{R}^{p_i \times n_{\theta_i}}$ and $\psi_i(k) \in \mathbb{R}^{n_{\theta_i}}$. $\square$

Proof of Proposition 1. Recursion (10) is the recursive solution of the exponentially weighted least-squares problem $\hat{\Theta}_i(k) = \arg \min_{\Theta} \sum_{s=1}^k (\prod_{r=s+1}^k \lambda_i(r)) \|y_i(s) - \Theta^T \bar{\psi}_i(s)\|^2$. Because the *same* regressor $\bar{\psi}_i(s)$ multiplies every one of the p_i output equations (the rows of Θ_i^T), the normal equations decouple across the columns of Θ_i and share the single Gram matrix $\Pi_i^{-1}(k) = \lambda_i(k) \Pi_i^{-1}(k-1) + \bar{\psi}_i(k) \bar{\psi}_i^T(k)$; hence, a *common* covariance $\Pi_i(k)$ and a vector gain $L_i(k) = \Pi_i(k) \bar{\psi}_i(k)$ suffice, and the parameter update is the rank-one outer product $L_i(k) \varepsilon_i^T(k)$. Applying the matrix inversion lemma $(A + BC)^{-1} = A^{-1} - A^{-1} B (I + C A^{-1} B)^{-1} C A^{-1}$ to $\Pi_i^{-1}(k) = \lambda_i(k) \Pi_i^{-1}(k-1) + \bar{\psi}_i(k) \bar{\psi}_i^T(k)$, with $A = \lambda_i(k) \Pi_i^{-1}(k-1)$, $B = \bar{\psi}_i(k)$, $C = \bar{\psi}_i^T(k)$, yields the covariance update (10)b and the gain $L_i(k) = \Pi_i(k-1) \bar{\psi}_i(k) / [\lambda_i(k) + \bar{\psi}_i^T(k) \Pi_i(k-1) \bar{\psi}_i(k)]$ of (10)a. Summing the information recursion over a window of length L and invoking the persistent-excitation bounds of Assumption 9 gives, for $\lambda_i(k) \rightarrow \lambda_i^I \in (0, 1]$, $\alpha \sum_{j=0}^{L-1} (\lambda_i^I)^j I \preceq \Pi_i^{-1}(k) - (\lambda_i^I)^L \Pi_i^{-1}(k-L) \preceq \beta \sum_{j=0}^{L-1} (\lambda_i^I)^j I$, whence $\Pi_i^{-1}(k)$ is uniformly bounded above and below; the lower bound prevents wind-up, and the implementation enforces the upper bound directly during weakly excited phases (Remark 4). Convergence of $\hat{\Theta}_i$ to a neighborhood of Θ_i then follows from the standard analysis of recursive (extended) least squares under persistent excitation [24,27,28] and is quantified in Lemma 2. $\square$

Proof of Lemma 2. Fix i and drop the subscript where unambiguous.

Step 1: exact error model. By Lemma 1, $y(k) = \Theta^T(k) \psi(k) + e(k)$, where the ideal regressor ψ contains the unmeasured noises $e(k - \ell)$, $\ell = 1, \dots, n_C$. The implemented

vector $\bar{\psi}$ replaces them by the residuals $\varepsilon(k - \ell)$ and leaves every other entry unchanged, so with $z(k) := e(k) - \varepsilon(k)$,

$$y(k) = \Theta^T(k) \bar{\psi}(k) + e(k) + \sum_{\ell=1}^{n_C} C_\ell z(k - \ell).$$

Subtracting the one-step prediction and using $\varepsilon(k) = y(k) - \hat{\Theta}^T(k-1) \bar{\psi}(k)$ and $b(k) := \tilde{\Theta}^T(k-1) \bar{\psi}(k)$,

$$z(k) + \sum_{\ell=1}^{n_C} C_\ell z(k - \ell) = -b(k), \quad \text{i.e.,} \quad C(q^{-1})z(k) = -b(k) :$$

the residual-substitution error is the output of the *inverse noise filter* driven by the regression error. This is precisely the structure that invalidates off-the-shelf RLS bounds and calls for the passivity argument below.

Step 2: passivity. Let $u_s := -b$ and $w_s := H(q^{-1})u_s = C^{-1}(q^{-1})u_s - \frac{1}{2}u_s = z + \frac{1}{2}b$. By Assumption 10, H is SPR with margin ς_{SPR} , so by the discrete-time Kalman–Yakubovich–Popov (positive-real) lemma applied to a minimal realization of H [24] (Appendix C), there exist a storage function $S(k) = x_s^T(k)P_s x_s(k) \geq 0$ (x_s the filter state) and $c_s \geq 0$ with, for every k ,

$$u_s^T(k) w_s(k) \geq S(k+1) - S(k) + \varsigma_{\text{SPR}} \|u_s(k)\|^2,$$

which, after substituting $u_s = -b$ and $w_s = z + \frac{1}{2}b$, yields the key per-step inequality

$$-b^T(k) z(k) \geq \left(\frac{1}{2} + \varsigma_{\text{SPR}}\right) \|b(k)\|^2 + S(k+1) - S(k). \quad (\text{A1})$$

In words: the cross term between the regression error and the residual-substitution error, which standard RLS analyses cannot control, is dominated by a strictly negative multiple of $\|b\|^2$ up to a telescoping storage increment.

Step 3: Lyapunov recursion. Consider $V(k) = \text{tr}(\tilde{\Theta}^T(k)\Pi^{-1}(k)\tilde{\Theta}(k)) + 2S(k)$. Subtracting $\Theta(k) = \Theta(k-1) + \Delta(k)$, $\|\Delta(k)\|_F \leq \mu_\theta$ (Assumption 5), from the update (10)a, using the information recursion $\Pi^{-1}(k) = \lambda(k)\Pi^{-1}(k-1) + \bar{\psi}(k)\bar{\psi}^T(k)$ and the standard completion of squares [24] (§3.3), and noting that the innovation entering the update is $\varepsilon(k) = e(k) - z(k) - b(k)$, one obtains

$$\begin{aligned} V(k) &\leq \lambda(k)V(k-1) - \frac{\lambda(k)\|b(k)\|^2}{\lambda(k) + \bar{\psi}^T(k)\Pi(k-1)\bar{\psi}(k)} + 2b^T(k)z(k) + 2\left(\frac{1}{2} + \varsigma_{\text{SPR}}\right)\|b(k)\|^2 \\ &\stackrel{(\text{A1})}{\leq} \lambda(k)V(k-1) - \frac{\lambda(k)\|b(k)\|^2}{\lambda(k) + \bar{\psi}^T(k)\Pi(k-1)\bar{\psi}(k)} + 2\text{tr}(\tilde{\Theta}^T(k)\Pi^{-1}(k)\Delta(k)) + \nu(k), \end{aligned}$$

where the SPR surplus has absorbed both the cross term $2b^T z$ and the $\|b\|^2$ -inflation caused by the residual substitution (this is the standard ELS mechanism, here written with forgetting), and $\nu(k)$ collects the terms that are zero-mean given $\mathcal{F}_{k-1}$ together with the squared-innovation contribution, $\mathbb{E}[\nu(k) \mid \mathcal{F}_{k-1}] \leq c \bar{\pi} \sigma_e^2$ by Assumption 4 and $\sup_k \mathbb{E}\|\bar{\psi}(k)\|^2 \leq c_\psi$. The projection step satisfies $\|\mathcal{P}_\mathcal{C}(\tilde{\Theta}) - \Theta(k)\|_F \leq \|\tilde{\Theta} - \Theta(k)\|_F$, since $\Theta(k) \in \mathcal{C}$ and $\mathcal{P}_\mathcal{C}$ is nonexpansive on the convex $\mathcal{C}$ (Assumption 9); hence, it cannot increase V and every inequality above survives the projection. Taking expectations, using $\underline{\pi}I \preceq \Pi^{-1}(k) \preceq \bar{\pi}I$ (Proposition 1, enforced in implementation by the ceiling of Remark 4), the Cauchy–Schwarz estimate $2\text{tr}(\tilde{\Theta}^T \Pi^{-1} \Delta) \leq \epsilon \bar{\pi} \mathbb{E}\|\tilde{\Theta}\|_F^2 + \epsilon^{-1} \bar{\pi} \mu_\theta^2$, and $\lambda(k) \rightarrow \lambda^I < 1$, we get $\mathbb{E}V(k) \leq \eta \mathbb{E}V(k-1) + c_1 \sigma_e^2 + c_2 \mu_\theta^2$ with $\eta := \lambda^I + \epsilon \bar{\pi} / \underline{\pi} \in (0, 1)$ for $\epsilon < (1 - \lambda^I) \underline{\pi} / \bar{\pi}$ —a choice possible precisely because $\lambda^I < 1$. Iterating, using $S \geq 0$ and $\mathbb{E}\|\tilde{\Theta}\|_F^2 \leq \mathbb{E}V / \underline{\pi}$, yields (11) with $\kappa_1 = c_1 / [(1 - \eta) \underline{\pi}]$ and $\kappa_2 = c_2 / [(1 - \eta) \underline{\pi}]$; the constants depend on (α, β, L) through $(\underline{\pi}, \bar{\pi})$, on c_ψ through c_1 , and on ς_{SPR} through the absorption in Step 2. $\square$

Proof of Lemma 3. Write $D_i(q^{-1}) := p_i(q^{-1})C_i(q^{-1}) = \sum_{\ell \geq 0} D_{i,\ell} q^{-\ell}$, which is monic ($D_0 = I$) of degree $n_P + n_{C_i}$, and seek $F_i(q^{-1}) = \sum_{r=0}^{d_i} F_{i,r} q^{-r}$ and $A_i(q^{-1}) = \sum_{s=0}^{n_{A_i}} A_{i,s} q^{-s}$ with $A_{i,0} = I$. The coefficient of q^{-t} in $A_i F_i$ is $\sum_{r+s=t} A_{i,s} F_{i,r}$. Matching the coefficients of q^{-t} on both sides of (12) for $t = 0, 1, \dots, d_i$ (for which the remainder $q^{-(d_i+1)} G_i$ contributes nothing) gives

$$F_0 = D_0 = I, \quad F_t = D_t - \sum_{s=1}^{\min(t, n_{A_i})} A_{i,s} F_{t-s}, \quad 1 \leq t \leq d_i,$$

which determines $F_0, \dots, F_{d_i}$ uniquely and recursively, with $F_0 = I$ so that F_i is monic of degree d_i . Defining G_i by the remainder, $q^{-(d_i+1)} G_i(q^{-1}) := D_i(q^{-1}) - A_i(q^{-1}) F_i(q^{-1})$, i.e., $G_{i,n} = (D_i - A_i F_i)_{n+d_i+1}$ for $n \geq 0$, fixes G_i uniquely; since $\deg(D_i - A_i F_i) = \max\{n_P + n_{C_i}, n_{A_i} + d_i\}$, we obtain $n_{G_i} = \max\{n_{A_i} - 1, n_{C_i} + n_P - d_i - 1\}$.

The construction is a block lower-triangular linear system: writing the unknowns $(F_1, \dots, F_{d_i})$ as a stacked vector, the relations above read $\mathcal{L} \text{vec}(F_{1:d_i}) = \text{vec}(D_{1:d_i}) - \text{vec}(\dots)$ with $\mathcal{L}$ unit lower triangular (identity diagonal blocks coming from $A_{i,0} = I$), hence invertible irrespective of the $A_{i,\ell}$; this is why monicity of A_i —and nothing more—guarantees solvability.

Worked illustration ($d_i = 1$). With $A_i = I + A_{i,1} q^{-1}$, $p_i = 1$ and $C_i = I + C_{i,1} q^{-1}$ one has $D_i = I + C_{i,1} q^{-1}$, so $F_0 = I$, $F_1 = D_1 - A_{i,1} F_0 = C_{i,1} - A_{i,1}$ (monic $F_i = I + (C_{i,1} - A_{i,1}) q^{-1}$), and $q^{-2} G_i = D_i - A_i F_i$ gives $G_i = G_{i,0} = -A_{i,1}(C_{i,1} - A_{i,1})$, matching $n_{G_i} = \max\{0, 0\} = 0$. This is exactly the predictor data used in Section 4.

Uniqueness. If (F'_i, G'_i) is another solution with F'_i monic of degree d_i , then $A_i(F_i - F'_i) = q^{-(d_i+1)}(G_i - G'_i)$. The right-hand side has no terms of order lower than $q^{-(d_i+1)}$, while the left-hand side, because A_i is monic, has the same coefficients as $F_i - F'_i$ in orders $0, \dots, d_i$; as $\deg(F_i - F'_i) \leq d_i$, these must all vanish, so $F_i = F'_i$ and hence $G_i = G'_i$.

On handedness. The factorization places A_i on the left of F_i , matching the left action of A_i on y_i in (2); since matrix polynomials do not commute, the right factorization $C_i F'_i + \dots$ would solve a different problem and would not produce the cancellation used in Theorem 1. The scalar weighting $P_i = p_i I$ is what allows $p_i C_i$ (rather than a noncommuting $P_i C_i$) to appear on the left, keeping the identity well posed. $\square$

Proof of Proposition 2. $\Phi_i \in C^1$ implies $h_i \in C^1$, with Jacobian $\partial h_i / \partial u = \rho_i(k) (\partial \Phi_i / \partial u)(u) = B_{i,1}(k) (\Xi_i \partial \Phi_i / \partial u)$, and by Assumption 6 this Jacobian obeys $\sigma_{\min}(\partial h_i / \partial u) \geq \underline{\sigma}_i > 0$ on $\mathcal{U}_i$.

Regime (I). Here, $\mathcal{U}_i = \mathbb{R}^{m_i}$ and the bound holds globally, so $\sup_{u \in \mathbb{R}^{m_i}} \|(\partial h_i / \partial u)^{-1}(u)\| \leq 1/\underline{\sigma}_i < \infty$. By the Hadamard–Lévy global inverse function theorem [25,26], a C^1 map $\mathbb{R}^n \rightarrow \mathbb{R}^n$ whose Jacobian is everywhere nonsingular with uniformly bounded inverse is a C^1 -diffeomorphism of $\mathbb{R}^n$ onto $\mathbb{R}^n$ (in particular it is automatically proper, so properness need not be hypothesized). Hence, $u_i(k) = h_i^{-1}(b_i(k))$ exists and is unique for every $b_i(k) \in \mathbb{R}^{m_i}$, and $\rho_i(k)$ has full row rank.

Regime (II). On the compact convex $\mathcal{U}_i$, h_i is by hypothesis injective and has a nonsingular Jacobian. An injective continuous map from a compact set onto its (Hausdorff) image is a homeomorphism, and the inverse is C^1 at every image point by the local inverse function theorem; hence, h_i is a C^1 -diffeomorphism of $\mathcal{U}_i$ onto $h_i(\mathcal{U}_i)$, and the equivalence “(19) uniquely solvable in $\mathcal{U}_i \iff b_i(k) \in h_i(\mathcal{U}_i)$ ” is immediate. For the stated sufficient condition for injectivity, suppose $M_i \partial h_i / \partial u + (M_i \partial h_i / \partial u)^\top \succeq 2\delta_i I$ on $\mathcal{U}_i$ and let $u \neq u' \in \mathcal{U}_i$. Because $\mathcal{U}_i$ is convex, the segment $[u, u']$ lies in $\mathcal{U}_i$ and

$$\langle u - u', M_i(h_i(u) - h_i(u')) \rangle = \int_0^1 (u - u')^\top M_i \frac{\partial h_i}{\partial u}(u' + t(u - u')) (u - u') dt \geq \delta_i \|u - u'\|^2 > 0,$$

so $h_i(u) \neq h_i(u')$; moreover the same display gives the quantitative inverse bound $\|h_i^{-1}(b) - h_i^{-1}(b')\| \leq (\|M_i\|/\delta_i)\|b - b'\|$ on $h_i(\mathcal{U}_i)$.

Lipschitz in b . In regime (I), the inverse function theorem gives $\partial h_i^{-1}/\partial b = [(\partial h_i/\partial u)(h_i^{-1}(b))]^{-1}$, so $\|\partial h_i^{-1}/\partial b\| \leq 1/\underline{\sigma}_i =: L_b$ uniformly, and the mean-value inequality along $[b, b'] \subset \mathbb{R}^{m_i}$ (all of $\mathbb{R}^{m_i}$ being convex) gives the first bound in (20). In regime (II) the segment $[b, b']$ need not lie in $h_i(\mathcal{U}_i)$, so the mean-value argument is not available; instead, define the injectivity modulus $c_i := \inf\{\|h_i(u) - h_i(u')\|/\|u - u'\| : u \neq u' \in \mathcal{U}_i\}$. The ratio is continuous and positive off the diagonal of the compact $\mathcal{U}_i \times \mathcal{U}_i$ (by injectivity), and along the diagonal its liminf is at least $\underline{\sigma}_i$ (uniform C^1 regularity), so $c_i > 0$ and $L_b = 1/c_i$ serves; under the monotonicity condition above, one has the explicit value $c_i \geq \delta_i/\|M_i\|$.

Lipschitz in ρ . Differentiating $\rho \Phi_i(u) = b$ in a direction $\delta\rho$ and solving, $\partial u/\partial \rho [\delta\rho] = -(\rho \partial \Phi_i/\partial u)^{-1} \delta\rho \Phi_i(u)$, so $\|\partial u/\partial \rho\| \leq L_b \|\Phi_i(u)\| \leq L_b \Phi_{\max} =: L_\rho$. The bound is obtained by integrating along the segment $[\rho, \rho']$, and this is legitimate precisely because the admissible set $\mathcal{C}_i$ of Assumption 9 is convex, so that $[\rho, \rho'] \subset \mathcal{C}_i$ and the uniform bound $\underline{\sigma}_i$ applies at every point of the segment. (In the original version, this convexity was used implicitly; it is now an explicit standing hypothesis, as one reviewer required.) $\square$

Proof of Proposition 3. The plant produces $v_i(k) = f_i(u_i(k)) = \Xi_i \Phi_i(u_i(k)) + \Delta_i(u_i(k))$. The control law enforces $\hat{\sigma}_i(k + d_i + 1 | k) = 0$ using the basis part, i.e., $B_{i,1}(k) \Xi_i \Phi_i(u_i(k)) + \xi_i(k) - R_i r_i(k + d_i + 1) = 0$. Substituting the true $v_i(k)$ into the predictor decomposition (13) of Theorem 1 gives

$$\sigma_i(k + d_i + 1) = \underbrace{B_{i,1} \Xi_i \Phi_i(u_i(k)) + \xi_i(k) - R_i r_i(k + d_i + 1)}_{=0} + B_{i,1} \Delta_i(u_i(k)) + F_i e_i(k + d_i + 1),$$

since the leading control term acts through $B_{i,1} v_i(k) = B_{i,1} \Xi_i \Phi_i(u_i(k)) + B_{i,1} \Delta_i(u_i(k))$. Shifting the time index by $-(d_i + 1)$ yields (27). Because $F_i e_i(k)$ (future innovations, zero mean, independent of $\mathcal{F}_k$) and $B_{i,1} \Delta_i(u_i(k - d_i - 1))$ ($\mathcal{F}_k$ -measurable, with norm $\leq \|B_{i,1}\| \delta_i$) are uncorrelated, the cross term vanishes and $\mathbb{E}\|P_i y_i(k) - R_i r_i(k)\|^2 = \mathbb{E}\|F_i e_i(k)\|^2 + \mathbb{E}\|B_{i,1} \Delta_i(u_i(k - d_i - 1))\|^2 \leq J_i^* + \|B_{i,1}\|^2 \delta_i^2$, which is (28). Boundedness follows as in Theorem 3, the term $B_{i,1} \Delta_i$ acting as a bounded exogenous disturbance in (22). $\square$

Proof of Proposition 4. By the predictor decomposition (13) and the definitions of Theorem 2, the realized filtered error obeys, for the *applied* input $u_i(k)$ of (29),

$$\sigma_i(k + d_i + 1) = [\rho_i(k) \Phi_i(u_i(k)) - b_i(k)] + F_i e_i(k + d_i + 1),$$

where ρ_i, b_i are the true quantities. Decompose the bracket as

$$\rho_i \Phi_i(u_i) - b_i = \underbrace{\hat{\rho}_i \Phi_i(u_i^N) - \hat{b}_i}_{= r_i(k)} + \underbrace{\hat{\rho}_i [\Phi_i(u_i^N + \zeta \zeta_i) - \Phi_i(u_i^N)]}_{=: \pi_i^d(k)} + \underbrace{(\rho_i - \hat{\rho}_i) \Phi_i(u_i) + (\hat{b}_i - b_i)}_{=: \pi_i^e(k)}.$$

At success steps $\|r_i(k)\| \leq \tau$; at fallback steps $\mathbb{E}\|r_i(k)\|^2 \leq \bar{r}_F^2 + o(1)$ by hypothesis, the event having asymptotic frequency p_F , so $\limsup \mathbb{E}\|r_i(k)\|^2 \leq \tau^2 + p_F \bar{r}_F^2$. For the dither term, Φ_i is C^1 , hence Lipschitz with constant L_Φ on the bounded actuator set, and $\|\hat{\rho}_i\|$ is bounded on the compact $\mathcal{C}_i$; hence, $\mathbb{E}\|\pi_i^d(k)\|^2 \leq c_\zeta^2 \zeta^2$ with $c_\zeta := \sup_{\mathcal{C}_i} \|\hat{\rho}_i\| L_\Phi (\mathbb{E}\|\zeta_i\|^2)^{1/2}$. For the estimation term, exactly as in Theorem 4(b), $\|\pi_i^e(k)\| \leq c \|\hat{\Theta}_i(k)\|_F (1 + \|\chi_i(k)\|)$. *Boundedness.* Write $v_i = f_i(u_i^*) + \delta_i$ with $\delta_i = f_i(u_i) - f_i(u_i^*)$; combining the local Lipschitz property of f_i , the Lipschitz control map (Proposition 2) applied to the perturbed right-hand sides, and the three bounds above, gives $\mathbb{E}\|\delta_i(k)\|^2 \leq c_\delta (\bar{\sigma}^2 + \tau^2 + \zeta^2 + p_F \bar{r}_F^2) (1 + \max_j \sup_{s < k} \mathbb{E}\|v_j(s)\|^2)$. Substituting this perturbation into the BIBO-stable ideal internal

dynamics (22) and repeating verbatim the contraction argument of Theorem 4(c) (with $\bar{e}^2$ replaced by $\omega := \bar{e}^2 + \tau^2 + \varsigma^2 + p_F \bar{r}_F^2$) shows that there is $\omega^* > 0$ such that $\omega \leq \omega^*$ implies uniform mean-square boundedness of all signals; this also discharges the finiteness of $\bar{r}_F$. *Performance.* The innovation term $F_i e_i(k + d_i + 1)$ is a moving average of $e_i(k + 1), \dots, e_i(k + d_i + 1)$, independent of $\mathcal{F}_k$ and of the dither sequence (Assumption 4 and independence of ζ_i), while the bracket is measurable with respect to $\sigma(\mathcal{F}_k \cup \{\zeta_i(k)\})$; the cross term therefore vanishes in expectation and

$$\mathbb{E}\|P_i y_i - R_i r_i\|^2 = J_i^* + \mathbb{E}\|\rho_i \Phi_i(u_i) - b_i\|^2 \leq J_i^* + 3[(\tau^2 + p_F \bar{r}_F^2) + c_\zeta^2 \varsigma^2 + c^2(1 + c_\chi)\bar{e}^2(1 + \sup_k V)] + o(1),$$

which is (30) with $\tilde{\Gamma}_i := 3 \max\{1, c_\zeta^2, c^2(1 + c_\chi)(1 + \sup_k V)\}$. $\square$

References

- Šiljak, D.D. *Decentralized Control of Complex Systems*; Academic Press: Boston, MA, USA, 1991.
- Bakule, L. Decentralized control: An overview. *Annu. Rev. Control* **2008**, *32*, 87–98. <https://doi.org/10.1016/j.arcontrol.2008.03.004>.
- Lunze, J. *Feedback Control of Large-Scale Systems*; Prentice Hall: Hemel Hempstead, UK, 1992.
- Kamoun, S. Design of optimal self-tuning regulators for large-scale stochastic systems. *Int. J. Inf. Syst. Sci.* **2010**, *6*, 201–219.
- Ioannou, P.A.; Sun, J. *Robust Adaptive Control*; Prentice Hall: Upper Saddle River, NJ, USA, 1996.
- Narendra, K.S.; Annaswamy, A.M. *Stable Adaptive Systems*; Prentice Hall: Englewood Cliffs, NY, USA, 1989.
- Yang, S.; Wu, W.; Tong, S. Adaptive fuzzy decentralized output-feedback inverse optimal control for nonstrict-feedback interconnected large-scale systems. *Optim. Control Appl. Methods* **2025**, *46*, 948–959. <https://doi.org/10.1002/oca.3240>.
- Giri, F.; Bai, E.-W. (Eds.) *Block-Oriented Nonlinear System Identification*; Lecture Notes in Control and Information Sciences; Springer: London, UK, 2010; Volume 404. <https://doi.org/10.1007/978-1-84996-513-2>.
- Schoukens, M.; Tiels, K. Identification of block-oriented nonlinear systems starting from linear approximations: A survey. *Automatica* **2017**, *85*, 272–292. <https://doi.org/10.1016/j.automatica.2017.06.044>.
- Bai, E.-W. An optimal two-stage identification algorithm for Hammerstein–Wiener nonlinear systems. *Automatica* **1998**, *34*, 333–338. [https://doi.org/10.1016/S0005-1098\(97\)00198-2](https://doi.org/10.1016/S0005-1098(97)00198-2).
- Ding, F.; Chen, T. Identification of Hammerstein nonlinear ARMAX systems. *Automatica* **2005**, *41*, 1479–1489. <https://doi.org/10.1016/j.automatica.2005.03.026>.
- Santos, L.H.; Braga, M.F.; Ricco, R.A. A new approach for robust control based on parametric Hammerstein models. *Int. J. Dyn. Control* **2025**, *13*, 323. <https://doi.org/10.1007/s40435-025-01826-z>.
- Borisson, U. Self-tuning regulators for a class of multivariable systems. *Automatica* **1979**, *15*, 209–215. [https://doi.org/10.1016/0005-1098\(79\)90071-2](https://doi.org/10.1016/0005-1098(79)90071-2).
- Koivo, H.N. A multivariable self-tuning controller. *Automatica* **1980**, *16*, 351–366. [https://doi.org/10.1016/0005-1098\(80\)90020-5](https://doi.org/10.1016/0005-1098(80)90020-5).
- Elloumi, M.; Kamoun, S. Adaptive control scheme for large-scale interconnected systems described by Hammerstein models. *Asian J. Control* **2017**, *19*, 1075–1088. <https://doi.org/10.1002/asjc.1443>.
- Wang, D.; Mao, L.; Ding, F. Recasted models-based hierarchical extended stochastic gradient method for MIMO nonlinear systems. *IET Control Theory Appl.* **2017**, *11*, 476–485. <https://doi.org/10.1049/iet-cta.2016.0017>.
- Filipovic, V.Z. Outlier robust stochastic approximation algorithm for identification of MIMO Hammerstein models. *Nonlinear Dyn.* **2017**, *90*, 1427–1441. <https://doi.org/10.1007/s11071-017-3736-2>.
- Xia, H.; Ji, Y.; Yang, Y.; Ding, F.; Hayat, T. Improved least-squares identification for multiple-output non-linear stochastic systems. *IET Control Theory Appl.* **2020**, *14*, 964–971. <https://doi.org/10.1049/iet-cta.2019.0915>.
- Li, F.; Sun, X.; Wang, T.; Liu, R. Identification methodology for MIMO Hammerstein nonlinear model with process noise. *Trans. Inst. Meas. Control* **2025**, *47*, 2190–2202. <https://doi.org/10.1177/01423312241276072>.
- Qiu, F.; Wang, L.; Mu, W.; Ji, Y. Hierarchical least squares parameter estimation for the multiple-input nonlinear systems by using the data filtering. *Int. J. Robust. Nonlinear Control* **2025**, *36*, 1727–1741. <https://doi.org/10.1002/rnc.70232>.
- Zhao, Y.; Ji, Y. Weighted multi-innovation parameter estimation for a time-varying Volterra–Hammerstein system with colored noise. *Optim. Control Appl. Methods* **2025**, *46*, 271–291. <https://doi.org/10.1002/oca.3210>.
- Li, F.; Zhu, Y.; Zhao, C.; Liu, R. Parameter estimation of multi-input multi-output Hammerstein nonlinear system with deep GRU networks. *Int. J. Robust. Nonlinear Control* **2026**, *36*, 4792–4810. <https://doi.org/10.1002/rnc.70461>.
- Vörös, J. Recursive identification of Hammerstein systems with polynomial nonlinearities. *J. Electr. Eng.* **2006**, *57*, 42–46.
- Goodwin, G.C.; Sin, K.S. *Adaptive Filtering Prediction and Control*; Prentice Hall: Englewood Cliffs, NJ, USA, 1984.
- Krantz, S.G.; Parks, H.R. *The Implicit Function Theorem: History, Theory, and Applications*; Birkhäuser: New York, NY, USA, 2013. <https://doi.org/10.1007/978-1-4612-0059-8>.

26. Plastock, R. Homeomorphisms between Banach spaces. *Trans. Am. Math. Soc.* **1974**, *200*, 169–183. <https://doi.org/10.1090/S0002-9947-1974-0356122-6>.
27. Ljung, L. *System Identification: Theory for the User*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1999.
28. Söderström, T.; Stoica, P. *System Identification*; Prentice Hall: Hemel Hempstead, UK, 1989.
29. Vidyasagar, M. *Input–Output Analysis of Large-Scale Interconnected Systems*; Lecture Notes in Control and Information Sciences; Springer: Berlin, Germany, 1981; Volume 29. <https://doi.org/10.1007/BFb0044477>.
30. Åström, K.J.; Wittenmark, B. On self-tuning regulators. *Automatica* **1973**, *9*, 185–199. [https://doi.org/10.1016/0005-1098\(73\)90073-3](https://doi.org/10.1016/0005-1098(73)90073-3).
31. Clarke, D.W.; Gawthrop, P.J. Self-tuning control. *Proc. IEE* **1979**, *126*, 633–640. <https://doi.org/10.1049/piee.1979.0145>.
32. Kučera, V. *Discrete Linear Control: The Polynomial Equation Approach*; Wiley: Chichester, UK, 1979.
33. Salgado, M.E.; Goodwin, G.C.; Middleton, R.H. Modified least squares algorithm incorporating exponential resetting and forgetting. *Int. J. Control* **1988**, *47*, 477–491. <https://doi.org/10.1080/00207178808906026>.
34. Åström, K.J.; Wittenmark, B. *Adaptive Control*, 2nd ed.; Addison-Wesley: Reading, MA, USA, 1995.
35. Johansson, K.H. The quadruple-tank process: A multivariable laboratory process with an adjustable zero. *IEEE Trans. Control Syst. Technol.* **2000**, *8*, 456–465. <https://doi.org/10.1109/87.845876>.

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.