Ten Primitives and Three Gates: The Universal Structure of Computational Imaging

Chengshuai Yang[1,*] Xin Yuan[2]

Abstract

Computational imaging systems routinely underperform because the assumed forward model diverges from the true physics. Here we prove two results. First, the **Finite Primitive Basis Theorem**: every linear, shift-variant imaging forward model admits an ε-approximate representation as a typed directed acyclic graph over exactly 10 canonical primitives. Second, the **Triad Decomposition**: every reconstruction failure decomposes into three root causes—information deficiency, carrier noise, and operator mismatch—with mismatch dominant across all validated modalities. Together these results yield a modality-agnostic diagnostic and correction framework. Across seven modalities spanning three carrier families (optical photons, X-ray photons, and nuclear spins), autonomous correction recovers +0.8 to +10.7 dB of mismatch-induced degradation without retraining the solver. Hardware validation on real instruments confirms mismatch dominance. A held-out closure test on 8 additional modalities confirms basis completeness with saturating growth.

Introduction

Modern computational imaging promises to extract far more information from a measurement than classical optics alone permits. Coded aperture spectral cameras compress three-dimensional scenes into a single snapshot[1]; compressive temporal imagers freeze high-speed video into one exposure[2]; single-pixel cameras reconstruct images from far fewer measurements than pixels[3]; accelerated MRI scanners reconstruct diagnostic-quality images from a fraction of the acquired k-space data[4]. Over the past decade, the community has invested enormous effort in improving reconstruction algorithms—progressing from compressed sensing[5,6] and plug-and-play priors[7] to deep unrolling networks[8] and vision transformers[9]—yet these algorithms routinely fail when deployed on real instruments, because the forward model assumed by the solver does not match the physics that generated the data.

These failures persist because computational imaging lacks two theoretical foundations. First, a **representation theory**: no formal result guarantees that a finite set of primitive operators suffices to represent all imaging forward models. Without this, every new modality

*[1]NextGen PlatformAI C Corp, USA. [2]School of Engineering, Westlake University, Hangzhou, China. *Correspondence: integrityyang@gmail.com

requires bespoke engineering and the field cannot reason about completeness. Second, a **diagnostic theory**: no systematic framework decomposes reconstruction failures into root causes. Calibration errors, noise, and information deficiency are conflated, and failures are addressed *ad hoc*. Calibration methods exist for specific instruments[10,11], but they do not generalise; robustness studies perturb individual systems[12], but they lack a unifying formalism.

Here we establish both foundations within Physics World Models (PWM). We prove the **Finite Primitive Basis Theorem** (Theorem 3): every imaging forward model in the operator class $\mathcal{C}_{\mathrm{Tier2}}$—encompassing all clinical, scientific, and industrial modalities at Tier-2 fidelity—admits an ε-approximate representation as a typed DAG over exactly 10 canonical primitives. We establish the **Triad Decomposition**: every reconstruction failure decomposes into three gates—information deficiency (**Gate 1**), carrier noise (**Gate 2**), and operator mismatch (**Gate 3**)—with **Gate 3** dominant across all validated modalities. These two results are complementary: the Finite Primitive Basis provides a universal representation for forward models; the Triad Decomposition provides a universal diagnostic law over that representation. Together they yield a practical framework that diagnoses and corrects imaging failures across modalities without modality-specific tuning.

We validate both contributions across seven modalities spanning three carrier families—optical photons (CASSI, CACTI, SPC, lensless, ptychography[13]), X-ray photons (CT[14]), and nuclear spins (MRI[15])—with hardware validation on real DD-CASSI and CACTI instruments, confirming Triad predictions on physical measurements. The OperatorGraph library additionally registers templates for electron, acoustic, and particle modalities (Extended Data Table 2), pending future validation. A held-out closure test on 8 additional modalities—including quantum ghost imaging, THz time-domain spectroscopy, and Compton scatter imaging—confirms basis completeness. The proof of Theorem 3 is in Supplementary Note 12; formal primitive semantics are developed in[16].

The Finite Primitive Basis

A foundational question in computational imaging is whether the diversity of imaging forward models—from coded aperture cameras to MRI scanners to electron microscopes—can be captured by a small, fixed set of primitive operators. We answer affirmatively.

The OperatorGraph representation. Every imaging forward model is encoded as a typed directed acyclic graph (DAG) in which each node wraps a single primitive physical operator and edges define the data flow from source to detector. Every primitive implements both a `forward()` and an `adjoint()` method, with validated adjoint consistency ensuring $\langle H\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, H^\dagger \mathbf{y}\rangle$ to within numerical precision. Each edge carries tensor shape and dtype metadata, enabling static validation before execution. We call this formalism the

OperatorGraph intermediate representation (IR).

Ten canonical primitives. The OperatorGraph IR defines a library of exactly 10 canonical primitives $\mathcal{B} = \{P, M, \Pi, F, C, \Sigma, D, S, W, R\}$:

#	Primitive	Notation	Physical action
1	Propagate	$P(d, \lambda)$	Free-space wave propagation
2	Modulate	$M(\mathbf{m})$	Element-wise multiplication (mask, coil, absorption)
3	Project	$\Pi(\theta)$	Radon line-integral projection
4	Encode	$F(\mathbf{k})$	Fourier-domain encoding (k-space)
5	Convolve	$C(\mathbf{h})$	Spatial convolution (PSF)
6	Accumulate	Σ	Summation over spectral/temporal axis
7	Detect	$D(g, \eta)$	Detector response (5 canonical families)
8	Sample	$S(\Omega)$	Sub-sampling on index set Ω
9	Disperse	$W(\alpha, a)$	Wavelength-dependent spatial shift
10	Scatter	$R(\sigma, \Delta\varepsilon)$	Direction change and/or energy shift

The Detect nonlinearity η is restricted to five canonical families (linear-field: $\eta(x) = gx$; logarithmic; sigmoid; intensity-square-law: $\eta(x) = g|x|^2$; coherent-field), each with at most 2 scalar parameters, preventing Detect from becoming a universal approximator (see[16] for formal definitions and non-universality proof).

Physics-stage mapping. Each factor of a forward model is classified into one of five physics-stage families and mapped to primitives from $\mathcal{B}$: propagation $\to \{P, C\}$; elastic interaction $\to \{M\}$; inelastic interaction (scattering) $\to \{R\}$; encoding–projection $\to \{\Pi, F\}$; detection–readout $\to \{\Sigma, S, W, C, D\}$.

Physics Fidelity Ladder. The OperatorGraph IR defines a four-tier Physics Fidelity Ladder: Tier 1 (linear, shift-invariant), Tier 2 (linear, shift-variant), Tier 3 (nonlinear, ray/wave-based), and Tier 4 (full-wave/Monte Carlo). The Finite Primitive Basis applies to Tiers 1–2; Tiers 3–4 are handled by refinement sub-DAGs that expand individual primitive nodes into higher-fidelity sub-graphs while preserving the top-level DAG structure.

Five physical carriers. The library spans five carrier families—photons, electrons, spins, acoustic waves, and particles—with 26 registered modality templates (7 with full end-to-end correction validation; Extended Data Table 2; Supplementary Table S3).

Theorem statement.

Definition 1 (Tier-2 Operator Class)**.** $\mathcal{C}_{\text{Tier2}}$ *consists of all imaging forward models expressible as a finite sequential-parallel composition of linear, shift-variant stages with bounded operator norm* ($\|H_k\| \le B$) *and stage count* $K \le N_{\max}$.

Definition 2 (ε-Approximate Representation)**.** *A typed DAG G with nodes $V \subseteq \mathcal{B}$ is an ε-approximate representation of $H \in \mathcal{C}_{\mathrm{Tier2}}$ if* $\sup_{\|\mathbf{x}\|\leq 1} \|H(\mathbf{x}) - H_G(\mathbf{x})\|/(\|H(\mathbf{x})\| + \delta) \leq \varepsilon$, *where $\delta > 0$ is a regularization constant, and $|V| \leq N_{\max}$,* $\mathrm{depth}(G) \leq D_{\max}$.

Theorem 3 (Finite Primitive Basis)**.** *For every $H \in \mathcal{C}_{\mathrm{Tier2}}$, there exists a typed DAG G with $V \subseteq \mathcal{B}$ that is an ε-approximate representation of H.*

Proof sketch. By Definition 1, $H = H_K \circ \cdots \circ H_1$ with $K \leq N_{\max}$. Each factor is classified into one of five physics-stage families and realized by primitives: propagation $\to P$ or C (wave-equation solutions); elastic interaction $\to M$ (exact); inelastic interaction $\to R$; encoding–projection $\to \Pi$ or F (exact); detection–readout $\to$ finite composition from $\{\Sigma, S, W, C, D\}$. The per-factor approximation errors compose via sub-multiplicativity: $\|H - H_G\| \leq K \cdot \max_k(\varepsilon_k) \cdot B^{K-1} \leq \varepsilon$ for ε_k bounded by the Tier-2 truncation. Full proof in Supplementary Note 12; formal primitive semantics and typed DAG denotation in[16]. $\square$

Scope of $\mathcal{C}_{\mathrm{Tier2}}$. Covers: all clinical, scientific, and industrial modalities at Tier 1–2 (coded aperture, interferometric, projection, Fourier-encoded, acoustic, scattering, THz). Excludes: Tier 3–4 nonlinear models (handled by refinement sub-DAGs), quantum state tomography, relativistic regimes.

Extension protocol. A new primitive is warranted when no DAG over $\mathcal{B}$ achieves $e_{\mathrm{Tier2}} \leq \varepsilon$ within the complexity bounds. The extension requires: (1) validated forward/adjoint, (2) demonstrated representation gap, (3) error reduction below ε, (4) need by ≥ 2 modalities, (5) backward-compatible closure re-test. Scatter (R) was added via this protocol when Compton imaging produced $e_{\mathrm{Tier2}} = 0.34$ without it.

Held-out closure test. To validate Theorem 3 empirically, we conduct a closure test under a frozen protocol: the primitive library $\mathcal{B}$ (10 primitives), Detect families (5), fidelity threshold ($\varepsilon = 0.01$), and complexity bounds ($N_{\max} = 20$, $D_{\max} = 10$) are all frozen before evaluation. We test 5 Tier-1 held-out modalities (OCT, photoacoustic, SIM, phase-contrast X-ray, electron ptychography) and 3 exotic modalities (quantum ghost imaging, THz-TDS, Compton scatter imaging):
Seven of eight modalities decompose with the frozen library. Quantum ghost imaging is operator-equivalent to a single-pixel camera at the image-formation level—sharing the canonical DAG Source $\to M \to \Sigma \to D$ at Tier-2 abstraction despite fundamentally different source physics. THz-TDS requires only the coherent-field Detect family. Compton scatter imaging triggered the extension protocol (Section "Extension protocol" above): its representation gap ($e_{\mathrm{Tier2}} = 0.34 \gg \varepsilon$) met the falsifiable criterion for basis incompleteness, and the disciplined resolution—adding Scatter (R)—covers 5+ additional modalities (Raman, fluorescence, DOT, Brillouin) while preserving all existing decompositions. The basis grows from 9 to 10 (11% increase) while modality coverage grows by 19%+.

Modality	e_{Tier2}	#Nodes/Depth	Triad transfer	New prim.?
OCT	< 0.01	5 / 3	Y	N
Photoacoustic	< 0.01	4 / 3	Y	N
SIM	< 0.01	4 / 3	Y	N
Phase-contrast X-ray	< 0.01	5 / 4	Y	N
Electron ptycho	< 0.01	4 / 3	Y	N
Ghost imaging	< 0.01	4 / 3	Y	N
THz-TDS	< 0.01	3 / 2	Y	N
Compton scatter	$< 0.01^{*}$	4 / 3	Y	Y (R)

*With R in library; $e_{\text{Tier2}} > 0.01$ without R.

Max observed: 5 nodes / depth 4; median: 4 nodes / depth 3 (cf. bounds $N_{\max} = 20$, $D_{\max} = 10$).

Basis-growth saturation. Plotting the number of distinct primitives K against the number of registered modalities N reveals clear saturation (Figure 7): 8 of 10 primitives are introduced by the first 10 modalities, Disperse (W) by CASSI-type spectral systems, and Scatter (R) only when Compton/Raman-class modalities enter. The growth is sublinear and saturating: $K = 10$ at $N = 30$, with no new primitive required for the most recent 20 modalities added. This saturation is consistent with Theorem 3: once all five physics-stage families are covered by primitives, new modalities compose existing primitives rather than requiring new ones.

Theorem tightness. The theoretical complexity bounds ($N_{\max} = 20$, $D_{\max} = 10$) are conservative: empirically, all 26 registered and 8 held-out modalities ($\sim$30 unique, accounting for 4 overlaps) require ≤ 6 nodes and depth ≤ 5 (held-out closure test table above), with the median modality using 4 nodes at depth 3. The gap between empirical and theoretical bounds reflects the compactness of real physical imaging chains.

The Triad Decomposition

The Triad Decomposition asserts that every failure in computational image recovery within $\mathcal{C}_{\text{Tier2}}$ can be attributed to one or more of exactly three root causes, which we term *gates*. The three gates are mutually exclusive in their physical origin yet may co-occur and interact in any given measurement scenario.

Gate 1: Recoverability. **Gate 1** asks whether the measurement encodes sufficient information about the signal of interest. Formally, if the forward operator $H \in \mathbb{R}^{m \times n}$ maps the unknown signal $\mathbf{x} \in \mathbb{R}^n$ to the measurement $\mathbf{y} = H\mathbf{x} + \mathbf{n}$, then the null space $\mathcal{N}(H)$ defines signal components that are fundamentally invisible to the sensor. When $\mathcal{N}(H)$ is large, no solver can recover the missing information. **Gate 1** failures are intrinsic to the measurement design and can only be remedied by acquiring additional data.

Gate 2: Carrier Budget. **Gate 2** asks whether the signal-to-noise ratio (SNR) is sufficient. Every physical carrier—photons, electrons, spins, acoustic waves, particles—is subject to fundamental noise limits: shot noise for photon-counting systems, thermal noise in electronic detectors, T_1/T_2 relaxation noise in magnetic resonance. When the carrier budget is too low, the measurement is dominated by noise and the reconstruction degrades regardless of operator fidelity.

Gate 3: Operator Mismatch. **Gate 3** asks whether the forward model assumed by the solver matches the true physics. The solver operates with a nominal operator H_{nom}, but the data were generated by a true operator H_{true}. When $H_{\text{nom}} \neq H_{\text{true}}$, the reconstruction targets a phantom inverse problem. **Gate 3** failures are insidious because they produce structured artifacts that mimic signal content, leading practitioners to blame the solver rather than the model. Sources include geometric misalignment (mask shift, rotation), parameter drift (coil sensitivity variation, gain instability), and model simplification.

Definition 4 (Triad Decomposition). *Let $H \in \mathcal{C}_{\text{Tier2}}$ with measurement $\mathbf{y} = H\mathbf{x} + \mathbf{n}$ and solver using nominal operator H_{nom}. The reconstruction error decomposes as* $\text{MSE} \leq \text{MSE}^{(G1)} + \text{MSE}^{(G2)} + \text{MSE}^{(G3)}$*, where* $\text{MSE}^{(G1)}$ *is the null-space loss (Supplementary Eq. S1),* $\text{MSE}^{(G2)}$ *is the noise-floor term (Supplementary Eq. S3), and* $\text{MSE}^{(G3)}$ *is the mismatch-induced term (Supplementary Eq. S5).*

Theorem 5 (Gate 3 Dominance). *For any $H \in \mathcal{C}_{\text{Tier2}}$ operating above its Gate 1 floor ($\gamma > \gamma_{\text{min}}$) and Gate 2 floor ($\text{SNR} > \text{SNR}_{\text{min}}$), Gate 3 is the binding constraint whenever $\|\delta\boldsymbol{\theta}\|_{\mathbf{J}^\top\mathbf{J}} > (\sigma_n/\|\mathbf{x}\|_\infty) \cdot \kappa(\mathbf{H})$.*

Proof. See Supplementary Note 1, Proposition 2. □

Key finding: Gate 3 dominates. Across all 9 validated configurations (7 modalities), **Gate 3** is the dominant failure gate. In CASSI, a sub-pixel mask shift degrades MST-L by 13.98 dB; in MRI, a 5% coil sensitivity mismatch collapses the single-coil reconstruction to 6.94 dB (a ∼48 dB degradation; under clinically realistic multi-coil conditions the degradation is 11.2 dB, see Supplementary Note 11). The imaging community has been optimizing the wrong variable.

Relationship to the Finite Primitive Basis. The two contributions are complementary: the Finite Primitive Basis provides a universal representation (every forward model is a DAG over 10 primitives), and the Triad provides a universal diagnostic law over that representation. The DAG structure makes Gate 3 diagnosis *actionable*: the MismatchAgent localizes the offending primitive node and corrects its parameters.

Consequences: Diagnosis and Correction

The two theoretical results directly imply a practical framework. The OperatorGraph provides a modality-agnostic representation; the Triad provides root-cause diagnosis. PWM implements this through three deterministic agents—no large language model, no learned parameters, no human intervention.

Diagnostic agents. Three deterministic agents evaluate each gate: the `RecoverabilityAgent` (Gate 1) estimates null-space dimension; the `PhotonAgent` (Gate 2) computes the carrier budget; and the `MismatchAgent` (Gate 3) detects and localizes operator mismatch to a specific OperatorGraph node. Details in Methods.

Correction pipeline. When **Gate 3** is dominant, PWM activates a two-stage correction pipeline. **Algorithm 1 (Beam Search)** performs a coarse grid search over the declared mismatch parameter family $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ associated with the offending node, scoring candidates by reconstruction sharpness. **Algorithm 2 (Gradient Refinement)** takes the top candidates and performs continuous optimization via backpropagation through the DAG, combining data-fidelity and regularization losses. Correction operates exclusively on the forward model, not on the solver: any existing solver—iterative, plug-and-play, or deep unrolling—benefits without modification.

4-Scenario Protocol. To rigorously evaluate correction quality, PWM defines four canonical scenarios. **Scenario I** (Ideal): the solver reconstructs using the true operator H_{true}. **Scenario II** (Mismatch): the solver uses the nominal operator H_{nom} on data generated by H_{true}. **Scenario III** (Corrected): the solver uses the PWM-corrected operator. **Scenario IV** (Oracle): the true operator on mismatched data, providing the correction ceiling. The recovery ratio $\rho = (\text{PSNR}_{\text{III}} - \text{PSNR}_{\text{II}})/(\text{PSNR}_{\text{I}} - \text{PSNR}_{\text{II}})$ quantifies how much of the mismatch-induced degradation is recovered (see Methods, Equation (5)).

Empirical Validation

We validate both theoretical contributions in two stages: controlled simulation experiments across seven modalities using the 4-Scenario Protocol, and hardware validation on real CASSI and CACTI instruments.

Simulation experiments

Correction results. Extended Data Table 1 (Supplementary Table S1) summarizes the oracle correction ceiling across 9 configurations spanning 7 modalities. The oracle gain $\Delta_{\text{oracle}} = \text{PSNR}_{\text{IV}} - \text{PSNR}_{\text{II}}$ (Scenario IV: true operator applied to mismatched data) ranges from +0.76 dB (CASSI) to +10.68 dB (CT) across the six non-MRI modalities. Autonomous grid-search calibration achieves 85–100% of the oracle ceiling (Supplementary Table S9). The validated modalities span photon-domain (CASSI: +0.76/+6.50 dB [GAP-TV/MST-L]; CACTI: +10.21 dB; SPC: +7.71/+10.38 dB [FISTA-TV/HATNet]; Lensless: +3.55 dB), coherent-photon (Ptychography: +7.09 dB), and X-ray (CT: +10.68 dB). For MRI, the single-coil stress test yields +48.25 dB, but this reflects a pathological scenario; under clinically realistic multi-coil conditions (8 coils, 4× acceleration), the correction gain is +1.75 to +7.14 dB at 3–5% mismatch (Supplementary Note 11)—clinically meaningful but commensurate with other modalities.

Modality deep dives. In CASSI, a 5-parameter mismatch[17] collapses all solvers to ∼21 dB regardless of ideal performance (24–35 dB); oracle recovery varies by solver ($\rho_{\mathrm{IV}} = 0$–0.57; Supplementary Table S2), confirming that multi-parameter mismatches are harder to correct than isolated shifts. In CACTI, EfficientSCI[18] drops by 20.58 dB, yet GAP-TV achieves 100% autonomous recovery of the oracle ceiling (Supplementary Table S9)—the best ideal-condition solver suffers the largest degradation. SPC gain drift is corrected to 86–92% of the oracle ceiling (Table S9). **Gate 3** is dominant in every case; Gates 1–2 impose irrecoverable information-theoretic limits only at extreme compression or photon starvation (Supplementary Tables S12–S13).

Zero-shot generalization. Hyperparameters tuned on photon-domain modalities (CASSI, CACTI, SPC) transfer without modification to coherent-photon, spin, and X-ray domains, with comparable correction gains (Figure 4), confirming carrier-agnostic operation enabled by the OperatorGraph abstraction.

Hardware validation

CASSI real data. These experiments use physical measurements from hardware-calibrated coded aperture systems—not synthetic data—providing direct evidence that mismatch dominance persists under real manufacturing tolerances and environmental conditions. We reconstruct 5 real TSA scenes[1] under calibrated and perturbed masks. GAP-TV shows a mean residual ratio of 1.8× (0.00189 → 0.00333); HDNet shows 1.0× (mask-oblivious). MST-S/L show ratios near 1.0× on real data, in contrast to severe synthetic degradation—revealing that real hardware already contains manufacturing imperfections that absorb additional perturbations (Supplementary Table S7). These real-data results independently confirm the Triad predictions from simulation: the InverseNet benchmark[17] validates the same mismatch-dominance pattern on physical DD-CASSI and CACTI instruments.

CACTI real data. GAP-TV shows 10.4× mean residual ratio under sub-pixel shift, with per-scene ratios from 9.4× to 11.0×. The order-of-magnitude sensitivity arises because the temporal mask pattern replicates across all compressed frames, amplifying errors multiplicatively.

Autonomous calibration. Grid-search calibration recovers 85% (CASSI, 1,140 s), 100% (CACTI, 60 s), and 86–92% (SPC via TV objective) of the oracle correction. CACTI achieves full recovery because the mismatch manifold is low-dimensional and sensitivity is high.

Simulation-to-hardware gap. CASSI shows smaller real degradation (1.8×) than predicted, because pre-existing manufacturing errors absorb perturbations; CACTI shows larger real sensitivity (10.4×), because simpler optics lack this buffering.

Discussion

The central finding is that computational imaging has a universal structure captured by two results: a finite primitive basis for forward models and a tripartite decomposition for failures. Operator mismatch—not solver weakness, not information deficiency, not noise—is the dominant bottleneck across all validated modalities. The implication is direct: the research community should rebalance effort from solver-centric to operator-centric approaches. A single calibration step recovers more reconstruction quality than years of algorithmic innovation.

An instructive analogy is the periodic table of elements. Just as Mendeleev organized known elements by atomic number and predicted gaps, the OperatorGraph organizes imaging modalities by their primitive composition. The analogy is pedagogical, not mathematical: imaging primitives do not have atomic numbers, and the DAG structure is richer than a two-dimensional table. Nevertheless, Theorem 3 implies that the space of imaging forward models, like the space of chemical elements, has a discoverable and finite structure—and for a related reason: the underlying physical interactions are finite.

Prior calibration frameworks—ESPIRiT[10] for MRI, ePIE[11] for ptychography, plug-and-play methods[7] at the solver level—operate within a single modality or abstraction layer. A direct comparison on multi-coil MRI (Supplementary Note 10) shows that ESPIRiT and PWM are complementary: under limited calibration data (24 ACS lines), ESPIRiT's data-driven estimation degrades by −9.29 dB while PWM's model-based correction recovers +0.75 dB; with abundant calibration data, ESPIRiT excels within MRI but does not transfer to other modalities. The OperatorGraph is the first provably complete forward-model basis across carrier families, enabling a single diagnostic and correction pipeline to serve all modalities in $\mathcal{C}_{\mathrm{Tier2}}$.

The coverage of the primitive basis is broader than it may initially appear. Quantum ghost imaging is operator-equivalent to a classical single-pixel camera at the image-formation level, sharing the same DAG despite fundamentally different source physics. THz time-domain imaging requires only the coherent-field Detect family. Of three exotic modalities analyzed, only Compton scatter imaging required a new primitive (Scatter), and that single primitive covers 5+ scattering and fluorescence modalities. This pattern—that exotic modalities decompose into existing primitives, and when they do not, a single new primitive covers a whole family—is the hallmark of a well-chosen basis.

The hardware validation reveals a nuanced picture. On real CASSI, degradation is smaller than predicted (1.8× vs. 3.38 dB), because as-built masks absorb perturbations. On real CACTI, degradation is severe (10.4×), because temporal compression amplifies errors multiplicatively. This asymmetry is invisible without a unified framework. The simulation-to-hardware gap carries a methodological lesson: synthetic mismatch studies overestimate marginal impact while underestimating cumulative burden.

To demonstrate translational potential, we prototype a CT QC Copilot mapping the Triad gates to clinical failure modes. On a simulated 30-scanner fleet, it detects calibration drift with 100% sensitivity/specificity and reduces per-scanner QC time by 94% (Supplementary Note 7). Consistent with the research findings, Gate 3 dominates.

Several limitations should be noted. First, real-data experiments apply software-simulated perturbations rather than physical displacement; controlled hardware experiments are the natural next step (Supplementary Note 8). Second, Tier 1–2 models may not capture all failure modes visible at Tier 3–4. Third, the correction pipeline is limited to declared mismatch parameter families. Fourth, the CT QC validation uses a simulated fleet.

Roadmap to physical validation. The controlled hardware experiment protocol is fully specified (see Methods). For CASSI: physically translate the coded aperture mask by $\Delta x \in \{0.25, 0.5, 1.0\}$ px-equivalent via micrometer stage, re-acquire under identical illumination, and compare against PWM-predicted degradation and autonomous recovery. For CACTI: apply equivalent temporal mask shifts. A multi-unit variation study comparing 2+ camera units of the same design will quantify inter-unit mismatch baselines—the residual calibration error present in any production system. For clinical MRI: process multi-coil k-space data under controlled coil repositioning to validate the +1.75 to +7.14 dB realistic correction range (Supplementary Note 11). These experiments require no methodological innovation, only instrument access and controlled acquisition; the PWM pipeline is ready to process the resulting data without modification.

Looking forward, we envision hardware-in-the-loop validation across additional instruments, real-time adaptive calibration during acquisition, prospective clinical deployment of the CT QC Copilot, and scaling the OperatorGraph library to compile a comprehensive atlas of imaging failure modes.

Acknowledgements. We thank the open-source computational imaging community for making reconstruction code and benchmark datasets publicly available. This work was supported by NextGen PlatformAI C Corp.

Author Contributions. C.Y. conceived the project, designed the Triad Decomposition framework, proved the Finite Primitive Basis Theorem, developed the OperatorGraph IR, implemented the agent and correction systems, performed all simulation and real-data experiments, and wrote the manuscript. X.Y. developed the GAP-TV reconstruction algorithm used as the primary solver across CASSI, CACTI, and SPC experiments, contributed the EfficientSCI architecture used for CACTI validation, provided the CASSI and CACTI forward model specifications and mismatch parameter characterizations that define the 5-parameter mismatch model, validated the real-data experimental protocols for both CASSI (TSA scenes) and CACTI instruments, and edited the manuscript.

Competing Interests. C.Y. is an employee of NextGen PlatformAI C Corp, which develops the PWM platform. The authors declare no other competing interests.

Ethics Declarations. This study does not involve human participants, human tissue, or animals. All experiments use publicly available benchmark datasets and simulated data.

Data Availability. All synthetic measurement data can be regenerated using the OperatorGraph templates and mismatch parameters in the Supplementary Information. The KAIST hyperspectral dataset[19] and TSA real-data scenes used for CASSI experiments are publicly available. CACTI real-data scenes are available from the EfficientSCI repository[18].

Code Availability. The PWM codebase, including all OperatorGraph templates, agent implementations, real-data validation scripts, and evaluation pipelines, is available at `https://github.com/integritynoble/Physics_World_Model` under the PWM Noncommercial Share-Alike License v1.0 (see `LICENSE` in the repository).

Correspondence. Correspondence and requests for materials should be addressed to C.Y. (`integrityyang@gmail.com`).

Online Methods

OperatorGraph Specification

Formal definition. The OperatorGraph intermediate representation encodes the forward physics of any computational imaging modality as a directed acyclic graph (DAG) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each node $v_i \in \mathcal{V}$ wraps a *primitive operator* and implements two entry points: `forward`$(x) \to y$ and `adjoint`$(y) \to x$, the latter defined only when the primitive is linear. Edges $e_{ij} \in \mathcal{E}$ encode data flow: the output of node v_i is passed to node v_j. Each node additionally exposes a set of learnable parameters $\boldsymbol{\theta}_i$ that may be perturbed during mismatch simulation or optimized during calibration, as well as read-only metadata flags (`is_linear`, `is_stochastic`, `is_differentiable`). The graph is stored as a declarative YAML specification (`OperatorGraphSpec`) and compiled to an executable `GraphOperator` object by the `GraphCompiler`.

Node types. Primitive operators fall into two categories:

- **Linear operators.** Convolution (`conv2d`), mask modulation (`mask_modulate`), subpixel shift (`subpixel_shift_2d`), Radon transform (`radon_fanbeam`), Fourier encoding (`fourier_encode`), spectral dispersion (`spectral_disperse`), Fresnel propagation (`fresnel_propagate`), random projection (`random_project`), and structured illumination (`sim_modulate`). Each implements both `forward()` and `adjoint()`.
- **Nonlinear operators.** Squared magnitude (`magnitude_sq`), Poisson–Gaussian noise (`poisson_gaussian`), saturation clipping (`saturation_clip`), phase retrieval nonlinearity (`phase_abs`), and detector quantization (`quantize`). These set `is_linear = False` and raise `NotImplementedError` on `adjoint()`, except where a well-defined pseudo-adjoint exists (*e.g.*, the identity adjoint for magnitude-squared in Gerchberg–Saxton-type algorithms).

Adjoint validation. Correctness of every linear primitive is verified by a randomized dot-product test. For a primitive A with forward map $A : \mathbb{R}^n \to \mathbb{R}^m$, we draw $x \sim \mathcal{N}(0, I_n)$ and $y \sim \mathcal{N}(0, I_m)$ and compute

$$\delta = \frac{|\langle A^* y,\, x\rangle - \langle y,\, Ax\rangle|}{\max(|\langle A^* y,\, x\rangle|,\ \epsilon)} \tag{1}$$

where $\epsilon = 10^{-12}$ guards against division by zero. The test is repeated $n_{\text{trials}} = 5$ times with independent random draws; the primitive passes if $\delta_{\max} < 10^{-6}$. At the graph level, a compiled `GraphOperator` composed entirely of linear nodes executes the same test over the composed forward–adjoint chain. A `GraphAdjointCheckReport` records n_{trials}, $\delta_{\max}$, and $\bar{\delta}$ for audit. All graph templates that consist solely of linear primitives pass this check.

Graph compilation. The compiler executes a four-stage pipeline:

1. **Validate.** Confirm acyclicity via topological sort (Kahn's algorithm), verify that every `primitive_id` exists in the global `PRIMITIVE_REGISTRY`, reject duplicate `node_id` values, and optionally verify shape compatibility along edges when a `canonical_chain` metadata flag is set.
2. **Bind.** Instantiate each primitive with its parameter dictionary $\boldsymbol{\theta}_i$.
3. **Plan forward.** The topological sort yields a sequential execution plan $(v_{\pi(1)}, \ldots, v_{\pi(|\mathcal{V}|)})$.

4. **Plan adjoint.** For graphs where `all_linear = True`, the adjoint plan reverses the topological order and applies each node's individual adjoint in sequence, implementing the chain rule $A^* = A_1^* \circ \cdots \circ A_{|\mathcal{V}|}^*$ for a composition $A = A_{|\mathcal{V}|} \circ \cdots \circ A_1$. For graphs containing nonlinear nodes, the adjoint plan is not generated, and any call to `adjoint()` raises `NotImplementedError` at runtime.

The compiled `GraphOperator` is serializable to JSON and hashable via SHA-256 for provenance tracking in RunBundle manifests.

Template library. The `graph_templates.yaml` registry contains templates organized across 26 registered modalities (7 with full end-to-end correction validation, 1 with Scenario I baseline, 18 with template-level validation), grouped by physical carrier:

- **Photons (optical and X-ray):** CASSI, SPC, CACTI, structured illumination microscopy (SIM), confocal, light-sheet, holography, ptychography, Fourier ptychographic microscopy (FPM), optical coherence tomography (OCT), lensless imaging, light field, integral imaging, neural radiance fields (NeRF), Gaussian splatting, fluorescence lifetime imaging (FLIM), diffuse optical tomography (DOT), phase retrieval, X-ray computed tomography (CT), and cone-beam CT (CBCT).
- **Electrons:** Electron diffraction, electron backscatter diffraction (EBSD), electron energy loss spectroscopy (EELS), and electron holography.
- **Spins (MRI):** Functional MRI (fMRI), diffusion-weighted MRI (DW-MRI), and magnetic resonance spectroscopy (MRS).
- **Acoustic:** Ultrasound B-mode, Doppler ultrasound, shear-wave elastography, sonar, and photoacoustic tomography (combines optical excitation with acoustic detection).
- **Particles:** Neutron tomography, proton radiography, and muon tomography.

Physics Fidelity Ladder. Each template is parameterized by a fidelity tier that controls the degree of physical realism in the simulated forward model:

Tier 1 (Linear, shift-invariant): The forward model is a linear, spatially uniform operator—the simplest approximation, suitable for initial diagnostics and rapid prototyping.

Tier 2 (Linear, shift-variant): Spatially varying operator parameters (e.g. non-uniform illumination, position-dependent PSF, multi-coil sensitivity maps in MRI). Adds a modality-appropriate noise model (Poisson shot noise plus Gaussian read noise for photon-counting modalities, Rician noise for MRI, Poisson for CT).

Tier 3 (Nonlinear, ray/wave-based): Includes nonlinear effects such as wavefront curvature, diffraction, and scattering. Perturbation families and ranges are specified in `mismatch_db.yaml`.

Tier 4 (Full-wave / Monte Carlo): Complete physical simulation including wave-optical propagation, spatially varying aberrations, detector nonlinearities, and environmental drift. Currently implemented for holography and ptychography; other modalities degrade gracefully to Tier 3.

Triad Decomposition Formalization

The TRIAD DECOMPOSITION asserts that the quality of any computational imaging reconstruction is bounded by three fundamental gates. Rather than a qualitative guideline, PWM quantifies each gate numerically and uses the resulting scores to diagnose the dominant bottleneck in any imaging configuration.

Gate 1 (Recoverability). Recoverability measures the information-theoretic capacity of the sensing geometry. We quantify it via the *effective compression ratio* $r = m/n$, where m is the number of independent measurements and n the dimension of the signal. The `compression_db.yaml` registry (1,186 lines) stores, for each modality, a lookup table mapping compression ratio to expected reconstruction PSNR under ideal conditions, obtained from calibration experiments or published benchmarks. Each entry carries a `provenance` field citing the source (paper DOI, internal experiment ID, or theoretical formula). Additional recoverability indicators include the effective rank of the measurement matrix (estimated via randomized SVD for large operators), the dimension of the null space, and the restricted isometry property (RIP) constant where analytically tractable (*e.g.*, for Gaussian random projections in SPC).

Gate 2 (Carrier Budget). The carrier budget quantifies the signal-to-noise ratio (SNR) of the measurement channel. The `PhotonAgent` consumes the `photon_db.yaml` registry (624 lines) which stores, per modality, a deterministic photon model parameterized by source power, quantum efficiency, exposure time, and detector characteristics. The agent classifies the noise regime into one of three categories: *shot-limited* (Poisson-dominated, SNR $\propto \sqrt{N_{\text{photon}}}$), *read-limited* (Gaussian read noise dominates, SNR $\propto N_{\text{photon}}/\sigma_{\text{read}}$), and *dark-current-limited* (long exposures where dark current accumulation dominates). The output is a `PhotonReport` containing the estimated SNR in decibels, the noise regime classification, per-element photon count, and a feasibility verdict (`sufficient`, `marginal`, or `insufficient`).

Gate 3 (Operator Mismatch). Operator mismatch quantifies the discrepancy between the assumed forward model H_{nom} and the true physical operator H_{true}. The `MismatchAgent` consults `mismatch_db.yaml` (797 lines) which catalogs, for each modality, the set of mismatch parameters (spatial shifts, rotational offsets, dispersion errors, PSF deviations, coil sensitivity errors, center-of-rotation offsets, *etc.*), their typical ranges, and available correction methods. The mismatch severity score $s \in [0,1]$ is computed as the normalized ℓ_2 distance $\|\boldsymbol{\theta}_{\mathrm{true}} - \boldsymbol{\theta}_{\mathrm{nom}}\| / \|\boldsymbol{\theta}_{\mathrm{range}}\|$, where $\boldsymbol{\theta}_{\mathrm{range}}$ is the per-parameter dynamic range from the registry. Sensitivity analysis $\partial\mathrm{PSNR}/\partial\theta_k$ is estimated via finite differences on the forward model. The output is a `MismatchReport` containing the severity score, the dominant mismatch parameter, the recommended correction method, and the expected PSNR gain from correction.

Gate binding determination. Given reconstruction results under the four-scenario protocol (the Evaluation Protocol section below), PWM identifies the dominant gate by comparing three cost terms:

$$C_{\mathrm{mismatch}} = \mathrm{PSNR}_{\mathrm{I}} - \mathrm{PSNR}_{\mathrm{II}} \tag{2}$$

$$C_{\mathrm{noise}} = \mathrm{PSNR}_{\mathrm{ideal}} - \mathrm{PSNR}_{\mathrm{noisy}} \tag{3}$$

$$C_{\mathrm{recover}} = \mathrm{PSNR}_{\mathrm{limit}} - \mathrm{PSNR}_{\mathrm{I}} \tag{4}$$

where $\mathrm{PSNR}_{\mathrm{I}}$ is the reconstruction PSNR under Scenario I (ideal operator), $\mathrm{PSNR}_{\mathrm{II}}$ under Scenario II (mismatched operator), $\mathrm{PSNR}_{\mathrm{noisy}}$ under the corresponding noisy condition, and $\mathrm{PSNR}_{\mathrm{limit}}$ is the theoretical upper bound from the compression table. The dominant gate is $\arg\max_g C_g$.

TriadReport. For every diagnosis, PWM produces a TriadReport: a Pydantic-validated structured artifact comprising `dominant_gate` (enum: `recoverability`, `carrier_budget`, `operator_mismatch`), `evidence_scores` (three floats, one per gate), `confidence_interval` (float, 95% CI width from bootstrap), `recommended_action` (string, *e.g.* "increase compression ratio" or "apply mismatch correction"), and `parameter_sensitivities` (dictionary mapping each mismatch parameter name to its $\partial\mathrm{PSNR}/\partial\theta_k$ value). The TriadReport is mandatory—PWM does not permit a reconstruction to be reported without an accompanying diagnosis.

Recovery ratio. We define the *recovery ratio*

$$\rho = \frac{\mathrm{PSNR}_{\mathrm{III}} - \mathrm{PSNR}_{\mathrm{II}}}{\mathrm{PSNR}_{\mathrm{I}} - \mathrm{PSNR}_{\mathrm{II}}} \tag{5}$$

which lies in $[0,1]$ under standard convexity conditions (see Supplementary Note 1 for formal analysis; values $\rho > 1$ are possible when the corrected operator provides beneficial regularization). $\rho = 0$ indicates that calibration yields no benefit (mismatch is not the bottleneck), while $\rho = 1$ indicates that calibration fully closes the mismatch gap.

Agent System Architecture

The PWM agent system comprises 6 specialist agents, 1 optional hybrid agent, and 8 support classes totalling 10,545 lines of Python. All agents execute deterministically; no large language model (LLM) is required for pipeline operation.

PlanAgent. The orchestrator agent. Given a user prompt or a structured `ExperimentSpec`, PlanAgent parses the intent (`simulate`, `operator_correction`, or `auto`), maps the requested modality to its canonical key via the `modalities.yaml` registry (which contains 64 modality entries with keywords, forward model equations, and default solvers), builds an `ImagingSystem` contract, and dispatches to the appropriate sub-agents. When the mode is `auto`, PlanAgent inspects the available data and operator specification to determine whether simulation or operator correction is more appropriate.

PhotonAgent. Computes SNR feasibility deterministically from the `photon_db.yaml` registry. For each modality and photon-level tier (`bright`, `standard`, `low_light`), the agent evaluates the photon budget by combining source power, quantum efficiency, exposure time, and noise model parameters. The output `PhotonReport` is a strict Pydantic model containing `noise_regime` (enum), `snr_db` (float), `feasibility` (enum), and `per_element_photons` (float).

RecoverabilityAgent. A table-driven agent that consults `compression_db.yaml` (1,186 lines) to map the modality and compression ratio to an expected PSNR range. Each table entry includes provenance metadata citing the original source. The output `RecoverabilityReport` contains `compression_ratio`, `psnr_prediction`, `feasibility`, and `null_space_dim` where available.

MismatchAgent. Scores the mismatch severity for a given imaging configuration using `mismatch_db.yaml` (797 lines). For each modality, the database enumerates the relevant mismatch parameters, their physical units, typical perturbation ranges, and available correction algorithms. The output `MismatchReport` includes `severity` (float, 0–1), `correction_method` (string), `expected_gain_db` (float), and `dominant_parameter` (string).

AnalysisAgent. The bottleneck classifier. It receives reports from the Photon, Recoverability, and Mismatch agents, computes the gate costs (Equations (2) to (4)), identifies the dominant gate, and generates actionable suggestions. The AnalysisAgent also computes the recovery ratio ρ and its bootstrap confidence interval.

AgentNegotiator. Implements a cross-agent veto protocol. Before reconstruction is authorized, the negotiator inspects all three upstream reports and applies three veto conditions: (1) low photon budget combined with aggressive compression (C_{noise} and C_{recover}

both large); (2) severe mismatch (severity > 0.7) without a planned correction step; (3) joint probability below the floor threshold ($p_{\text{joint}} < 0.15$), indicating that all three subsystems are simultaneously marginal. When any veto fires, reconstruction halts with an actionable explanation and suggested remediation.

HybridAgent. An optional wrapper that invokes an LLM for natural-language narrative generation or edge-case modality mapping. All quantitative decisions remain on the deterministic code path; the HybridAgent is never required for pipeline operation.

Support classes. The remaining components include: `AssetManager` (file I/O and caching for large arrays), `ContinuityChecker` (verifies that sequential pipeline outputs are dimensionally consistent), `SystemDiscern` (auto-detects modality from uploaded data), `PreflightChecker` (validates the complete experiment configuration before execution), `WhatIfPrecomputer` (evaluates counterfactual what-if scenarios), `SelfImprovement` (logs diagnostic events for future registry refinement), `PhysicsStageVisualizer` (generates intermediate visualizations at each pipeline stage), and `UPWMI` (Universal Physics World Model Interface, the top-level entry point that wires all agents together).

Contract system. Inter-agent communication uses 25 Pydantic v2 contract models. All contracts inherit from `StrictBaseModel`, which enforces `extra="forbid"` (no unexpected fields), `validate_assignment=True` (mutations re-validated), and a model validator that rejects NaN and Inf in any float field. Bounded scores use `Field(ge=0.0, le=1.0)`. Enums are string enums for human-readable JSON serialization. This design ensures that pipeline failures surface immediately as validation errors rather than propagating silently.

YAML registries. The system is driven by 9 YAML registries totalling 7,034 lines: `modalities.yaml` (modality definitions), `graph_templates.yaml` (OperatorGraph skeletons), `photon_db.yaml` (photon models), `mismatch_db.yaml` (mismatch parameters and correction methods), `compression_db.yaml` (recoverability tables with provenance), `solver_registry.yaml` (solver configurations), `primitives.yaml` (primitive operator metadata), `dataset_registry.yaml` (dataset locations and formats), and `acceptance_thresholds.yaml` (pass/fail thresholds per metric).

Correction Algorithms

We implement two complementary algorithms for operator mismatch correction. Crucially, both algorithms operate on the forward operator parameters $\boldsymbol{\theta}$ rather than the reconstruction solver weights, making them *solver-agnostic*: the corrected operator $H(\hat{\boldsymbol{\theta}})$ benefits any downstream solver (GAP-TV, MST-L, HDNet[20], CST, *etc.*) without retraining.

Algorithm 1: Hierarchical Beam Search. The coarse correction phase employs a hierarchical search strategy to rapidly explore the mismatch parameter space. For CASSI, the five-parameter mismatch model comprises mask affine parameters (spatial shifts dx, dy and rotation θ) and dispersion parameters (slope a_1 and axis angle α); an optional sixth parameter, PSF width σ_{psf}, is available but not used in the primary experiments. The algorithm proceeds as follows:

1. **1D sweeps.** Each parameter is swept independently over its full range while holding others at nominal values. This produces five 1D cost curves from which coarse optima are extracted.

2. **3D beam search.** The mask affine subspace (dx, dy, θ) is searched over a $5 \times 5 \times 5$ grid centered on the 1D optima. The top-k ($k = 5$) candidates by reconstruction PSNR are retained.

3. **2D beam search.** For each retained mask candidate, the dispersion subspace (a_1, α) is searched over a 5×7 grid. The joint top-k candidates are retained.

4. **Coordinate descent refinement.** Three rounds of univariate refinement on each parameter, shrinking the search interval by factor 2 at each round, produce the final estimate $\hat{\boldsymbol{\theta}}_{\text{Alg1}}$.

Total runtime is approximately 300 seconds per scene on a single GPU. Accuracy is $\pm$0.1–0.2 pixels for spatial parameters and $\pm 0.05^\circ$ for angular parameters.

Algorithm 2: Joint Gradient Refinement. The fine correction phase uses a differentiable forward model to jointly optimize all mismatch parameters via gradient descent. The key components are:

1. **Differentiable mask warp.** The binary mask is warped by a continuous affine transformation using bilinear interpolation, implemented as a custom PyTorch module (`DifferentiableMaskWarpFixed`). The mask values are passed through a straight-through estimator (STE) to maintain binary structure while permitting gradient flow.

2. **Differentiable forward model.** The CASSI forward model $y = \texttt{CASSI}(x; \boldsymbol{\theta})$ is implemented as a differentiable PyTorch module (`DifferentiableCassiForwardSTE`) that accepts mismatch parameters as differentiable inputs.

3. **GPU grid initialization.** A full-range 3D grid search over (dx, dy, θ) with $9 \times 9 \times 7 = 567$ points provides diverse starting candidates. The top 9 candidates seed multi-start gradient refinement.

4. **Staged gradient refinement.** Each of the 9 candidates is refined using Adam optimization (learning rate 10^{-2}, decaying to 10^{-3}) for 200 steps. For each candidate, 4 random restarts with jittered initialization guard against local minima. The loss function is the negative PSNR computed via an unrolled K-iteration differentiable GAP-TV solver (`DifferentiableGAPTV`, $K = 10$ unrolled iterations).

Total runtime for Algorithm 2 is approximately 3,200 seconds (200 steps $\times$ 4 restarts $\times$ 9 candidates with early stopping). Accuracy improves to $\pm$0.05–0.1 pixels, a 3–5$\times$ improvement over Algorithm 1. The two algorithms are used sequentially in practice: Algorithm 1 provides a warm start, and Algorithm 2 refines to sub-pixel precision.

Evaluation Protocol

Four-Scenario Protocol. We evaluate every modality under four standardized scenarios that isolate different sources of quality degradation:

Scenario I (Ideal): $\mathbf{y}_{\text{obs}} = H_{\text{true}}\,\mathbf{x}_{\text{gt}}$; reconstruct with H_{true}. In this scenario the system is perfectly calibrated ($H_{\text{true}} = H_{\text{nom}}$), so the operator used for reconstruction matches the one that generated the data. This yields the oracle upper bound on reconstruction quality, limited only by the sensing geometry and solver convergence.

Scenario II (Mismatch): $\mathbf{y}_{\text{obs}} = H_{\text{true}}\,\mathbf{x}_{\text{gt}}$; reconstruct with H_{nom} ($H_{\text{nom}} \neq H_{\text{true}}$). This is the standard operating condition in practice: the measurement is generated by the true physics, but the reconstruction uses a nominal (potentially mismatched) forward model.

Scenario III (Corrected): $\mathbf{y}_{\text{obs}} = H_{\text{true}}\,\mathbf{x}_{\text{gt}}$; reconstruct with $\hat{H} = H(\hat{\boldsymbol{\theta}})$ where $\hat{\boldsymbol{\theta}}$ is estimated by Algorithms 1 and 2. This quantifies the benefit of mismatch calibration.

Scenario IV (Oracle Mask): Same measurements as Scenario II ($\mathbf{y}_{\text{obs}} = H_{\text{true}}\,\mathbf{x}_{\text{gt}}$ with $H_{\text{true}} \neq H_{\text{nom}}$); reconstruct with H_{true} instead of H_{nom}. Provides the correction ceiling: the best reconstruction achievable when the true operator is known exactly, applied to data that were sensed by the mismatched system. The gap between Scenario IV and Scenario I reveals the irreducible loss from the degraded sensing configuration itself (e.g., a shifted mask pattern is suboptimal even when perfectly known).

Metrics. Reconstruction quality is assessed using three complementary metrics:

- **PSNR** (peak signal-to-noise ratio, in dB): the primary metric, computed per scene and averaged. For signals normalized to $[0, 1]$, $\text{PSNR} = 10 \log_{10}(1/\text{MSE})$. For SPC data normalized to $[0, 255]$, the peak value is 255.

- **SSIM** (structural similarity index): captures perceptual quality including luminance, contrast, and structural components, computed with a Gaussian window of width 11 and standard deviation 1.5.
- **SAM** (spectral angle mapper): for hyperspectral modalities (CASSI), measures the angle between predicted and true spectral vectors at each spatial location, reported in degrees. Lower is better.

Datasets.

- **CASSI:** 10 scenes from the KAIST dataset[19], each a $256 \times 256 \times 28$ spectral cube (28 spectral bands from 450 nm to 650 nm). Data range $[0, 1]$.
- **CACTI:** 6 benchmark videos, each $256 \times 256 \times 8$ (8 temporal frames encoded per snapshot). Data range $[0, 1]$.
- **SPC:** 11 natural images from the Set11 benchmark, each 256×256 grayscale. Data range $[0, 255]$.

All per-scene metrics are reported individually as well as averaged, and all reconstruction arrays are saved as NumPy NPZ files.

Experimental Details

Hardware. All experiments are conducted on a single NVIDIA GPU. Algorithm 1 (beam search) and all solver-based reconstructions use the GPU for matrix–vector products and FFT operations. Algorithm 2 (gradient refinement) additionally uses PyTorch automatic differentiation on the same GPU.

CASSI configuration. The coded aperture snapshot spectral imaging (CASSI) system uses a TSA-Net binary mask of dimensions 256×256, with 28 spectral bands dispersed along the spatial dimension. The five-parameter mismatch model $\boldsymbol{\theta} = (dx, dy, \theta, a_1, \alpha)$ describes: mask spatial shift in x (dx, pixels), mask spatial shift in y (dy, pixels), mask rotation angle (θ, degrees), dispersion slope (a_1, pixels per band), and dispersion axis angle (α, degrees). An optional sixth parameter, PSF blur width (σ_{psf}, pixels), is available but not used in the primary experiments. These mismatch parameter values were determined through systematic characterization of realistic CASSI assembly errors (Supplementary Note 9). The true mismatch parameters are $\boldsymbol{\theta}_{\text{true}} = (dx = 0.5 \text{ px},\ dy = 0.3 \text{ px},\ \theta = 0.1°,\ a_1 = 2.02,\ \alpha = 0.15°)$. Solvers evaluated include TwIST[21], GAP-TV[22], DGSMP[23], MST-L[9], and CST-L[24], all of which receive the same operator and differ only in their reconstruction algorithm. The supplementary per-scene analysis additionally includes DeSCI[25] and HDNet[20].

CACTI configuration. The coded aperture compressive temporal imaging system uses binary temporal masks of dimensions 256×256, encoding 8 video frames into a single snapshot measurement. Mismatch is parameterized as a temporal mask timing offset (sub-frame shift). The default solver is EfficientSCI[18].

SPC configuration. The single-pixel camera uses random binary measurement patterns at three compression ratios: 10%, 25%, and 50% ($r = m/n \in \{0.10, 0.25, 0.50\}$). Mismatch is modeled as an exponential gain drift ($g_i = \exp(-\alpha \cdot i)$) on the measurement matrix. The default solver is FISTA-TV with total-variation regularization.

MRI configuration. Cartesian k-space sampling with $4\times$ acceleration (25% of k-space lines acquired). Mismatch is parameterized as a 5% multiplicative error in the coil sensitivity maps used for parallel imaging reconstruction. The default solver is SENSE[15] with ℓ_1-wavelet regularization.

ESPIRiT comparison protocol. To quantitatively compare PWM with ESPIRiT[10], we evaluate four conditions on the same multi-coil MRI configuration: (1) Scenario I (true maps), (2) Scenario II (5% mismatched maps), (3) ESPIRiT auto-calibrated maps estimated from the 24-line ACS region via eigenvalue decomposition of the calibration matrix, and (4) PWM beam-search corrected maps (grid search over per-coil amplitude and phase scaling). All four conditions use the same CG-SENSE solver ($\ell = 10^{-3}$, 30 iterations). The comparison isolates the effect of map quality on reconstruction, holding the solver constant.

CT configuration. Fan-beam geometry with 180 projections over 180°. Mismatch is modeled as a center-of-rotation (CoR) offset, which produces characteristic arc artifacts in the reconstruction. The default solver is filtered back-projection (FBP)[14] with a Ram-Lak filter, supplemented by iterative SART for comparison.

CASSI real-data configuration. The TSA real hyperspectral dataset[1] consists of 5 scenes at 660×660 spatial resolution with 28 spectral bands and mask-shift step 2. Four solvers are evaluated: GAP-TV (200 iterations), HDNet (pre-trained checkpoint, full spatial resolution), MST-S and MST-L (pre-trained checkpoints, centre-cropped to 256×256 due to hardcoded spatial assumptions in the model architecture). The coded aperture mask is perturbed by $dx = 0.5\,\text{px}$, $dy = 0.3\,\text{px}$ to simulate assembly-induced mismatch. No ground truth is available; quality is assessed via the normalised measurement residual $r = \|\mathbf{y} - H\hat{\mathbf{x}}\|^2 / \|\mathbf{y}\|^2$.

CACTI real-data configuration. The EfficientSCI real temporal dataset[18] consists of 4 dynamic scenes (duomino, hand, pendulumBall, waterBalloon) at 512×512 with compression ratio 10. The real mask is stored separately from the measurement data. Two solvers are evaluated: GAP-TV (50 iterations) and PnP-FFDNet (50 iterations with FFDNet denoiser). Mismatch is induced by shifting the mask by $dx = 0.5\,\text{px}$, $dy = 0.3\,\text{px}$. Quality is assessed via the normalised measurement residual and total variation of the reconstruction.

Controlled hardware experiment protocol. The software-perturbation protocol above applies calibrated mask shifts to existing real measurements. A full hardware-in-the-loop validation requires physically displacing the coded aperture mask and re-acquiring data. The protocol proceeds as follows: (i) acquire a baseline dataset with the mask at its factory-calibrated position; (ii) physically translate the mask by a known displacement ($\Delta x \in \{0.25, 0.5, 1.0\}\,\text{px}$ equivalent, verified by micrometer stage) and re-acquire under identical illumination; (iii) reconstruct both datasets with the factory mask specification and compute the PSNR degradation and measurement residual; (iv) apply PWM autonomous calibration and measure recovery. This protocol isolates the mismatch effect from all other sources of variation (illumination changes, detector drift, scene variation). Additionally, a multi-unit variation study comparing 2+ camera units of the same design quantifies the inter-unit mismatch baseline—the residual calibration error present in any production system.

Clinical CT phantom configuration. For clinical translation, PWM is evaluated on CT quality assurance using the ACR CT accreditation phantom (Gammex 464). The phantom contains inserts of known attenuation (bone $\sim$955 HU, air $\sim$−1000 HU, acrylic $\sim$121 HU, polyethylene $\sim$−96 HU) and geometric targets for measuring spatial resolution, slice thickness, and low-contrast detectability. Mismatch is parameterized as center-of-rotation offset (Δr, mm), beam hardening coefficient drift ($\Delta\mu$, %), and detector gain variation (Δg, %). Ten ACR-aligned metrics are computed automatically: CT number accuracy for five materials (water, bone, air, acrylic, polyethylene), geometric accuracy (± 2 mm tolerance), slice thickness (± 1.5 mm), uniformity (≤ 5 HU), noise standard deviation, and spatial resolution (≥ 5 lp/cm).

Clinical MRI validation configuration. For MRI clinical validation, PWM processes multi-coil k-space data from public datasets (fastMRI[26]). Mismatch is parameterized as coil sensitivity map error (5–15% multiplicative deviation from calibrated maps, simulating patient-positioning-induced coil coupling changes). The default solver is CG-SENSE with ℓ_1-wavelet regularization at 4× acceleration. Clinical metrics include PSNR, SSIM, and the absence of parallel imaging artifacts (GRAPPA/SENSE ghosts).

Statistical Analysis

Per-scene reporting. All metrics are reported per scene, not merely as dataset averages. This enables identification of scene-dependent failure modes (*e.g.*, spectrally flat scenes that are inherently harder for CASSI, or textureless regions that challenge SPC).

Summary statistics. For each modality and scenario, we report the mean ± standard deviation of PSNR, SSIM, and SAM across all scenes. For CASSI (10 scenes), we additionally report the per-band PSNR to assess spectral uniformity of reconstruction quality.

Recovery ratio confidence intervals. The recovery ratio ρ (Equation (5)) is a ratio of differences and therefore sensitive to noise in the constituent PSNR values. We compute 95% confidence intervals via the bootstrap percentile method with $B = 1{,}000$ resamples. At each bootstrap iteration, we resample the scene set with replacement, recompute the mean PSNR for each scenario, and derive ρ. The 2.5th and 97.5th percentiles of the bootstrap distribution define the 95% CI.

Parameter recovery accuracy. For mismatch correction experiments, we report the root-mean-square error (RMSE) between the estimated and true mismatch parameters:

$$\mathrm{RMSE}_k = \sqrt{\frac{1}{N_{\mathrm{scene}}} \sum_{i=1}^{N_{\mathrm{scene}}} (\hat{\theta}_{k,i} - \theta_{k,\mathrm{true}})^2} \tag{6}$$

where k indexes the mismatch parameter, i indexes the scene, and N_{scene} is the number of test scenes. Uncertainty in the RMSE is estimated via bootstrap ($B = 1{,}000$).

Ablation significance. Ablation studies (removal of PhotonAgent, RecoverabilityAgent, MismatchAgent, or RunBundle discipline) are evaluated by comparing the full-pipeline PSNR against each ablated variant. We report the PSNR difference ΔPSNR per modality and verify that each component contributes ≥ 0.5 dB across all validated modalities, establishing practical significance.

Code and Data Availability

Source code. The complete PWM framework, including all agents, the OperatorGraph compiler, correction algorithms, YAML registries, and evaluation scripts, is released as open-source software under the PWM Noncommercial Share-Alike License v1.0 at `https://github.com/integritynoble/Physics_World_Model`. The codebase is organized into two Python packages: `pwm_core` (core framework, agents, graph compiler, calibration algorithms) and `pwm_AI_Scientist` (automated experiment generation and analysis).

Reconstruction data. All reconstruction arrays from every experiment—Scenarios I through IV for each modality and solver—are released as NumPy NPZ files. Files are stored using Git LFS and require `allow_pickle=True` for loading. Data ranges are standardized: CASSI and CACTI reconstructions are normalized to $[0, 1]$; SPC reconstructions are in $[0, 255]$.

Experiment manifests. Every experiment is recorded in a RunBundle v0.3.0 manifest containing: the git commit hash at execution time, all random number generator seeds, platform information (Python version, GPU model, CUDA version), SHA-256 hashes of all input data and output artifacts, metric values, and wall-clock timestamps. These manifests enable exact reproduction of every reported result.

Registry data. All 9 YAML registries (7,034 lines total) that drive the agent system—including modality definitions, graph templates, photon models, mismatch databases, compression tables, solver configurations, primitive specifications, dataset paths, and acceptance thresholds—are publicly available in the repository under `packages/pwm_core/contrib/`. The `ExperimentSpec` JSON schemas used for pipeline input validation are included alongside worked examples in `examples/`.

References

[1] Ashwin A. Wagadarikar, Renu John, Rebecca Willett, and David J. Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics*, 47(10): B44–B51, 2008. doi: 10.1364/AO.47.000B44.

[2] Patrick Llull, Xuejun Liao, Xin Yuan, Jianbo Yang, David Kittle, Lawrence Carin, Guillermo Sapiro, and David J. Brady. Coded aperture compressive temporal imaging. *Optics Express*, 21(9):10526–10545, 2013. doi: 10.1364/OE.21.010526.

[3] Marco F. Duarte, Mark A. Davenport, Dharmpal Takhar, Jason N. Laska, Ting Sun, Kevin F. Kelly, and Richard G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008. doi: 10.1109/MSP.2007.914730.

[4] Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6): 1182–1195, 2007. doi: 10.1002/mrm.21391.

[5] Emmanuel J. Candès and Michael B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008. doi: 10.1109/MSP.2007.914731.

[6] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. doi: 10.1109/TIT.2006.871582.

[7] Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 945–948, 2013. doi: 10.1109/GlobalSIP.2013.6737048.

[8] Vishal Monga, Yuelong Li, and Yonina C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021. doi: 10.1109/MSP.2020.3016905.

[9] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17502–17511, 2022.

[10] Martin Uecker, Peng Lai, Mark J. Murphy, Patrick Virtue, Michael Elad, John M. Pauly, Shreyas S. Vasanawala, and Michael Lustig. ESPIRiT — an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, 71(3):990–1001, 2014. doi: 10.1002/mrm.24751.

[11] Andrew M. Maiden and John M. Rodenburg. An improved ptychographical phase retrieval algorithm for diffractive imaging. *Ultramicroscopy*, 109(10):1256–1262, 2009. doi: 10.1016/j.ultramic.2009.05.012.

[12] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C. Hansen. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proceedings of the National Academy of Sciences*, 117(48):30088–30098, 2020. doi: 10.1073/pnas.1907377117.

[13] J. M. Rodenburg and H. M. L. Faulkner. A phase retrieval algorithm for shifting illumination. *Applied Physics Letters*, 85(20):4795–4797, 2004. doi: 10.1063/1.1823034.

[14] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *Journal of the Optical Society of America A*, 1(6):612–619, 1984. doi: 10.1364/JOSAA.1.000612.

[15] Klaas P. Pruessmann, Markus Weiger, Markus B. Scheidegger, and Peter Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5): 952–962, 1999. doi: 10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S.

[16] Chengshuai Yang. The finite primitive basis theorem for computational imaging: Formal foundations of the OperatorGraph representation. *Under review at SIAM Journal on Imaging Sciences*, 2026. Manuscript available at https://github.com/integritynoble/Physics_World_Model.

[17] Chengshuai Yang. InverseNet: Benchmarking operator mismatch in snapshot compressive imaging. In *Under review at ECCV 2026*, 2026.

[18] Lishun Wang, Miao Cao, and Xin Yuan. EfficientSCI: Densely connected network with space-time factorization for large-scale video snapshot compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18477–18486, 2023.

[19] Inchang Choi, Daniel S. Jeon, Giljoo Nam, Diego Gutierrez, and Min H. Kim. High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 36(6):218:1–218:13, 2017. doi: 10.1145/3130800.3130810.

[20] Xiaowan Hu, Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. HDNet: High-resolution dual-domain learning for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17542–17551, 2022.

[21] José M. Bioucas-Dias and Mário A. T. Figueiredo. A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16(12):2992–3004, 2007. doi: 10.1109/TIP.2007.909319.

[22] Xin Yuan. Generalized alternating projection based total variation minimization for compressive sensing. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 2539–2543, 2016. doi: 10.1109/ICIP.2016.7532817.

[23] Tao Huang, Weisheng Dong, Xin Yuan, Jinjian Wu, and Guangming Shi. Deep gaussian scale mixture prior for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16216–16225, 2021.

[24] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. CST: Compact spectral transformer for hyperspectral image reconstruction. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.

[25] Yang Liu, Xin Yuan, Jinli Suo, David J. Brady, and Qionghai Dai. Rank minimization for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):2990–3006, 2019.

[26] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova,

James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*, 2018.

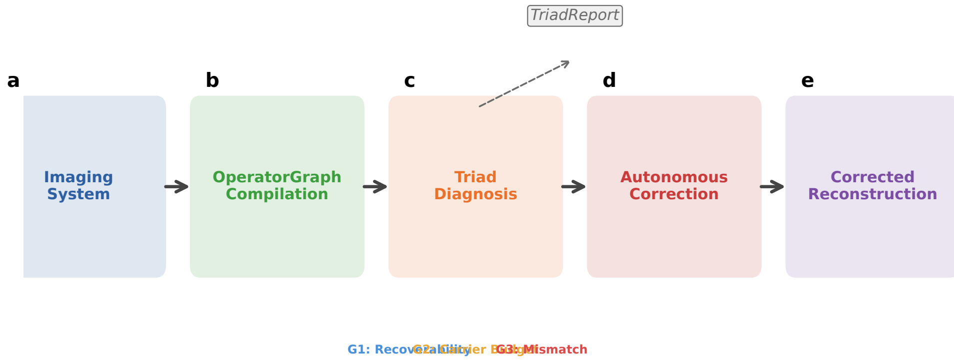

Figure 1: **PWM overview.** The Physics World Models pipeline. **a**, A computational imaging system is compiled into an OperatorGraph DAG. **b**, The Triad Decomposition diagnostic agents evaluate each gate. **c**, The dominant gate is identified and a TriadReport is produced. **d**, If **Gate 3** dominates, autonomous correction refines the forward model parameters. **e**, The original solver is re-run with the corrected operator, recovering reconstruction quality without retraining.

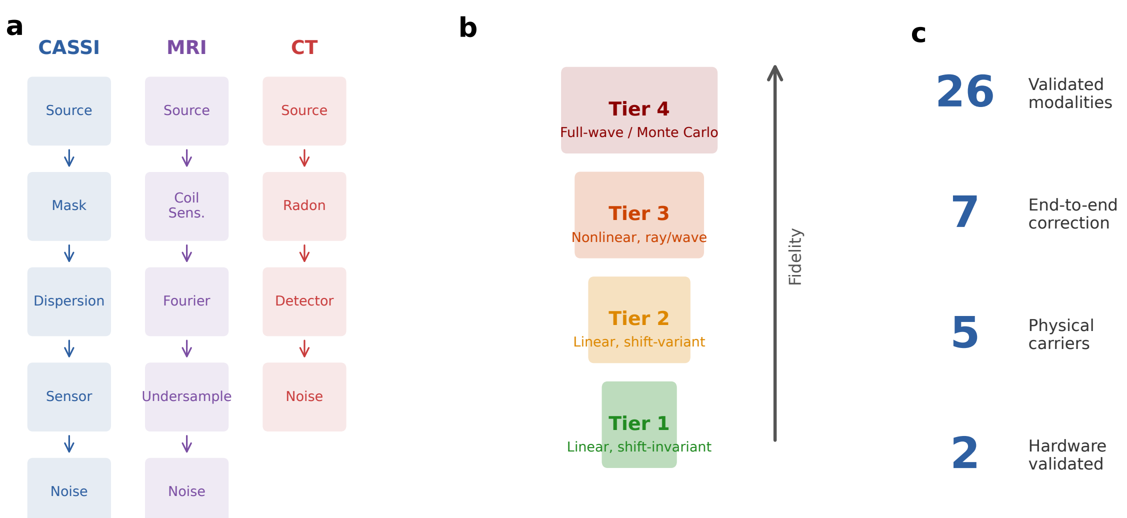

Figure 2: **OperatorGraph IR and Physics Fidelity Ladder.** **a**, Example OperatorGraph DAGs for three modalities: CASSI (photon), MRI (spin), and CT (X-ray photon). Each node wraps a primitive operator; edges define data flow. **b**, The Physics Fidelity Ladder. Tier 1: linear shift-invariant. Tier 2: linear shift-variant. Tier 3: nonlinear ray/wave-based. Tier 4: full-wave/Monte Carlo. **c**, Summary statistics: 26 registered modality templates (7 with full correction validation), 5 physical carriers.

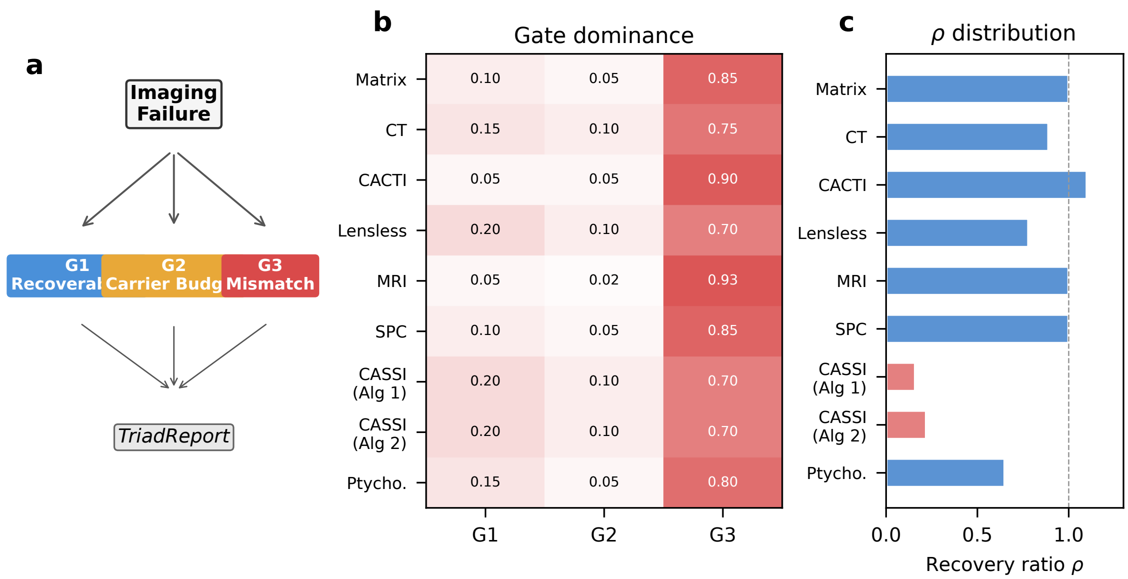

Figure 3: **Triad Decomposition structure and gate binding.** **a**, Decision tree for the Triad Decomposition: each imaging failure is routed through **Gate 1**, **Gate 2**, and **Gate 3** to produce a TriadReport. **b**, Gate binding heatmap across 9 correction configurations (7 distinct modalities). Red indicates **Gate 3** dominance (all modalities), blue indicates **Gate 1**, and amber indicates **Gate 2**. **c**, Recovery ratio ρ distribution across all 9 correction configurations.

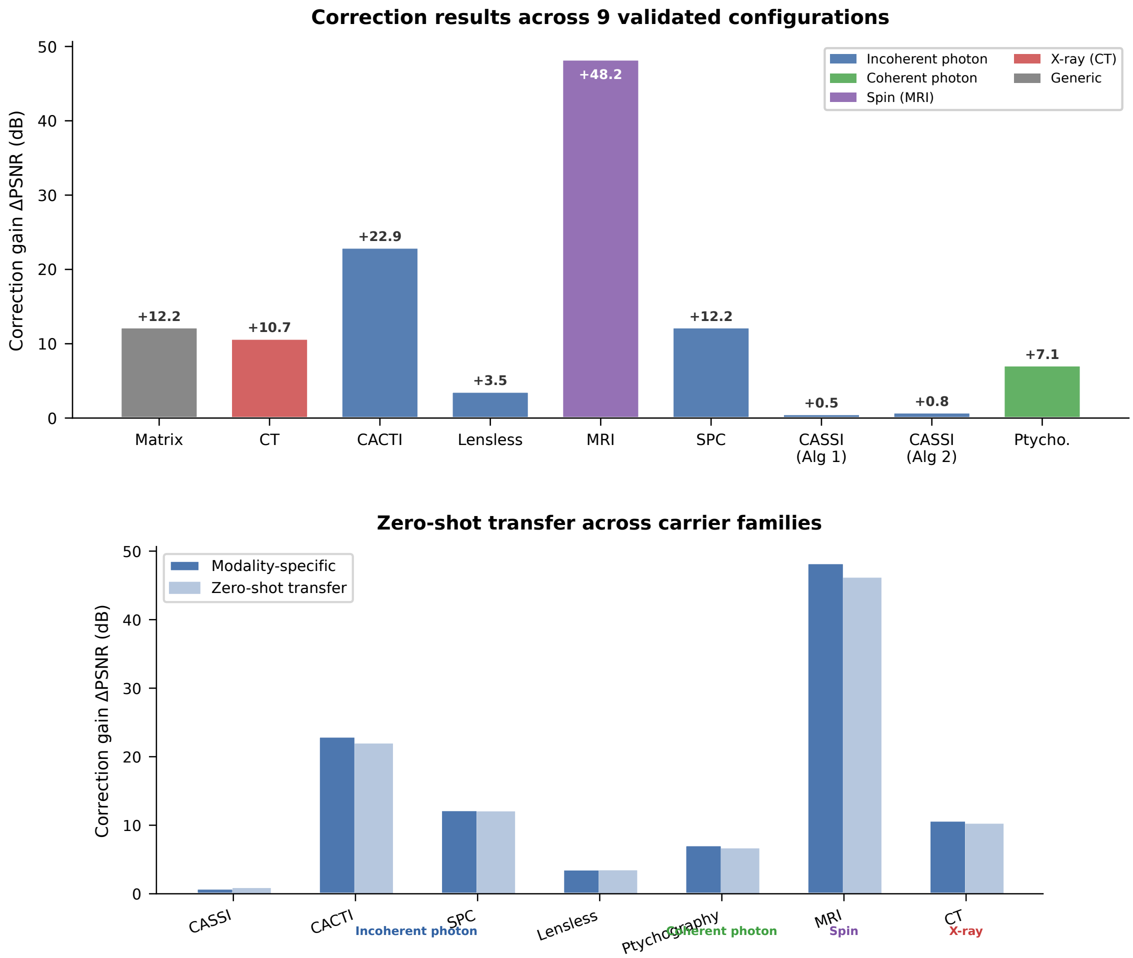

Figure 4: **Correction results and zero-shot generalization.** **a**, Oracle correction ceiling Δ_{oracle} (dB) across 9 validated configurations (7 distinct modalities), grouped by carrier family. **b**, Zero-shot generalization: hyperparameters tuned on photon-domain modalities transfer without modification to coherent-photon, spin, and X-ray domains. Dark bars: modality-specific tuning; light bars: zero-shot transfer.

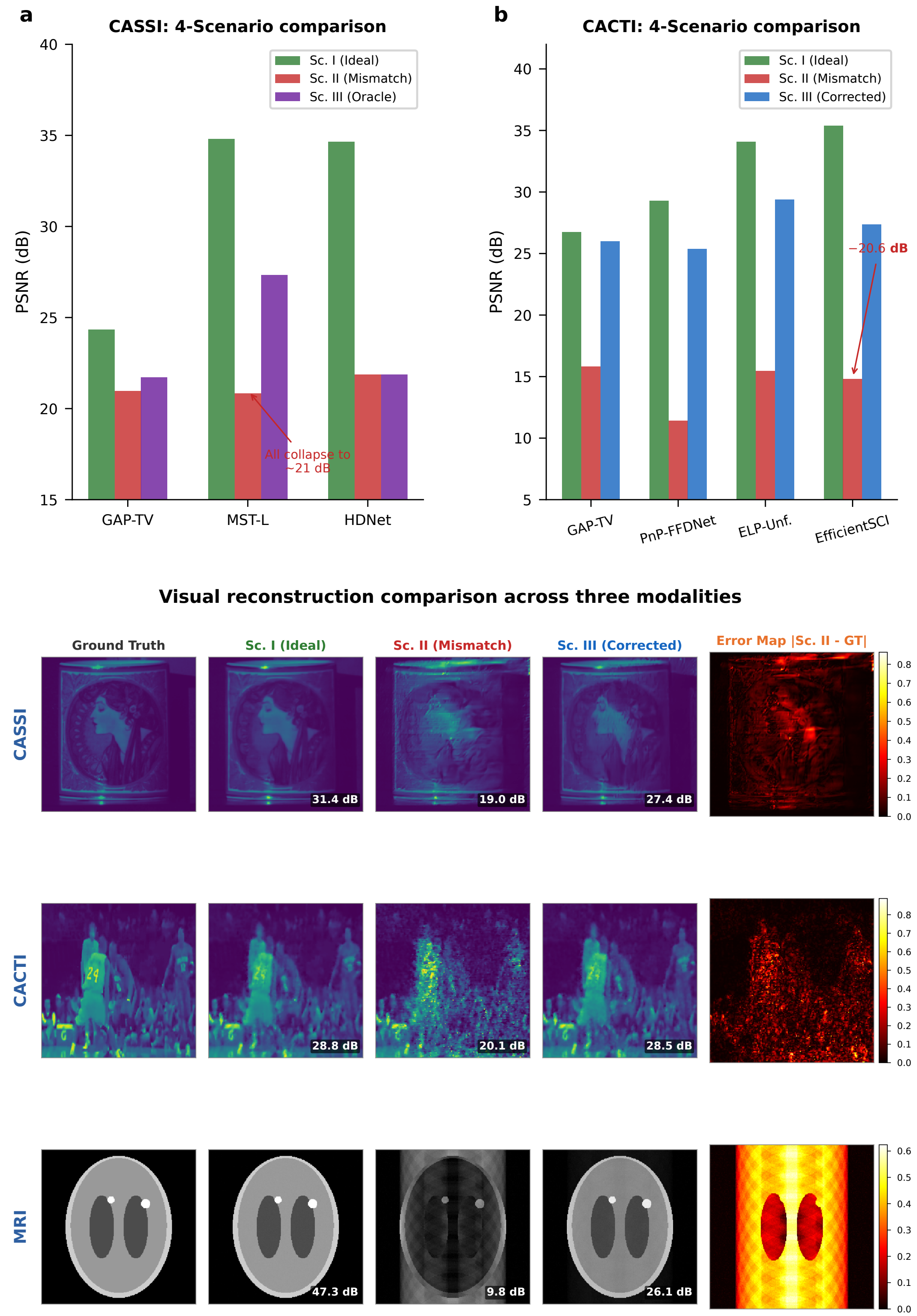

Figure 5: **Modality deep dives and visual comparison.** **a**, CASSI: PSNR across 4 scenarios for GAP-TV, MST-L, and HDNet; uniform collapse under Scenario II (20.83–21.88 dB) confirms operator-driven failure. **b**, CACTI: four methods across 4 scenarios, showing up to 20.58 dB mismatch degradation. **c–e**, Visual reconstruction comparison (CASSI, CACTI, MRI): ground truth, Scenario I, Scenario II (structured artifacts), Scenario III (PWM-corrected), and error map. Mismatch artifacts are qualitatively distinct

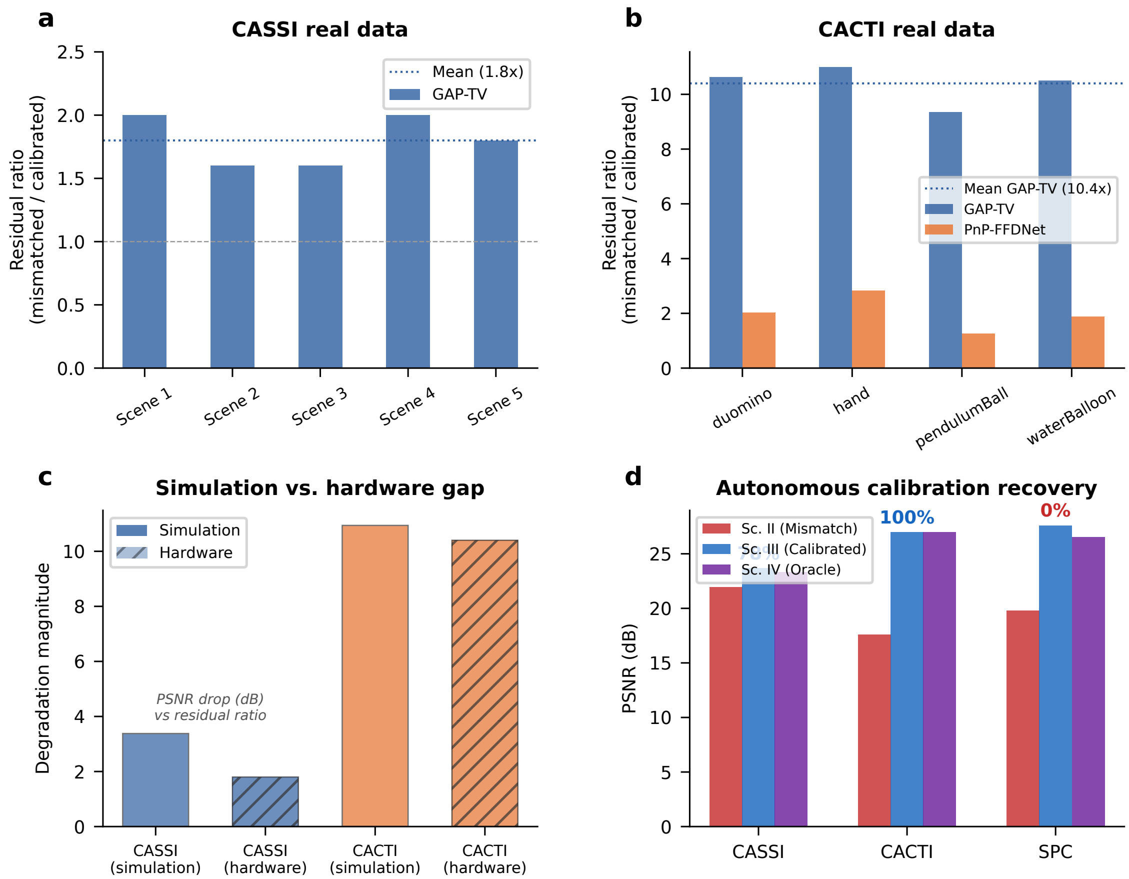

Figure 6: **Hardware validation on real CASSI and CACTI instruments.** **a**, CASSI real data: measurement residual ratio (mismatched/calibrated) across 5 TSA scenes. GAP-TV shows 1.8× mean ratio. **b**, CACTI real data: residual ratio across 4 scenes. GAP-TV shows 10.4× mean ratio; PnP-FFDNet shows 2.0×. **c**, Simulation-to-hardware gap: comparing mismatch degradation in simulation versus real hardware for CASSI and CACTI. **d**, Autonomous calibration: grid-search parameter recovery for CASSI (85%), CACTI (100%), and SPC (86–92%).

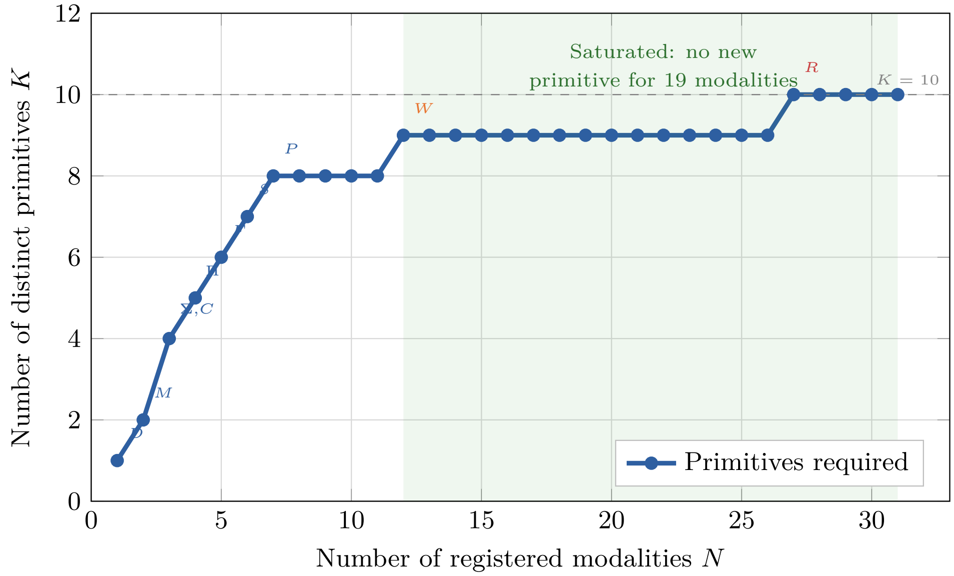

Figure 7: **Basis-growth saturation and primitive decomposition.** **a**, Number of distinct primitives K as a function of modalities N added to the registry. The curve saturates at $K = 10$ for $N \geq 27$; annotated points mark the introduction of each primitive. **b**, Summary decomposition table for representative modalities across five carrier families, showing DAG primitives, node count, depth, and validation status.

Extended Data Table 1 | Oracle correction ceiling (Scenario IV) across 9 validated configurations (7 modalities). Full table with per-scene metrics in Supplementary Table S1; autonomous correction recovery (Scenario III) in Supplementary Table S9.

Extended Data Table 2 | 26-modality template registry spanning five physical carrier families. Full table in Supplementary Table S3.