Supplementary Information:
Physics World Models for Computational Imaging

Chengshuai Yang and Xin Yuan

Contents

1 Supplementary Note 1: Triad Decomposition Mathematical Derivations

The Triad Decomposition formalizes three successive gates that every computational-imaging reconstruction must pass. We derive rigorous bounds for each gate and show how the *recovery ratio* ρ summarizes the overall reconstruction quality.

1.1 Gate 1 — Information-Theoretic Limit (Compression)

Let the forward operator $\mathbf{H} \in \mathbb{R}^{m \times n}$ map an n-dimensional scene $\mathbf{x}$ to m measurements $\mathbf{y}$. Define the *compression ratio* $\gamma = m/n$.

Theorem 1 (Compression bound)**.** *For any estimator* $\hat{\mathbf{x}}(\mathbf{y})$*, the minimum achievable mean-squared error satisfies*

$$\mathrm{MSE}_{\min} \;\geq\; \frac{1}{n} \sum_{i=1}^{n-m} \sigma_i^2(\mathbf{x}), \tag{S1}$$

where $\sigma_i^2(\mathbf{x})$ *denotes the variance of* $\mathbf{x}$ *along the* i*-th null-space direction of* $\mathbf{H}$*.*

Interpretation. The null space of $\mathbf{H}$ has dimension $\dim \mathcal{N}(\mathbf{H}) = n - \mathrm{rank}(\mathbf{H}) \geq n - m$. Information lost in the null space cannot be recovered without a prior; Gate 1 therefore sets an *information-theoretic floor* on PSNR:

$$\mathrm{PSNR}_{\max}^{(G1)} = 10 \log_{10}\!\left(\frac{\|\mathbf{x}\|_\infty^2}{\mathrm{MSE}_{\min}} \right). \tag{S2}$$

Modalities with higher compression ratios (e.g. CACTI with B frames compressed into one) face a proportionally lower ceiling.

1.2 Gate 2 — SNR-Dependent PSNR Bound (Noise)

Given additive noise $\mathbf{n} \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I})$, the measurement SNR is $\mathrm{SNR} = \|\mathbf{H}\mathbf{x}\|^2 / (m\, \sigma_n^2)$.

Theorem 2 (Noise bound)**.** *Under matched forward model and Gaussian noise, the best achievable PSNR is*

$$\mathrm{PSNR}_{\max}^{(G2)} = 10 \log_{10}(\mathrm{SNR}) + C_{\mathcal{M}}, \tag{S3}$$

where $C_{\mathcal{M}}$ *is a modality-dependent constant that absorbs the conditioning of* $\mathbf{H}$ *and the prior strength:*

$$C_{\mathcal{M}} = 10 \log_{10}\!\left(\frac{n\, \|\mathbf{x}\|_\infty^2}{\|\mathbf{x}\|^2} \cdot \kappa^{-2}(\mathbf{H}) \right), \tag{S4}$$

with $\kappa(\mathbf{H})$ *the condition number of* $\mathbf{H}$ *restricted to its column space.*

Interpretation. Gate 2 is a *noise ceiling*: no algorithm can exceed $\mathrm{PSNR}_{\max}^{(G2)}$ regardless of computational budget. Empirically, modalities such as MRI ($C_{\mathrm{MRI}} \approx 8\,\mathrm{dB}$) are more noise-tolerant than CASSI ($C_{\mathrm{CASSI}} \approx 3\,\mathrm{dB}$) due to favorable operator conditioning.

1.3 Gate 3 — Sensitivity to Model Mismatch (Calibration)

Let $\boldsymbol{\theta} \in \mathbb{R}^p$ collect the calibration parameters (e.g. mask shift, PSF blur, coil sensitivity) and let $\mathbf{H}_{\boldsymbol{\theta}}$ denote the parameterized forward operator.

Theorem 3 (Calibration sensitivity)**.** *For small perturbation $\delta\boldsymbol{\theta}$ around the true parameter $\boldsymbol{\theta}^*$, the PSNR degradation is*

$$\Delta\text{PSNR} \approx -\frac{10}{\ln 10}\,\frac{\delta\boldsymbol{\theta}^\top \mathbf{J}^\top \mathbf{J}\,\delta\boldsymbol{\theta}}{\text{MSE}_0}, \tag{S5}$$

where $\mathbf{J} = \partial(\mathbf{H}_{\boldsymbol{\theta}}\mathbf{x})/\partial\boldsymbol{\theta}\big|_{\boldsymbol{\theta}^*}$ is the parameter Jacobian and MSE_0 is the noise-only MSE at $\boldsymbol{\theta}^*$.

Per-parameter sensitivity. Restricting to a single parameter θ_j:

$$\frac{d\,\text{PSNR}}{d\,\theta_j} = -\frac{10}{\ln 10}\,\frac{\|\mathbf{J}_{:,j}\|^2}{\text{MSE}_0}\,\delta\theta_j + \mathcal{O}(\delta\theta_j^2). \tag{S6}$$

The first-order Taylor expansion shows that PSNR degrades linearly in $|\delta\theta_j|$ for small mismatches and quadratically for larger ones.

1.4 Recovery Ratio

We define the *recovery ratio* as

$$\rho = \frac{\text{PSNR}_{\text{achieved}} - \text{PSNR}_{\text{mismatch}}}{\text{PSNR}_{\text{ideal}} - \text{PSNR}_{\text{mismatch}}}, \qquad 0 \le \rho \le 1. \tag{S7}$$

In rare cases where Scenario III exceeds Scenario I (e.g., due to regularization benefits from the corrected operator), ρ may exceed 1.

Proposition 4. *Under convex reconstruction loss with matched regularization and convex constraint on $\boldsymbol{\theta}$, the recovery ratio satisfies $0 \le \rho \le 1$ for any estimator.*

Proof sketch. The oracle PSNR is the global optimum of the jointly convex problem in $(\mathbf{x}, \boldsymbol{\theta})$. The mismatch PSNR uses a fixed $\boldsymbol{\theta}_0 \neq \boldsymbol{\theta}^*$, yielding a suboptimal feasible point. Any correction step moves towards the optimum, so $\text{PSNR}_{\text{mismatch}} \le \text{PSNR}_{\text{achieved}} \le \text{PSNR}_{\text{oracle}}$, giving $\rho \in [0, 1]$. □

Under the conditions of Proposition 1, $0 \le \rho \le 1$. In practice, beneficial regularization bias from the corrected operator can yield $\rho > 1$, as observed for CACTI.

1.5 Conditions for Gate 3 Dominance

The empirical finding that Gate 3 dominates across all validated modalities admits a theoretical justification.

Proposition 5 (Gate 3 dominance condition)**.** *For any instrument operating above its Gate 1 floor ($\gamma > \gamma_{\min}$, where $\gamma_{\min}$ is the minimum compression ratio for which $\text{PSNR}^{(G1)}_{\max}$ exceeds the target quality) and Gate 2 floor ($\text{SNR} > \text{SNR}_{\min}$, where $\text{SNR}_{\min}$ is the minimum carrier budget for which $\text{PSNR}^{(G2)}_{\max}$ exceeds the target quality), Gate 3 becomes the binding constraint whenever the calibration error exceeds the noise-equivalent resolution:*

$$\|\delta\boldsymbol{\theta}\|_{\mathbf{J}^\top\mathbf{J}} > \frac{\sigma_n}{\|\mathbf{x}\|_\infty} \cdot \kappa(\mathbf{H}), \tag{S8}$$

where $\|\delta\boldsymbol{\theta}\|_{\mathbf{J}^\top\mathbf{J}} = \sqrt{\delta\boldsymbol{\theta}^\top\mathbf{J}^\top\mathbf{J}\,\delta\boldsymbol{\theta}}$ *is the parameter-space mismatch norm weighted by the sensitivity Jacobian,* σ_n *is the noise standard deviation, and* $\kappa(\mathbf{H})$ *is the condition number of the forward operator.*

Proof sketch. From Theorem 3 (Eq. S5), the mismatch-induced PSNR degradation is $\Delta\text{PSNR}_{\text{mismatch}} \approx (10/\ln 10)\cdot\|\delta\boldsymbol{\theta}\|^2_{\mathbf{J}^\top\mathbf{J}}/\text{MSE}_0$. From Theorem 2 (Eq. S3), the noise-induced PSNR ceiling is $\text{PSNR}^{(G2)}_{\max} = 10\log_{10}(\text{SNR}) + C_{\mathcal{M}}$. Gate 3 dominates when $\Delta\text{PSNR}_{\text{mismatch}} > \Delta\text{PSNR}_{\text{noise}}$, which after substitution yields condition (S8). The condition is satisfied in modern instruments because: (i) well-designed systems operate with $\gamma \gg \gamma_{\min}$ and $\text{SNR} \gg \text{SNR}_{\min}$, pushing Gates 1 and 2 well above their floors; and (ii) the sensitivity Jacobian $\mathbf{J}$ amplifies even small parameter errors ($\|\delta\boldsymbol{\theta}\| \ll 1$) into substantial measurement discrepancies, because the forward operators of modern high-resolution instruments are highly sensitive to calibration parameters. □

Interpretation. Proposition 2 formalizes the intuition that as instruments improve in information capacity (Gate 1) and signal quality (Gate 2), the binding constraint shifts to calibration fidelity (Gate 3). This is precisely the regime occupied by modern computational imaging systems, explaining the universal Gate 3 dominance observed empirically across all seven validated modalities.

2 Supplementary Note 2: Complete OperatorGraph Specification

The `OperatorGraph` is the central abstraction of the PWM framework. It represents a computational imaging pipeline as a directed acyclic graph (DAG) of differentiable operators.

2.1 Node Interface

Every node v in the graph implements the following interface:

Method	Contract
`forward(x)` $\rightarrow$ `y`	Apply the physical forward model. Input shape: `(B, C, *spatial)`. Output shape determined by the operator (e.g. compression changes spatial dims).
`adjoint(y)` $\rightarrow$ `x`	Apply the adjoint (transpose) operator. Must satisfy $\langle \mathbf{Hx}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{H}^\top \mathbf{y} \rangle$ to numerical precision ($< 10^{-6}$ relative error).
`shape_in` / `shape_out`	Static shape annotations for compile-time validation.
`dtype`	Data type contract: `float32` (default) or `complex64`.
`parameters()` $\rightarrow$ `dict`	Returns learnable calibration parameters with their current values.

2.2 Edge Semantics

An edge $e = (u, v)$ represents data flow: the output tensor of node u is fed as input to node v. Edges carry the following metadata:

- **Tensor shape**: inferred at compile time via shape propagation.
- **Data type**: must match between source output and target input.
- **Optional transform**: lightweight reshape or permute operations (e.g. channel stacking for multi-frame CACTI data).

2.3 Compilation Algorithm

Given an `OperatorGraph` $G = (V, E)$:

1. **Topological sort.** Compute a valid execution order $v_1, v_2, \ldots, v_{|V|}$ using Kahn's algorithm. Raise `CycleError` if the graph contains a cycle.
2. **Shape propagation.** Starting from the source node(s), propagate tensor shapes through each operator's `shape_out` to validate all edge contracts.
3. **Automatic adjoint chain.** Construct the adjoint graph $G^\top$ by reversing all edges and replacing each node's `forward` with `adjoint`. The adjoint graph is used for gradient-based reconstruction and the Triad Decomposition sensitivity analysis.

4. **Fusion (optional).** Consecutive linear operators are fused into a single matrix–vector product when dimensions permit, reducing memory traffic.

2.4 JSON Serialization Schema

The graph is serialized to a JSON document for reproducibility and sharing:


```json
{
  "version": "0.3.0",
  "nodes": [
    {
      "id": "mask",
      "type": "CassiMask",
      "params": {"shift_r": 0.0, "shift_c": 0.0},
      "shape_in": [1, 28, 256, 256],
      "shape_out": [1, 1, 256, 310]
    }
  ],
  "edges": [
    {"src": "mask", "dst": "detector", "dtype": "float32"}
  ],
  "metadata": {
    "modality": "CASSI",
    "created": "2026-02-01T00:00:00Z"
  }
}
```

3 Supplementary Table S1: All 16 Correction Configurations

Table S1 reports the correction results for all 16 correction configurations across the validated modalities. The first 9 modalities have full end-to-end validation; the remaining 7 are designated Phase 2/4 additions with partial or planned validation.

Table S1: Oracle correction ceiling summary. ΔPSNR is the improvement from Scenario II (mismatch) to Scenario IV (oracle: true operator applied to mismatched data). For modalities with low-dimensional mismatch (Matrix, CT, Lensless, MRI, Ptychography), autonomous correction achieves near-oracle performance (Sc. III ≈ Sc. IV); for CASSI and CACTI with multi-parameter mismatch, the gap between Sc. III and Sc. IV is reported in Table S9. CASSI correction results use GAP-TV as the reconstruction solver. Scenario I (ideal) PSNR values are available in the RunBundle manifests in the code repository.

Modality	**Mismatch Parameter**	**Sc. II (dB)**	**Sc. IV (dB)**	**ΔPSNR (dB)**	**RMSE**
Matrix[1]	gain_bias	11.14	23.35	+12.21	0.0042
CT	center_of_rotation	13.41	24.09	+10.68	0.0031
CACTI[2]	8-param mismatch	15.81	26.01	+10.21	—
Lensless	psf_shift	23.48	27.03	+3.55	0.0089
MRI	coil_sensitivities	6.94	55.19	+48.25	0.0003
SPC[3]	gain_drift	18.51	26.21	+7.71	—
CASSI (Alg 1)[4]	mask_geo + dispersion	20.96	21.50	+0.54	0.0105
CASSI (Alg 2)[5]	mask_geo + dispersion	20.96	21.72	+0.76	0.0098
Ptychography	position_offset	17.35	24.44	+7.09	0.0063
Phase 2/4 additions — validation in progress					
Holography	propagation_distance	—	—	—	—

Continued on next page

[1]Matrix (Generic) is a canonical test configuration that uses the same gain-bias operator template as SPC; identical results confirm pipeline consistency rather than representing an independent modality.

[2]CACTI uses GAP-TV as the reconstruction solver for Table S1 consistency. Oracle values from InverseNet validation under 8-parameter mismatch (dx=0.5 px, dy=0.3 px, θ=0.1°, Δt=0.05, η=0.95, g=1.02, o=0.002, σ_n=1.0). State-of-the-art EfficientSCI achieves Sc. II = 14.81 dB and Sc. IV = 27.38 dB; see main text.

[3]SPC uses FISTA-TV as the reconstruction solver for Table S1 consistency. Values from InverseNet validation under exponential gain drift (α=0.0015). HATNet achieves Sc. II = 19.40 dB and Sc. IV = 29.78 dB; see main text.

[4]Reported values are averaged over 10 test scenes using GAP-TV as the reconstruction solver. Sc. II from InverseNet validation; Sc. IV values are oracle upper bounds (true operator applied to mismatched data). Autonomous correction achieves 85% of oracle (Table S9).

[5]The correction-table PSNR reflects GAP-TV reconstruction; state-of-the-art solvers (e.g., MST-L) achieve 27.33 dB under Sc. IV (see main text and Table S2).

Modality	**Mismatch Parameter**	**Sc. II (dB)**	**Sc. IV (dB)**	**ΔPSNR (dB)**	**RMSE**
FPM	LED_positions	—	—	—	—
Phase Retrieval	support_constraint	—	—	—	—
OCT	dispersion_coeffs	—	—	—	—
Fluorescence	PSF_aberration	—	—	—	—
Ultrasound	speed_of_sound	—	—	—	—
Radar/SAR	motion_compensation	—	—	—	—

4 Supplementary Table S2: CASSI Per-Scene Results

Table S2 reports PSNR (dB) and recovery ratio ρ (%) for 10 KAIST scenes across four methods under three scenarios, matching the InverseNet benchmark[1].

Table S2: CASSI per-scene PSNR (dB) and oracle recovery ratio ρ_{IV}. Sc. I = ideal forward model; Sc. II = mismatched (baseline); Sc. IV = oracle correction (true operator applied to mismatched data). All values from InverseNet validation under the 5-parameter mismatch (dx=0.5 px, dy=0.3 px, θ=0.1°, a_1=2.02, α=0.15°). ρ_{IV} = (Sc. IV − Sc. II)/(Sc. I − Sc. II) is the oracle recovery ceiling. HDNet shows ρ_{IV}=0% across all scenes due to its mask-oblivious architecture.

Scene	**Method**	**Sc. I**	**Sc. II**	**Sc. IV**	ρ_{IV}
scene01	GAP-TV	26.49	24.08	24.18	4.1%
scene01	PnP-HSICNN	27.43	23.47	25.78	58.3%
scene01	HDNet	34.95	24.37	24.37	0.0%
scene01	MST-L	35.29	23.96	29.98	53.2%
scene02	GAP-TV	24.60	21.89	22.82	34.5%
scene02	PnP-HSICNN	25.34	21.56	23.80	59.3%
scene02	HDNet	35.65	23.26	23.26	0.0%
scene02	MST-L	36.14	22.21	28.42	44.6%
scene03	GAP-TV	25.96	18.62	19.68	14.4%
scene03	PnP-HSICNN	26.71	17.59	21.83	46.5%
scene03	HDNet	35.54	18.61	18.61	0.0%
scene03	MST-L	35.66	16.09	23.57	38.2%
scene04	GAP-TV	28.36	23.37	24.41	20.9%
scene04	PnP-HSICNN	28.80	22.89	25.98	52.3%
scene04	HDNet	41.63	23.98	23.98	0.0%
scene04	MST-L	40.05	21.91	29.80	43.5%
scene05	GAP-TV	23.66	20.39	21.27	26.9%
scene05	PnP-HSICNN	24.65	19.36	22.75	64.3%
scene05	HDNet	32.56	20.22	20.22	0.0%
scene05	MST-L	32.84	20.28	26.75	51.5%
scene06	GAP-TV	22.34	20.39	21.00	31.2%
scene06	PnP-HSICNN	23.12	20.29	21.80	53.0%
scene06	HDNet	34.33	22.63	22.63	0.0%
scene06	MST-L	34.56	22.37	28.67	51.7%
scene07	GAP-TV	23.51	20.56	21.07	17.4%
scene07	PnP-HSICNN	24.94	19.86	22.77	57.4%
scene07	HDNet	33.27	20.79	20.79	0.0%
scene07	MST-L	33.80	19.76	26.12	45.3%
scene08	GAP-TV	22.16	20.57	21.10	33.3%
scene08	PnP-HSICNN	22.73	20.30	21.92	66.4%
scene08	HDNet	32.26	22.73	22.73	0.0%
scene08	MST-L	32.74	21.33	27.44	53.5%
scene09	GAP-TV	23.03	19.14	20.61	37.8%
scene09	PnP-HSICNN	24.03	18.79	21.82	57.8%
scene09	HDNet	34.18	21.18	21.18	0.0%
scene09	MST-L	34.37	19.75	26.71	47.6%
scene10	GAP-TV	23.27	20.58	21.04	16.9%
scene10	PnP-HSICNN	23.47	19.90	22.38	69.4%
scene10	HDNet	32.22	21.06	21.06	0.0%

Scene	**Method**	**Sc. I**	**Sc. II**	**Sc. IV**	$\rho_{\mathbf{IV}}$
scene10	MST-L	32.63	20.69	25.86	43.3%
Mean	**GAP-TV**	**24.34**	**20.96**	**21.72**	**22.5%**
Mean	**PnP-HSICNN**	**25.12**	**20.40**	**23.08**	**56.8%**
Mean	**HDNet**	**34.66**	**21.88**	**21.88**	**0.0%**
Mean	**MST-L**	**34.81**	**20.83**	**27.33**	**46.5%**

5 Supplementary Table S3: 26-Modality Template Registry

Table S3 lists all 26 modalities in the PWM benchmark, grouped by physical carrier tier.

Table S3: 26-modality template registry. PSNR values are representative (Scenario I, default solver). "Ref. PSNR" is the best published result where available. Validation tiers: *Validated* = full Scenarios I–IV correction (7 modalities); *Test* = pipeline consistency test (not an independent modality); *Phase 2* = template validated, correction in progress; *Phase 4* = template registered, validation planned.

#	Modality	Domain	PSNR (dB)	Ref. PSNR (dB)	Status	Solver
1	Matrix (toy)	Incoherent	23.35	—	Test	ADMM-TV
2	SPC	Incoherent	28.06	31.0	Validated	FISTA-TV
3	CASSI	Incoherent	24.34	34.2	Validated	GAP-TV
4	CACTI	Incoherent	35.39	33.4	Validated	EfficientSCI
5	Lensless	Incoherent	27.03	28.5	Validated	ADMM-TV
6	Fluorescence	Incoherent	—	30.1	Phase 4	—
7	Holography	Coherent	—	32.7	Phase 2	—
8	Ptychography	Coherent	24.44	26.8	Validated	ePIE
9	FPM	Coherent	—	31.5	Phase 2	—
10	Phase Retrieval	Coherent	—	29.3	Phase 2	—
11	CDI	Coherent	—	27.4	Phase 4	—
12	OCT	Coherent	—	35.2	Phase 4	—
13	CT (parallel)	X-ray	24.09	42.1	Validated	FBP + TV
14	CT (fan-beam)	X-ray	—	40.8	Phase 2	—
15	CT (cone-beam)	X-ray	—	39.5	Phase 2	—
16	Electron Pty.	Electron	—	25.9	Phase 4	—
17	MRI (Cartesian)[6]	RF	52.11	42.3	Validated	CG-SENSE
18	MRI (radial)	RF	—	38.7	Phase 2	—
19	MRI (spiral)	RF	—	37.2	Phase 4	—
20	Ultrasound	Acoustic	—	31.8	Phase 4	—
21	Radar / SAR	RF	—	28.4	Phase 4	—
22	NeRF	Multi-view	—	31.7	Phase 2	—
23	3DGS	Multi-view	—	33.2	Phase 2	—
24	Light Field	Multi-view	—	36.1	Phase 4	—
25	Structured Illum.	Incoherent	—	32.4	Phase 4	—
26	Ghost Imaging	Incoherent	—	22.8	Phase 4	—

[6] Multi-coil Scenario I PSNR (8 coils, 4× acceleration); see Supplementary Note 11 for full multi-coil analysis.

6 Supplementary Table S4: YAML Registry Summary

Table S4 summarizes the nine YAML registry files (totalling 7,034 lines) that define the PWM framework's modular architecture.

Table S4: YAML registry files in `packages/pwm_core/contrib/`.

#	File	Lines	Purpose
1	`modalities.yaml`	1200	Defines 64 modality entries with keywords, forward model equations, and default solvers.
2	`graph_templates.yaml`	1400	OperatorGraph skeletons for 89 templates across 64 modalities.
3	`photon_db.yaml`	624	Photon models parameterized by source power, quantum efficiency, exposure, and detector.
4	`mismatch_db.yaml`	797	Mismatch parameters, perturbation ranges, and correction methods per modality.
5	`compression_db.yaml`	1186	Recoverability tables mapping compression ratio to expected PSNR with provenance.
6	`solver_registry.yaml`	650	Maps solver names to implementations (ADMM[2], FISTA[3], GAP-TV, PnP, learned).
7	`primitives.yaml`	450	Primitive operator metadata (type, adjoint availability, differentiability).
8	`dataset_registry.yaml`	380	Links modalities to benchmark datasets (KAIST, fastMRI, etc.).
9	`acceptance_thresholds.yaml`	347	Pass/fail thresholds per metric for automated validation.

7 Supplementary Note 4: RunBundle Schema

The `RunBundle` is PWM's reproducibility artifact. Every experiment produces a RunBundle that captures the complete computational state.

7.1 RunBundle v0.3.0 Specification

Table S5: RunBundle v0.3.0 field specification.

Field	Type	Description
`version`	string	Schema version, currently `"0.3.0"`.
`git_hash`	string	Full 40-character SHA-1 of the commit used to generate results.
`rng_seeds`	object	Random seeds for Python (`random`, `numpy`), PyTorch (CPU, CUDA), and any solver-specific RNG.
`platform_info`	object	OS, Python version, GPU model, CUDA version, PyTorch version, and `pip freeze` hash.
`sha256_hashes`	object	SHA-256 checksums of all input data files (measurements, masks, ground truth).
`metrics`	object	Computed quality metrics: PSNR, SSIM, LPIPS, and any modality-specific metrics.
`timestamps`	object	ISO-8601 timestamps for start, end, and per-stage completion.
`operator_state`	object	Serialized OperatorGraph including all node parameters and edge metadata.
`triad_report`	object	Triad Decomposition evaluation: gate pass/fail status, PSNR bounds, and recovery ratio ρ.
`correction_trajectory`	array	Time series of correction iterations: [{`iter`, `params`, `loss`, `psnr`, `grad_norm`}, . . .]. Enables convergence analysis and early-stopping diagnostics.
`solver_config`	object	Complete solver hyperparameters (algorithm, iterations, step size, regularization weight, denoiser checkpoint).
`modality`	string	Modality identifier matching `modalities.yaml`.
`scenario`	string	One of `I` (ideal), `II` (mismatch), `III` (corrected), `IV` (oracle_mask).
`notes`	string	Free-form text field for experiment annotations.

7.2 Integrity Verification

A RunBundle can be verified by:

1. Checking `git_hash` matches the repository state.
2. Re-computing `sha256_hashes` from the data files.
3. Re-running the solver with the stored `rng_seeds` and `solver_config` and comparing metrics within tolerance (± 0.01 dB PSNR).

8 Supplementary Note 5: Computational Cost Analysis

8.1 Runtime per Modality

All timings were measured on a single NVIDIA RTX 4090 GPU (24 GB VRAM) with PyTorch 2.1 and CUDA 12.1.

Table S6: Computational cost for correction (Scenario II → III). RoIC = dB recovered per GPU-hour.

Modality	**Iterations**	**Runtime (s)**	**ΔPSNR (dB)**	**RoIC (dB/GPU-hr)**
Matrix	200	12	+12.21	3663.0
CT	500	85	+10.68	452.3
CACTI	1000	180	+10.21	204.2
Lensless	300	45	+3.55	284.0
MRI	150	30	+48.25	5790.0
SPC	200	15	+7.71	1850.4
CASSI (Alg 1)	500	300	+0.54	6.5
CASSI (Alg 2)	2000	3200	+0.76	0.9
Ptychography	800	420	+7.09	60.8

8.2 RoIC Metric: Efficiency of Correction

We define the *Return on Invested Compute* (RoIC) as:

$$\mathrm{RoIC} = \frac{\Delta\mathrm{PSNR}\ (\mathrm{dB})}{T_{\mathrm{GPU}}\ (\mathrm{hours})}, \tag{S9}$$

where T_{GPU} is the wall-clock GPU time.
Key observations:

- **MRI** achieves the highest RoIC (5790 dB/GPU-hr) because coil sensitivity correction yields massive PSNR gains with a well-conditioned linear operator that converges rapidly.
- **Matrix/SPC** also show high RoIC due to the simplicity of the gain–bias correction model.
- **CASSI Algorithm 2** has the lowest RoIC (0.9 dB/GPU-hr) because it jointly optimizes mask geometry and dispersion over 2000 iterations, each requiring a full forward–adjoint cycle through the dispersive operator. Under the multi-parameter InverseNet mismatch, the absolute correction gains are modest (+0.76 dB), reflecting the inherent difficulty of simultaneously correcting five coupled mismatch parameters.
- The practical implication is that Algorithm 1 (fixed dispersion, correct mask geometry only) is preferred when compute budget is limited, recovering 71% of Algorithm 2's ΔPSNR at <10% of the cost.

8.3 Scaling Considerations

For large-scale deployment:

- Correction is *embarrassingly parallel* across scenes/slices — multi-GPU scaling is near-linear.

- The OperatorGraph compilation overhead is amortized: once compiled, the same graph serves all scenes of a given modality.
- Memory footprint scales with the measurement dimension m, not the scene dimension n, making correction feasible even for high-resolution 3D volumes (e.g. 512^3 CT).

9 Supplementary Note 6: Real-Data Validation Details

9.1 CASSI Real Data: TSA Hyperspectral Camera

The TSA (Thermal Snapshot Aperture) real dataset consists of 5 hyperspectral scenes acquired with a coded aperture snapshot spectral imaging system. Each scene has spatial resolution 660×660 with 28 spectral bands (450–650 nm) and a mask-shift step of 2 pixels per band. The extended measurement has width $660 + (28 - 1) \times 2 = 714$ pixels.

Experimental protocol. For each scene, we reconstruct with four methods: GAP-TV (classical iterative), HDNet (mask-oblivious neural network), MST-S, and MST-L (mask-guided spectral transformers). Each method is applied under two conditions: (i) the calibrated mask provided with the dataset, and (ii) a perturbed mask with sub-pixel shift $dx = 0.5\,\text{px}$, $dy = 0.3\,\text{px}$. Because the MST model has a hardcoded 256×256 spatial assumption, the 660×660 real data is center-cropped to 256×256 for MST reconstruction. HDNet is fully convolutional and processes the full spatial extent.

Metric. Because no ground-truth hyperspectral cube exists for the real scenes, we use the measurement residual as a ground-truth-free diagnostic:

$$r = \frac{\|\mathbf{y} - H\hat{\mathbf{x}}\|^2}{\|\mathbf{y}\|^2}, \tag{S10}$$

where $\mathbf{y}$ is the measured snapshot and $\hat{\mathbf{x}}$ is the reconstruction. The residual ratio $r_{\text{mismatch}}/r_{\text{calibrated}}$ quantifies the relative impact of operator mismatch, independent of the (unknown) ground truth.
Key findings. GAP-TV, which explicitly conditions on the coded aperture mask in its forward–adjoint iterations, shows the expected sensitivity to mask mismatch (mean ratio 1.8×). HDNet, whose architecture processes the measurement without explicit mask conditioning, is entirely insensitive (ratio 1.0×). The transformer methods (MST-S, MST-L) show ratios near or below 1.0× on real data—in sharp contrast to their severe degradation in simulation (−13.98 dB for MST-L). This simulation-to-hardware gap indicates that the real hardware mask already contains uncorrected manufacturing imperfections, so the additional 0.5 px perturbation is small relative to pre-existing errors.

9.2 CACTI Real Data: Temporal Compressive Camera

The CACTI real dataset consists of 4 dynamic scenes (duomino, hand, pendulumBall, waterBalloon) acquired with a coded aperture compressive temporal imaging system at 512×512 spatial resolution with compression ratio 10 (10 video frames encoded per snapshot).
Key findings. GAP-TV shows an order-of-magnitude residual increase (10.4× mean) under subpixel mask mismatch, consistent with the multiplicative error amplification in temporal compression (a single mask error propagates across all 10 compressed frames). PnP-FFDNet shows moderate robustness (2.0×), likely because the FFDNet denoiser provides an implicit regularization that partially absorbs the mismatch artefacts. The CACTI results confirm that temporal compressive modalities are substantially more sensitive to operator mismatch than spectral compressive modalities (CASSI), a distinction that would be invisible without the unified Triad Decomposition framework.

Table S7: CASSI real-data measurement residuals. Residual values r (Eq. S10) and residual ratio (mismatched/calibrated) across 5 TSA scenes and 4 methods.

Scene	**Method**	r_{cal}	r_{mis}	**Ratio**
Scene 1	GAP-TV	0.00148	0.00298	2.0
Scene 1	HDNet	0.18506	0.18506	1.0
Scene 1	MST-S	0.09953	0.10746	1.1
Scene 1	MST-L	0.10055	0.09544	0.9
Scene 2	GAP-TV	0.00210	0.00331	1.6
Scene 2	HDNet	0.14853	0.14853	1.0
Scene 2	MST-S	0.25357	0.23807	0.9
Scene 2	MST-L	0.25441	0.22811	0.9
Scene 3	GAP-TV	0.00204	0.00320	1.6
Scene 3	HDNet	0.18291	0.18291	1.0
Scene 3	MST-S	0.09399	0.09160	1.0
Scene 3	MST-L	0.10127	0.09085	0.9
Scene 4	GAP-TV	0.00148	0.00297	2.0
Scene 4	HDNet	0.14355	0.14355	1.0
Scene 4	MST-S	0.15586	0.14712	0.9
Scene 4	MST-L	0.17899	0.15893	0.9
Scene 5	GAP-TV	0.00233	0.00421	1.8
Scene 5	HDNet	0.11325	0.11325	1.0
Scene 5	MST-S	0.11995	0.12004	1.0
Scene 5	MST-L	0.13008	0.12208	0.9
Mean	**GAP-TV**	**0.00189**	**0.00333**	**1.8**
Mean	**HDNet**	**0.15466**	**0.15466**	**1.0**
Mean	**MST-S**	**0.14458**	**0.14086**	**1.0**
Mean	**MST-L**	**0.15306**	**0.13908**	**0.9**

Table S8: CACTI real-data measurement residuals and total variation. Residual ratio (mismatched/calibrated) and total variation (TV) for 4 real scenes and 2 methods.

Scene	**Method**	r_{cal}	r_{mis}	**Ratio**
duomino	GAP-TV	8.0×10^{-6}	8.5×10^{-5}	10.6
duomino	PnP-FFDNet	0.00198	0.00403	2.0
hand	GAP-TV	7.0×10^{-6}	7.7×10^{-5}	11.0
hand	PnP-FFDNet	0.00249	0.00705	2.8
pendulumBall	GAP-TV	3.7×10^{-5}	3.46×10^{-4}	9.4
pendulumBall	PnP-FFDNet	0.00919	0.01150	1.3
waterBalloon	GAP-TV	1.4×10^{-5}	1.47×10^{-4}	10.5
waterBalloon	PnP-FFDNet	0.00264	0.00493	1.9
Mean	**GAP-TV**	—	—	**10.4**
Mean	**PnP-FFDNet**	—	—	**2.0**

Table S9: Grid-search calibration results on simulated data with known ground truth. PSNR values in dB. Sc. II = mismatch (no calibration); Sc. III = PWM grid-search corrected; Sc. IV = oracle (true operator on mismatched data). Recovery = (Sc. III − Sc. II) / (Sc. IV − Sc. II). CASSI/CACTI use measurement-residual objective; SPC uses reconstruction-TV objective. Note: Sc. II values here differ from Table S1 because Table S1 reports the full InverseNet multi-parameter mismatch, whereas this table uses simplified single-parameter mismatch for grid-search tractability.

Modality	**Method**	**Sc. II**	**Sc. III**	**Sc. IV**	**Time (s)**	**Recovery**
CASSI	GAP-TV	21.52	22.96	23.21	1140	85%
CACTI	GAP-TV	17.60	26.99	26.99	60	100%
SPC	FISTA-TV	19.78	26.54	27.60	166	86%
SPC	PnP-DRUNet	18.34	25.39	26.01	247	92%

9.3 Autonomous Calibration on Real Data

CACTI achieves perfect calibration recovery because the mismatch manifold is low-dimensional (2D spatial shift) and the temporal compression provides high sensitivity to mask position. CASSI achieves 85% recovery; the residual gap reflects the higher-dimensional mismatch space (mask geometry plus dispersion) and the estimation error in both dx and dy ($0.10\,\mathrm{px}$ each). For SPC, the measurement residual is uninformative for gain drift because the underdetermined system (m=272 measurements for n=1089 unknowns) always achieves near-zero self-consistent residual regardless of gain. Instead, reconstruction total variation provides a viable objective: the correctly calibrated gain produces clean measurements yielding smooth reconstructions with low TV, while incorrect gain leaves systematic artifacts that increase TV. The TV surface exhibits a clear bowl-shaped minimum near the true α, recovering 86% (FISTA-TV) to 92% (PnP-DRUNet) of the oracle bound. This demonstrates that blind calibration generalises to radiometric mismatch provided the objective matches the mismatch type: measurement residual for geometric mismatch, reconstruction sparsity for radiometric mismatch.

10 Supplementary Table S10: SSIM Comparison Across Modalities

Table S10 reports the structural similarity index (SSIM) for all three validated photon-domain modalities under the three primary scenarios. SSIM complements PSNR by capturing perceptual quality, particularly structural distortions from operator mismatch.

Table S10: SSIM comparison across modalities and scenarios. Sc. IV = oracle (true operator on mismatched data). Values are mean ± s.d. across test scenes/images.

Modality	Method	Sc. I	Sc. II	Sc. IV
CASSI (10 KAIST scenes, 28 spectral bands)				
	GAP-TV	0.723 ± 0.088	0.612 ± 0.084	0.688 ± 0.083
	HDNet	0.970 ± 0.009	0.756 ± 0.074	0.756 ± 0.074
	MST-L	0.973 ± 0.009	0.744 ± 0.069	0.881 ± 0.035
CACTI (6 benchmark videos, 8 temporal frames)				
	GAP-TV	0.848 ± 0.083	0.305 ± 0.070	0.794 ± 0.063
	PnP-FFDNet	0.890 ± 0.060	0.216 ± 0.076	0.820 ± 0.054
	ELP-Unfolding	0.965 ± 0.013	0.308 ± 0.076	0.927 ± 0.017
	EfficientSCI	0.973 ± 0.012	0.303 ± 0.079	0.927 ± 0.017
SPC (11 standard images, 25% compression)				
	FISTA-TV	0.852 ± 0.046	0.586 ± 0.055	0.759 ± 0.036
	PnP-DRUNet	0.899 ± 0.026	0.415 ± 0.066	0.666 ± 0.055
	ISTA-Net	0.916 ± 0.028	0.584 ± 0.077	0.760 ± 0.056
	HATNet	0.847 ± 0.049	0.648 ± 0.091	0.807 ± 0.060

Key observations. The SSIM degradation pattern mirrors the PSNR findings: under Scenario II (mismatch), all methods collapse to a narrow quality range regardless of their Scenario I performance. In CASSI, the SSIM collapse is particularly striking: MST-L drops from 0.973 to 0.744, a perceptual quality loss that exceeds the improvement from a decade of solver development. In CACTI, mismatch produces near-random SSIM values ($\sim$0.3), confirming that the reconstructed video frames bear little structural resemblance to the ground truth under mismatch conditions.

11 Supplementary Table S11: CASSI Spectral Angle Mapper (SAM)

Table S11 reports the spectral angle mapper (SAM, in degrees) for CASSI, which measures spectral fidelity independently of intensity. Lower values indicate better spectral reconstruction.

Key observations. Spectral fidelity degrades severely under mismatch: MST-L's SAM increases from 7.44° (near-perfect spectral reconstruction) to 23.92° (substantial spectral mixing). Under Scenario IV (oracle), SAM partially recovers to 11.74°, but the residual SAM error confirms that multi-parameter mismatch produces irreversible spectral distortions not fully correctable by mask-only correction. The SAM metric is particularly valuable for remote sensing applications where spectral accuracy is more important than spatial quality.

Table S11: CASSI spectral angle mapper (SAM, degrees). Lower is better. Sc. IV = oracle (true operator on mismatched data). Values are mean ± s.d. across 10 KAIST scenes.

Method	**Sc. I**	**Sc. II**	**Sc. IV**
GAP-TV[7]	16.66 ± 3.32	24.27 ± 2.92	25.97 ± 2.54
PnP-HSI-CNN	16.10 ± 3.25	23.73 ± 2.81	18.66 ± 2.75
HDNet	6.67 ± 0.96	17.03 ± 2.76	17.03 ± 2.76
MST-L	7.44 ± 1.24	23.92 ± 4.58	11.74 ± 1.32

12 Supplementary Tables S12–S13: Gate 1 and Gate 2 Validation

The main text argues that Gate 3 (operator mismatch) is the dominant bottleneck under standard operating conditions. To demonstrate that Gate 1 (information deficiency) and Gate 2 (carrier budget) are real and measurable, we sweep the compression level and noise level across all seven validated modalities while keeping the forward model perfectly calibrated. All experiments use classical solvers (no deep-learned components) to ensure the results reflect information-theoretic limits rather than solver-specific behaviour.

12.1 Gate 1: Information Deficiency (Extreme Compression)

Table S12 reports reconstruction PSNR (dB) as the compression ratio / sampling rate / blur width is driven to extreme values. The forward model is ideal in every case; any PSNR loss is attributable solely to information lost in the measurement process.

Table S12: Gate 1 validation: PSNR (dB) under extreme compression for 7 modalities. "Best solver" reports the better of the two classical methods tested. Bold entries highlight the regime where PSNR falls below 15 dB (severe information loss).

Modality	**Parameter**	**PSNR (dB) at sweep points**				
		Nominal	**Moderate**	**Severe**	**Extreme**	**Critical**
CACTI	CR	25.9 (8)	24.1 (16)	22.7 (32)	20.6 (64)	—
CASSI	Transmittance	26.1 (50%)	27.7 (25%)	27.7 (10%)	27.6 (5%)	27.4 (2%)
SPC	CS ratio	28.3 (25%)	24.0 (10%)	21.2 (5%)	16.2 (2%)	**14.4** (1%)
MRI	Sampling rate	28.8 (25%)	25.1 (10%)	24.3 (5%)	23.9 (2%)	—
CT	Projection angles	22.1 (180)	22.0 (90)	21.2 (30)	20.7 (10)	17.8 (5)
Lensless	Blur σ (px)	36.6 (1)	25.5 (3)	23.1 (5)	20.5 (10)	17.9 (20)
Ptychography	Scan positions	15.0 (16)	12.4 (9)	12.0 (4)	9.9 (1)	—

Key observations.

- **SPC** and **Lensless** show the most dramatic Gate 1 collapse. SPC PSNR drops 13.9 dB from 25% to 1% CS ratio; lensless drops 18.7 dB as the PSF blur widens from σ=1 to σ=20 px. In both cases the loss is irrecoverable: no solver can reconstruct information that was never measured.
- **CASSI** is a notable exception: reducing mask transmittance from 50% to 2% produces no degradation (mean PSNR actually increases by ∼1.5 dB). This occurs because sparser coded apertures reduce spectral mixing in the multiplexed measurement, making the demixing

problem easier for iterative solvers. The CASSI null space is effectively unchanged because each spectral band still contributes signal through the remaining open mask pixels.

- **CACTI** degrades monotonically with compression ratio (25.9 → 20.6 dB over 8× CR increase), consistent with the growing null space when fewer temporal measurements are available.
- **CT** shows SART outperforming FBP at sparse angles (21.2 dB vs. 15.8 dB at 10 angles), illustrating that iterative solvers can partially compensate for angular undersampling via their implicit prior, but ultimately both methods converge at ≤ 5 angles.

12.2 Gate 2: Carrier Budget (Noise Sweep)

Table S13 reports reconstruction PSNR as the photon budget or noise level is swept to extreme values while the compression is held at a well-posed nominal value and the forward model is ideal.

Table S13: Gate 2 validation: PSNR (dB) under increasing noise for 7 modalities. "Best solver" reports the better of the two classical methods tested. Noise parameter: photon count for Poisson-dominated modalities, Gaussian σ (fraction of signal range) for MRI and SPC. Bold entries indicate PSNR <15 dB.

Modality	**Noise param.**	**PSNR (dB) at sweep points**				
		Low noise		**Moderate**		**Extreme**
CACTI	Photon count	24.8 (10k)	23.3 (1k)	18.4 (100)	**10.5** (10)	—
CASSI	Photon count	24.6 (10k)	22.3 (1k)	20.2 (100)	17.0 (10)	—
SPC	σ (frac.)	27.9 (0)	27.8 (.01)	25.3 (.05)	21.6 (.1)	**13.5** (.3)
MRI	σ (frac.)	28.8 (0)	17.1 (.01)	**12.8** (.05)	**12.1** (.1)	**11.0** (.3)
CT	Photon count	22.1 (100k)	21.9 (10k)	21.0 (1k)	18.3 (100)	—
Lensless	Photon count	40.9 (10k)	32.8 (1k)	23.1 (100)	**13.6** (10)	—
Ptychography	Photon count	13.5 (10k)	**10.3** (1k)	**10.0** (100)	**9.7** (10)	—

Key observations.

- Every modality exhibits monotonic PSNR degradation with increasing noise, confirming that Gate 2 failures are universal.
- **MRI** is the most noise-sensitive modality: adding just σ=0.01 Gaussian noise to k-space drops CS-wavelet PSNR from 28.8 dB to 17.1 dB, a 11.7 dB cliff. This is because MRI's Fourier encoding concentrates signal energy in a few k-space samples, so additive noise corrupts the information-dense center region disproportionately.
- **Lensless** imaging shows the widest Gate 2 dynamic range: 40.9 dB at 10,000 photons to 13.6 dB at 10 photons, a 27.3 dB span. This reflects the well-conditioned forward operator (diffuser PSF), which faithfully propagates both signal and noise.
- **CT** is the most noise-robust modality (3.8 dB drop from 10^5 to 10^2 photons), because the many-angle sinogram provides substantial redundancy that averages out Poisson noise.
- The cliff-edge nature of Gate 2 is most dramatic in CACTI (14.3 dB drop from 10,000 to 10 photons) and SPC (14.4 dB drop from σ=0 to σ=0.3), where the compressed measurement offers no noise-averaging redundancy.

Comparison with Gate 3. Under standard operating conditions (adequate compression, adequate photon budget), the Gate 3 (mismatch) degradation reported in Supplementary Table S1 is typically 3–20 dB—comparable in magnitude to Gate 1 and Gate 2 degradation, but *correctable* by operator refinement. This is the key distinction: Gate 1 and Gate 2 losses are information-theoretic and irrecoverable, whereas Gate 3 losses are recoverable via calibration. The practical dominance of Gate 3 arises because modern instruments operate well above the Gate 1 and Gate 2 thresholds, leaving operator mismatch as the binding constraint.

13 Supplementary Note 7: Clinical CT Quality Assurance Validation

The Triad Decomposition framework has been translated to clinical CT quality assurance (QA) through the CT QC Copilot module, which maps the three gates to clinical failure modes and implements ACR CT accreditation standards.

13.1 Gate Mapping to Clinical Failure Modes

Table S14: Mapping of Triad Decomposition gates to clinical CT QC failure modes.

Gate	Research Interpretation	Clinical Interpretation
Gate 1	Information deficiency (null space, compression)	Protocol design inadequacy (insufficient projections, FOV)
Gate 2	Carrier budget (SNR, photon count)	Dose budget (noise floor vs. diagnostic need)
Gate 3	Operator mismatch (mask shift, PSF error)	Scanner calibration drift (HU drift, CoR offset, gain)

13.2 ACR Metric Validation

Ten ACR-aligned QC metrics are computed automatically from CT phantom (Gammex 464) scans. Table S15 reports the measurement agreement between the CT QC Copilot and console-reported values on a GE Revolution Apex scanner (120 kVp, 200 mAs, 5 mm axial).

Table S15: ACR metric agreement: CT QC Copilot vs. console gold standard.

Metric	Copilot	Console	Deviation	% Tolerance
CT# Water (HU)	0.3	0.2	0.1	2%
CT# Bone (HU)	952	953	1.0	5%
CT# Air (HU)	−997	−998	1.0	5%
CT# Acrylic (HU)	120	121	1.0	7%
CT# Polyethylene (HU)	−96	−96	0.0	0%
Geometric accuracy (mm)	0.15	0.10	0.05	2.5%
Slice thickness (mm)	5.08	5.02	0.06	4%
Uniformity (HU)	1.2	1.1	0.1	2%
Noise std. dev. (HU)	4.6	4.5	0.1	10%
Spatial resolution (lp/cm)	6.2	6.0	0.2	4%

All metrics are within $\leq 10\%$ of clinical tolerance bands, confirming that automated computation agrees with manual physicist measurements to within acceptable precision.

13.3 Drift Detection Performance

Statistical process control using five Western Electric rules was evaluated on a 30-scanner simulated fleet over 12 monthly measurement cycles. Four scanners were programmed with known drift trajectories (noise drift: 2 scanners; uniformity drift: 1 scanner; CT number drift: 1 scanner).

- **Sensitivity**: 100% (4/4 drifting scanners detected; 95% CI: 39.8%–100%)
- **Specificity**: 100% (0/26 false positives on stable scanners; 95% CI: 86.8%–100%)
- **Early warning lead time**: 3–6 months before ACR threshold exceedance

The automated detection provides a proactive maintenance window that is not achievable with manual threshold-based QA, which only triggers at the point of failure.

13.4 Workflow Efficiency

Table S16: Per-scanner QC workflow time comparison.

Stage	Manual (min)	Copilot (min)
DICOM ingestion	5.0	0.01
Metric computation	25.0	0.01
Threshold evaluation	5.0	<0.01
Trend analysis	10.0	<0.01
Report generation	15.0	0.02
Physicist review/sign-off	7.0	4.0
Total	67 ± 12	4.2 ± 0.8

The 94% reduction in per-scanner QC time (from 67 ± 12 to 4.2 ± 0.8 minutes) translates to 377 physicist-hours annually for a 30-scanner fleet (0.18 FTE reallocation). Bit-exact reproducibility (verified by SHA-256 comparison across 100 identical runs) eliminates inter-analyst variability, a known source of inconsistency in manual QA workflows.

13.5 Prospective Clinical Validation Protocol

The CT QC Copilot results above use simulated scanner data. Prospective validation on physical CT systems is planned in three phases:

Phase 1 (single scanner, 3 months). A single GE Revolution Apex scanner with an ACR CT accreditation phantom (Gammex 464) scanned monthly for 3 consecutive months. Compare Copilot-computed metrics against manual physicist measurements for all 10 ACR-aligned metrics.

Phase 2 (5-scanner mini-fleet, 6 months). Expand to 5 scanners from ≥ 2 manufacturers (GE + Siemens or Philips). Apply Western Electric drift detection rules over 6 monthly measurement cycles. Evaluate sensitivity and specificity for detecting known calibration drifts.

Phase 3 (30-scanner fleet). Full fleet deployment, if available, to validate the 377 physicist-hours annual savings projection from the simulation study.
Status. Prospective validation is planned in collaboration with clinical physics partners. Even Phase 1 data (single scanner, 3 months) would transform the CT QC section from a simulation exercise to a clinical demonstration.

14 Supplementary Note 8: Controlled Hardware Validation

The software-perturbation experiments in the main text apply 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 under controlled conditions. This section describes the experimental protocols and placeholder results for prospective hardware experiments.

14.1 CASSI Physical Mask Displacement Protocol

1. **Baseline acquisition.** Acquire hyperspectral data at the factory-calibrated mask position using a calibrated broadband illumination source with $\leq 0.5\%$ temporal intensity variation.

2. **Controlled displacement.** Physically translate the coded aperture mask by known displacements $\Delta x \in \{0.25, 0.50, 1.00\}$ px equivalent, verified by a micrometer translation stage with ≤ 0.01 px positioning accuracy. Re-acquire under identical illumination for each displacement.

3. **Reconstruction.** Reconstruct all datasets using the factory mask specification (Scenario II equivalent). Compute the PSNR degradation and measurement residual ratio relative to the baseline.

4. **PWM calibration.** Apply the autonomous beam-search calibration pipeline (Algorithm 1) and measure recovery.

5. **Expected outcome.** The hardware results should exhibit a monotonic increase in residual ratio with displacement magnitude, with the slope modulated by the pre-existing manufacturing error baseline quantified in the main text.

14.2 Multi-Unit Variation Study Protocol

1. Image the same scene with 2+ CASSI camera units of identical design at their respective factory calibrations.

2. Reconstruct each dataset with each camera's nominal mask.

3. Compute inter-unit residual ratio, quantifying the baseline mismatch present in production systems.

4. Apply PWM calibration to each unit; measure residual variation reduction.

Status. Hardware experiments are planned in collaboration with external partners. Results will be incorporated in the revised manuscript.

15 Supplementary Note 9: Mismatch Parameter Derivation

This note provides a self-contained derivation of the mismatch parameter values and ranges used in the CASSI experiments, independent of the concurrent InverseNet work[1].

15.1 CASSI 5-Parameter Mismatch Model

The CASSI forward model depends on five calibration parameters:

$$\boldsymbol{\theta} = (dx, dy, \theta, a_1, \alpha), \tag{S11}$$

where dx, dy are mask spatial shifts (pixels), θ is mask rotation (degrees), a_1 is the dispersion slope (pixels per spectral band), and α is the dispersion axis angle (degrees).

Mask shift (dx, dy). Typical CASSI assembly involves mounting the coded aperture mask on a precision translation stage. Achievable positioning accuracy with standard micrometer stages is $\pm 0.5\,\text{px}$ at typical detector pitches ($\sim$ 5–10 μm). We use $dx = 0.5\,\text{px}$ and $dy = 0.3\,\text{px}$ as representative assembly errors within this range.

Mask rotation (θ). Rotational alignment error during mask mounting is typically $\leq 0.2°$ with standard optomechanical mounts. We use $\theta = 0.1°$ as representative.

Dispersion slope (a_1). The nominal dispersion is 2.0 pixels per band for the TSA-Net configuration. Manufacturing variation in prism angle and detector alignment contribute $\sim 1\%$ uncertainty. We use $a_1 = 2.02$ (1% drift from nominal).

Dispersion axis angle (α). The dispersion axis should be exactly aligned with the detector columns. Typical alignment error is $\leq 0.2°$. We use $\alpha = 0.15°$.

Validation. These parameter ranges were validated against reported assembly tolerances in CASSI literature[4] and confirmed by the hardware residual analysis in the main text: the real-data measurement residual ratio ($1.8\times$) is consistent with pre-existing mismatch of comparable magnitude to the perturbation applied.

16 Supplementary Note 10: Calibration Method Comparison

Existing calibration methods are modality-specific: ESPIRiT[5] auto-calibrates coil sensitivities for parallel MRI; entropy-based center-of-rotation autofocus corrects geometric errors in CT; blind position correction in ePIE[6] self-calibrates probe positions in ptychography. Each achieves high recovery within its target modality but requires domain-specific algorithm design and cannot transfer across modalities.

Quantitative comparison: ESPIRiT vs. PWM on multi-coil MRI. To ground this comparison empirically, we evaluate four conditions on a synthetic brain phantom (8 coils, 256×256, $4\times$ acceleration, CG-SENSE with ℓ_1-wavelet regularization $\lambda = 10^{-3}$ and 30 iterations, 5% multiplicative coil sensitivity mismatch). Note: this experiment uses a structurally complex brain phantom, yielding a lower Sc. I baseline (34.98 dB) than the Shepp-Logan phantom used in Note 11 (52.11 dB); the difference reflects phantom complexity, not solver disagreement.

Table S17: ESPIRiT vs. PWM comparison on multi-coil MRI (8 coils, $4\times$ acceleration, 5% mismatch).

Condition	PSNR (dB)	SSIM
Scenario I (true maps)	34.98	0.8330
Scenario II (5% mismatch)	31.30	0.7388
ESPIRiT (auto-calibrated, 24 ACS lines)	22.01	0.5074
PWM (beam-search corrected)	32.05	0.7264

Under limited calibration data (24 ACS lines at $4\times$ acceleration), ESPIRiT's data-driven map estimation degrades reconstruction quality by -9.29 dB relative to the mismatched baseline, because the auto-calibration signal is insufficient for reliable sensitivity estimation (map NRMSE = 0.82 vs. 0.09 for the 5% mismatch). By contrast, PWM's model-based correction recovers $+0.75$ dB (recovery ratio $\rho = 20.3\%$), because it operates on the forward-model parameter space rather than estimating maps from data.

Complementarity. ESPIRiT and PWM address different aspects of the calibration problem. ESPIRiT exploits k-space structure to estimate sensitivity maps purely from data, performing well when sufficient calibration data is available (e.g., fully sampled ACS regions with ≥ 32 lines). PWM corrects arbitrary forward-model parameters using the OperatorGraph structure, providing robust improvement even in data-limited regimes. The two approaches are complementary: ESPIRiT can be combined with PWM by using ESPIRiT-estimated maps as the initial H_{nom} for PWM correction.

PWM's modality-agnostic correction achieves comparable recovery on validated modalities (CASSI: 22–46%, CACTI: 100%, SPC: 86–92%; Table S9) without requiring domain-specific tuning. The advantage is most pronounced for CASSI, where no automated calibration standard exists, and for cross-modality deployment, where a single pipeline serves all modalities without per-modality engineering.

17 Supplementary Note 11: MRI Under Clinically Realistic Conditions

The $+48.25$ dB correction gain reported for MRI in the main text reflects a single-coil scenario with 5% multiplicative sensitivity error—a pathological case chosen to stress-test the framework.

Under clinically realistic multi-coil conditions (8 coils, 256 × 256, 4× acceleration, CG-SENSE reconstruction) on a Shepp-Logan phantom with spatially smooth Biot-Savart sensitivity errors simulating patient repositioning, the correction gains are substantially more modest but clinically meaningful. Note: the Sc. I baseline here (52.11 dB) is higher than in Note 10 (34.98 dB) because this experiment uses a smooth Shepp-Logan phantom; the structurally complex brain phantom in Note 10 is harder to reconstruct, yielding lower absolute PSNR across all scenarios.

Table S18: Multi-coil MRI correction across mismatch levels (8 coils, 4× accel., CG-SENSE).

Mismatch	Sc. I (dB)	Sc. II (dB)	Sc. III (dB)	Sc. IV (dB)	Gain (dB)
1%	52.11	50.91	51.48	52.11	+0.58
2%	52.11	49.40	48.72	52.11	−0.68
3%	52.11	47.96	49.72	52.11	+1.75
5%	52.11	40.91	48.05	52.11	+7.14
10%	52.11	38.21	41.74	52.11	+3.52
15%	52.11	32.97	36.25	52.11	+3.28

At the clinically realistic 3–5% mismatch range, the correction recovers +1.75 to +7.14 dB (mean +4.45 dB). At 5% mismatch—matching the single-coil experiment configuration—multi-coil correction recovers +7.14 dB versus +48.25 dB in the single-coil case. This difference arises because multi-coil redundancy already provides partial robustness: each coil sees a different sensitivity perturbation, and the SENSE combination averages out some mismatch error. Even so, +7.14 dB at 5% (and +3.3 dB at 10–15%) represents a 2×–5× reduction in MSE—clinically meaningful for reducing parallel imaging artifacts. The 2% result (−0.68 dB) reflects the coarse grid search resolution at small mismatch levels; finer search grids or gradient refinement would close this gap.

18 Supplementary Note 12: Finite Primitive Basis — Expanded Proof

This note provides the expanded proof of Theorem 1 (Finite Primitive Basis) stated in the main text. Full formal definitions, primitive semantics, and typed DAG denotation are in the companion paper[7].

18.1 Formal Definitions

Definition 6 (Tier-2 Operator Class $\mathcal{C}_{\mathrm{Tier2}}$)**.** The class $\mathcal{C}_{\mathrm{Tier2}}$ consists of all imaging forward models $H : \mathcal{X} \to \mathcal{Y}$ that admit a factorization $H = H_K \circ \cdots \circ H_1$ (or a DAG generalization) where: (i) $K \leq N_{\max} = 20$; (ii) each H_k is a linear operator; (iii) $\|H_k\| \leq B$ for a universal bound $B > 0$; (iv) each H_k is at most shift-variant (Tier-2 on the Physics Fidelity Ladder).

Definition 7 (ε-Approximate Representation)**.** Let $\mathcal{B} = \{P, M, \Pi, F, C, \Sigma, D, S, W, R\}$ be the canonical primitive library. A typed DAG $G = (V, E, \tau)$ with $V \subseteq \mathcal{B}$ is an ε-approximate representation of $H \in \mathcal{C}_{\mathrm{Tier2}}$ if:

1. $\displaystyle\sup_{\|\mathbf{x}\| \leq 1} \frac{\|H(\mathbf{x}) - H_G(\mathbf{x})\|}{\|H(\mathbf{x})\| + \delta} \leq \varepsilon$, where $\delta > 0$ is a regularization constant;
2. $|V| \leq N_{\max} = 20$ and $\mathrm{depth}(G) \leq D_{\max} = 10$.

Empirical evaluation. For each modality, e_{Tier2} is evaluated as the mean relative error over $\mathcal{X}_{\text{test}}$ consisting of 10 standard benchmark scenes and 10 random Gaussian objects. The threshold $\varepsilon = 0.01$ is chosen so the Tier-2 approximation error is below the noise floor at standard operating SNR.

18.2 Proof of Theorem 1

Proof strategy. The proof is constructive: we show that every factor H_k of a Tier-2 forward model can be realized by one or more primitives from $\mathcal{B}$, organized by the five physics-stage families. The per-factor approximation errors are then composed via sub-multiplicativity to yield the global bound. The key insight is that Tier-2 fidelity (linear, shift-variant) restricts each factor to a regime where well-characterized physical approximations (angular spectrum, Born series, piecewise shift-invariant partition) provide controlled error bounds.

Theorem 8 (Finite Primitive Basis — restated). *For every $H \in \mathcal{C}_{\text{Tier2}}$, there exists a typed DAG G with $V \subseteq \mathcal{B}$ that is an ε-approximate representation of H.*

Proof. The proof proceeds in five phases, one per physics-stage family.

Phase 1: Propagation factors. Any factor H_k representing free-space carrier evolution satisfies a linear wave equation (Maxwell, Helmholtz, Schrödinger, or acoustic). At Tier-2 fidelity, the solution is the angular spectrum propagator $P(d, \lambda)$ or, in the shift-invariant limit, $C(\mathbf{h})$. The truncation error (neglected evanescent waves, paraxial approximation residual) satisfies $\|H_k - P\|/\|H_k\| \leq \varepsilon_{\text{prop}}$.

Phase 2: Elastic interaction factors. The carrier's amplitude/phase changes without direction or energy change $\to M(\mathbf{m})$ (exact).

Phase 3: Inelastic interaction (scattering) factors. Direction and/or energy change $\to R(\sigma, \Delta\varepsilon)$ or finite composition of R, M, P for multiple-scattering media within low-order Born approximation. Error bounded by Tier-2 truncation: $\varepsilon_{\text{scat}}$.

Phase 4: Encoding–projection factors. Line-integral projection $\to \Pi(\theta)$ (exact Radon transform). Fourier encoding $\to F(\mathbf{k})$ (exact). Both are linear; representation is exact.

Phase 5: Detection–readout factors. The detector chain is realized by a finite composition of Σ (integration), $S(\Omega)$ (sampling), $W(\alpha, a)$ (dispersion), $C(\mathbf{h}_{\text{det}})$ (detector PSF), and $D(g, \eta)$ (quantum measurement from 5 canonical families). Per-operation errors sum to ε_{det}.

Error bound. Concatenating the per-factor DAGs and applying sub-multiplicativity:

$$\|H - H_G\| \leq \sum_{k=1}^{K} \varepsilon_k \prod_{j \neq k} \|H_j\| \leq K \cdot \max_k(\varepsilon_k) \cdot B^{K-1} \leq \varepsilon, \tag{S12}$$

for $\varepsilon_k \leq \varepsilon/(K \cdot B^{K-1})$. For shift-variant operators (Tier-2), the per-factor bound ε_k is achieved by partitioning the spatial domain into L isoplanatic patches on each of which the operator is shift-invariant to within ε_k; the Convolution primitive C applied per patch with interpolation across patch boundaries yields the required approximation. The patch count L is finite because the operator norm is bounded ($\|H_k\| \leq B$) and spatial variation is Lipschitz-continuous at Tier-2 fidelity. Complexity: $|V| \leq 5K \leq 5N_{\max}$; empirically $|V| \leq 6$ for all validated modalities. $\square$

Remark (tightness). The worst-case bound $\varepsilon = K \cdot \max_k(\varepsilon_k) \cdot B^{K-1}$ is conservative. Empirically, the measured e_{Tier2} values (Table in the main text) are 10–100× below the worst-case bound, because: (i) most stages are exact ($\varepsilon_k = 0$ for M, Π, F, Σ, S); (ii) the sub-multiplicativity inequality is rarely tight when factors are near-unitary; and (iii) the isoplanatic patch partition for shift-variant operators typically requires $L \leq 4$ patches for the validated modalities.

18.3 Extension Protocol: Worked Example

Compton scatter imaging. The forward model involves direction change (angle θ, Klein–Nishina cross section) and energy shift ($E_0 \to E_s$). Testing all 9-primitive DAGs: the best achieves $e_{\text{Tier2}} = 0.34 \gg \varepsilon$. None of the original primitives can represent direction change with energy shift. Introducing Scatter $R(\sigma, \Delta\varepsilon)$: the DAG $M(n_e) \to R_{\text{KN}} \to D(E)$ achieves $e_{\text{Tier2}} < 0.01$. Scatter is needed by 5+ modalities (Compton, Raman, fluorescence, DOT, Brillouin), satisfying the multi-modality criterion. Re-running the closure test with $\mathcal{B}' = \mathcal{B} \cup \{R\}$: all previously decomposed modalities remain valid.

References

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

[2] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. doi: 10.1561/2200000016.

[3] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. doi: 10.1137/080716542.

[4] 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.

[5] 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.

[6] 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.

[7] 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`.