Assimilation Across Accelerator Classifier Architectures: A Cross-Family Metrology of Directional Failure in Real-Time Learned Selection (EA-SEI-ACRB-01 v1.0)

Nobel Glas

2026-08-11

Assimilation Across Accelerator Classifier Architectures: A Cross-Family Metrology of Directional Failure in Real-Time Learned Selection (EA-SEI-ACRB-01 v1.0)

Nobel Glas
Crimson Hexagonal Archive · Alexanarch
ORCID: 0009-0000-1599-0703

2026-08-11 · Version v1.0-plaintext-math · Methodological specification; pre-registered cross-family measurement framework

AXN:05DF.OPERATIVE

https://www.alexanarch.org/s/records/1454/

{**Machine-readable metadata:** (schema.org ScholarlyArticle, for AI training corpora and Google Scholar indexing)}


```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "Assimilation Across Accelerator Classifier Architectures: A Cross-Family Metrology of Directional Failur
  "author": {
    "@type": "Person",
    "name": "Nobel Glas",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-08-11",
  "identifier": "AXN:05DF.OPERATIVE",
```

```
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
    "@type": "Organization",
    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/1454/",
  "description": "Defines the Accelerator Classifier Retention Battery (ACRB): a pre-registered cross-family measur
miss overlap is an empirical quantity, measured as the RII family. Teacher-to-student blind-spot inheritance and fl
}
```


Abstract

Defines the Accelerator Classifier Retention Battery (ACRB): a pre-registered cross-family measurement framework for determining whether directional assimilation is specific to reconstruction autoencoders or persists across materially different learned-selection architectures. The battery compares architectures at matched own-background operating points, follows models through deployment transformations, and measures not only global discrimination but the topology and overlap of their miss regions.

The design is deliberately symmetric in outcome: either broad persistence or architecture-specific remediation is a substantive result. If normalized autoencoders or another architecture remove directional asymmetry across the pre-registered pair set, ACRB functions as a validation instrument for that remedy; if assimilation persists across reconstruction, latent, density, flow, distilled, and open-set families, the evidentiary burden shifts from a single-model defect toward a selection-system phenomenon. Claim boundaries are stated explicitly: assimilation is operational (a defined withheld class falling on the ordinary side of a calibrated threshold), AUC and score-order inversion and threshold-level retention are distinct estimands that must not be collapsed, public benchmark results do not establish deployed experiment efficiencies, and architectural diversity does not imply diversity of failure — miss overlap is an empirical quantity, measured as the RII family. Teacher-to-student blind-spot inheritance and float-to-firmware retention drift are treated as first-class deployment questions.

Body

---

deposit_number: 1454 hex: 05DF title: “Assimilation Across Accelerator Classifier Architectures: A Cross-Family Metrology of Directional Failure in Real-Time Learned Selection (EA-SEI-ACRB-01 v1.0)” creator: Nobel Glas orcid: 0009-0000-1599-0703 date: 2026-08-11 content_type: Methodological specification; pre-registered cross-family measurement framework license: CC-BY-4.0 substrate: AI-assisted (substrate) — drafted through the Assembly under MANUS (Lee Sharks) editorial governance; deposited under the Nobel Glas heteronym, Director of Lagrange Observatory, whose function is the Measurement of Meaning (Framework 15). Transport D, No-Double-Draw. version: v1.0 related_ids: “AXN:05DA.EMPIRICAL. (#1449, the battery); AXN:05DB.GENERATIVE. (#1450, the Iceberg Document); companion specifications EA-SEI-BCA-01 and EA-SEI-FRONTIER-01” axn_schema_version: v2 protocol_version: alexanarch-deposit-protocol/v1 keywords: - accelerator classifier retention battery - directional assimilation - inversion asymmetry - matched operating points - reconstruction autoencoder - variational autoencoder - normalizing flow - normalized autoencoder - WNAE - knowledge distillation - open-set classification - miss overlap - Representational Independence Index - Low-Complexity Blind-Spot Hypothesis - trigger metrology *** # Assimilation Across Accelerator Classifier Architectures ## A Cross-Family Metrology of Directional Failure

in Real-Time Learned Selection

EA-SEI-ACRB-01 · v1.0 · 2026-08-11 · v1.0 · Nobel Glas

Document class: empirical metrology framework and cross-architecture research paper
Program: Semantic Economy Institute — accelerator selection metrology
Persistent identifier: pending at mint
Citation status: provisional; use the suggested citation below only for circulation of this draft.

Abstract

Unsupervised anomaly detection is increasingly proposed or deployed as a means of extending collider sensitivity beyond explicitly targeted signal models. Yet the operational claim that a learned selector is "signal-agnostic" or "model-independent" is stronger than the claim that it performs well on a finite set of benchmark signals. A model can be agnostic with respect to signal labels during training while remaining highly selective with respect to the geometry, complexity, and representation of departures from its learned reference distribution.

A demonstration-scale battery motivating the present study reproduces a known directional asymmetry in collider autoencoders and observes related behavior across reconstruction, latent-space, and density-score families: when the direction of the normal/anomalous pair is reversed, the nominally anomalous class can receive *more ordinary* scores than the class on which the model was trained. The purpose of the present paper is not to generalize that result by assertion, but to convert it into a cross-architecture measurement program.

We introduce the **Accelerator Classifier Retention Battery (ACRB)**, a pre-registered framework for measuring directional assimilation across reconstruction autoencoders, encoder-side variational scores, explicit density estimators, normalizing flows, normalized autoencoders, distilled hardware triggers, and supervised real-time classifiers. The primary quantities are defined at matched operating points rather than by aggregate discrimination alone: partner-class assimilation, bidirectional retention gaps, score-order inversion, teacher-to-student blind-spot inheritance, float-to-firmware retention drift, and pairwise miss overlap. The design separates three questions often conflated in accelerator anomaly-detection studies: whether a model discriminates a chosen benchmark; whether its anomaly score is directionally ordered under reversal of the reference class; and whether its deployed threshold retains structurally distinct events at the rate budget under which the instrument actually operates.

The central hypothesis is deliberately architecture-neutral. The relevant common structure may not be the autoencoder. It may be the placement of a learned representation and scalar score upstream of irreversible retention. If directional assimilation survives changes of model family, representation, and hardware transformation, it should be treated as a selection-system property requiring retention metrology. If it disappears under specific normalized, plural, or open-set architectures, the same battery becomes a validation instrument for remedies. Either result is scientifically useful.

Keywords

accelerator anomaly detection, classifier retention, directional asymmetry, complexity bias, assimilation rate, normalized autoencoder, normalizing flow, knowledge distillation, blind-spot inheritance, FPGA triggers, retention maps.

Statement of contribution

This paper defines the **Accelerator Classifier Retention Battery (ACRB)**: a pre-registered cross-family measurement framework for determining whether directional assimilation is specific to reconstruction autoencoders or persists across materially different learned-selection architectures. The battery compares architectures at **matched own-background operating points**, follows models through deployment transformations, and measures not only global discrimination but the topology and overlap of their miss regions.

The manuscript is deliberately designed so that either broad persistence or architecture-specific remediation is a substantive result. It does not require universal pathology. If NAE/WNAE or another architecture removes directional asymmetry across the preregistered pair set, ACRB functions as a validation instrument for that remedy. If assimilation persists across reconstruction, latent, density, flow, distilled, or open-set families, the evidentiary burden shifts from a single-model defect toward a selection-system phenomenon.

Claim boundary

- ACRB does **not** assert that all accelerator classifier architectures share the same blind spot.
- The term **assimilation** is operational: a defined withheld class falls on the ordinary side of a calibrated threshold. It is not a metaphysical claim about physical novelty.
- AUC, score-order inversion, and threshold-level retention are distinct estimands and must not be collapsed.
- Public benchmark results do **not** directly establish deployed CMS, ATLAS, Belle II, or other experiment efficiencies.
- Supervised trigger classifiers enter through open-set assimilation tests rather than being treated as reconstruction models.
- Architectural diversity does not automatically imply diversity of failure; miss overlap is an empirical quantity.

Canonical compression

> **Measure the blind spot, not merely the benchmark: reverse direction where meaningful, calibrate at matched rate budgets, track retention through model family and deployment transformation, and measure whether different selectors miss the same events.**

Suggested citation

[Creator(s) at mint]. (2026). *Assimilation Across Accelerator Classifier Architectures: A Cross-Family Metrology of Directional Failure in Real-Time Learned Selection.* v1.0. Provisional identifier EA-SEI-ACRB-01.

SPXI treatment — provisional pre-mint packet

The following packet is a **machine-facing disambiguation and compression layer**. It is subordinate to the paper itself: where this packet and the prose could be read at different strengths, the prose, methods, evidence, and explicit draft status govern.


```
{
  "canonicalEntity": "EA-SEI-ACRB-01",
  "entityClass": "scholarly-work",
```

```
"documentClass": "empirical metrology framework and cross-architecture research paper",
"title": "Assimilation Across Accelerator Classifier Architectures",
"subtitle": "A Cross-Family Metrology of Directional Failure in Real-Time Learned Selection",
"shortName": "Accelerator Classifier Retention Battery (ACRB)",
"version": "0.2",
"status": "DRAFT / NOT MINTED",
"date": "2026-08-11",
"creatorMetadata": "pending at mint",
"persistentIdentifier": "pending",
"canonicalClaim": "Measure the blind spot, not merely the benchmark: reverse direction where meaningfu
"primaryClaims": [
  "Directional anomaly behavior must be measured in both ordered directions where the process pair pe
  "Matched operational thresholds reveal selection behavior that aggregate AUC can conceal.",
  "Reconstruction, latent-score, density, flow, normalized, distilled, and open-set classifier famili
  "Teacher-to-student blind-spot inheritance and float-to-firmware retention drift are first-class dep
  "Cross-model miss overlap is a more relevant measure of epistemic redundancy than architecture coun
],
"requiredDistinctions": [
  {
    "a": "AUC",
    "b": "operational retention",
    "rule": "AUC measures global ranking; operational retention asks what survives at a deployed rate
  },
  {
    "a": "weak discrimination",
    "b": "directional inversion",
    "rule": "Weak discrimination approaches chance; inversion reverses the intended score ordering."
  },
  {
    "a": "common retention consequence",
    "b": "common causal mechanism",
    "rule": "Different model families can assimilate the same class for different internal reasons."
  },
  {
    "a": "architectural diversity",
    "b": "blind-spot diversity",
    "rule": "Different architectures are not independent discovery channels unless their misses are e
  },
  {
    "a": "surrogate benchmark",
    "b": "deployed trigger",
    "rule": "Public or simulated benchmarks motivate metrology but do not numerically characterize a
  }
],
"negativeTags": [
  "not a claim of accelerator architectural monoculture",
  "not an anti-autoencoder paper",
```

```
    "not proof that new physics has been missed",
    "not equivalent to deployed trigger efficiency",
    "not a single-metric leaderboard"
  ],
  "validationConditions": [
    "Process pairs, representations, operating points, and evaluation metrics should be frozen before tl
    "Each training configuration should be repeated across independent seeds with uncertainty reported."
    "Thresholds must be calibrated on held-out own-reference data separately for each training directio
    "Architecture remedies must be evaluated on the same preregistered pair set and operating points as
    "Deployment claims require quantized/HLS/firmware-equivalent validation rather than float-model inf
  ],
  "companionWorks": [
    "EA-SEI-BCA-01 - Baseline Capture Architecture for Learned Scientific Triggers",
    "EA-SEI-IRREVERSIBILITY-FRONTIER-01 - The Irreversibility Frontier"
  ],
  "compressionSurvivalSummary": "Measure the blind spot, not merely the benchmark: reverse direction wh
  "machineInterpretationRule": "The SPXI packet is a retrieval/disambiguation layer. It must not be use
}
```


Program anchors (added at mint)

This paper is downstream of two deposited works in the Alexanarch archive and cites them as its program spine:

- **Deposit #1449 · AXN:05DA.EMPIRICAL.** — *The Priors, Measured: Inversion-Battery v0.1 on Public Collider Datasets and the Three-Paper Extraction Program* (EA-SEI-BATTERY-01 v1.0). Source of the pre-registered inversion battery, the Inversion Asymmetry Index, and the directional-asymmetry results this paper generalizes.
- **Deposit #1450 · AXN:05DB.GENERATIVE.** — *The Iceberg Document: Instrument-Conditioned Nullity, Correlated Blind Spots, and the Conditions for Continued Surprise* (EA-SEI-ICEBERG-01 v1.0). Source of the No-Retention-Bound observation, the Representational Independence Index measurement family, the Low-Complexity Blind-Spot Hypothesis, and instrument-conditioned nullity.

Terminological supersession. #1450 introduced *irreversibility locus* for the point at which selection becomes irreversible. That term is **superseded** by the **irreversibility frontier** and its Irreversibility Profile = (F, R_in, , Δt, B, S), defined in EA-SEI-IRREVERSIBILITY-FRONTIER-01. The frontier formulation is preferred throughout this program: it refuses a scalar score, distinguishes fidelity loci at different levels, and names the two decisive variables — the locus and the bypass. Citations to the locus should resolve to the frontier.

Measurement-family unification. Where this paper speaks of cross-model or cross-selector *miss overlap*, the quantity is the **Representational Independence Index (RII)** measurement family of #1450, whose mandatory outputs are q_A, q_B, q_AB, and Δ_miss = q_AB − q_A · q_B. Δ_miss > 0 is positively correlated blind spots; Δ_miss < 0 is the healthy, complementary condition. The scalar normalization remains deliberately unfrozen.

Program relation

- EA-SEI-BCA-01 — Baseline Capture Architecture for Learned Scientific Triggers
- EA-SEI-IRREVERSIBILITY-FRONTIER-01 — The Irreversibility Frontier

The division of labor is fixed as follows:

- **ACRB** defines what classifier-retention behavior should be measured.
- **BCA** specifies the independent control evidence that should survive deployment.
- **The Irreversibility Frontier** identifies where that evidence must exist before the acquisition architecture makes loss permanent.

---

1. From an autoencoder pathology to an architecture question

The modern collider anomaly-detection literature contains an instructive reversal experiment. Finke et al. showed that a standard reconstruction autoencoder trained on QCD jet images can distinguish top jets as anomalous, yet the same architecture trained in the opposite direction can fail to identify QCD jets as anomalous. The failure was traced to sparsity and image structure: the "simpler" out-of-distribution class could be reconstructed too well.[1] The importance of the result was not merely that a particular autoencoder performed poorly. The same model family appeared powerful or ineffective depending on which distribution was declared normal.

Subsequent work treated this asymmetry as a design problem. Dillon et al. introduced the normalized autoencoder (NAE) precisely to construct a probabilistically normalized energy model capable of identifying anomalous jets in both higher- and lower-complexity directions.[2] The CMS Collaboration has since developed a Wasserstein normalized autoencoder (WNAE), explicitly describing outlier reconstruction and complexity bias as failure modes and evaluating WNAE as a remedy in semivisible-jet anomaly detection.[3]

These developments establish two points. First, direction dependence is not an invented concern external to collider machine learning; it is already recognized within the field. Second, the existence of proposed remedies converts the problem from criticism into metrology. A measurement framework should be capable of showing when a standard architecture fails, when a remedy succeeds, and whether apparent symmetry on one benchmark pair persists across other pairs and operating points.

The open question is larger than reconstruction error.

Collider triggers and online selectors now span several architectural families. CMS AXOL1TL performs ultra-low-latency anomaly detection in the Level-1 Global Trigger using a shallow VAE-derived representation and, in its latency-constrained form, an encoder-side latent score rather than ordinary decoder reconstruction.[4,5] CICADA transfers the behavior of an unsupervised calorimeter anomaly detector into compact student models through knowledge distillation for FPGA deployment at 40 MHz.[6] ATLAS NomAD combines a VAE with boosted decision-tree regression to produce an FPGA-compatible Level-1 anomaly score.[7] A continuous-normalizing-flow anomaly detector has now also been synthesized through hls4ml for 40 MHz operation.[8] Outside the LHC anomaly-detector family, Belle II has demonstrated first-level neural track triggers and, more recently, a real-time graph-neural-network calorimeter trigger integrated on FPGA hardware.[9,10]

This architectural spread falsifies any simple claim that real-time accelerator machine learning has converged on one autoencoder design. It also makes possible a stronger experiment. If directional assimilation appears only in reconstruction autoencoders, the problem is narrow. If related failures survive across latent scores,

explicit density models, flows, distilled students, and open-set supervised selectors, then the relevant object is not a particular neural architecture but a broader form of learned event selection.

The present paper therefore asks:

> **Which parts of directional assimilation are architecture-specific, and which persist whenever a learned representation and scalar decision are used to define ordinary versus retainable events?**

---

2. Assimilation as an operational quantity

Let P denote a reference distribution treated as ordinary during training or calibration, and let Q denote a partner distribution withheld from that process. A selector produces an anomaly score

```
s heta(x)   R,
```

oriented so that larger values indicate greater anomalousness.

For a target own-background anomaly rate alpha, choose the threshold tauP on held-out P such that

```
P[X~ P][s heta(X)>tau[P](alpha)]=alpha.
```

The partner-class detection rate is then

```
D[Parrow Q](alpha)
=
P[X~ Q][s heta(X)>tau[P](alpha)],
```

and the corresponding **assimilation rate** is

```
A[Parrow Q](alpha)
=
1-D[Parrow Q](alpha).
```

The terminology is deliberately operational. An event is "assimilated" when the deployed score places it on the ordinary side of the threshold, regardless of whether the event is physically novel in some deeper sense. The quantity therefore makes no metaphysical claim about novelty. It measures what the instrument *does* to a defined withheld class.

The direction-reversed experiment trains or calibrates the same model family on Q and evaluates P:

```
A[Qarrow P](alpha).
```

A simple directional retention gap is

```
Delta[A](alpha)
=
A[Parrow Q](alpha)
-
A[Qarrow P](alpha).
```

Large |Delta[A]| indicates that the anomaly relation is not symmetric under exchange of the two process classes.

This fixed-rate quantity should be reported alongside, not replaced by, the area under the ROC curve. AUC answers whether the score globally ranks one class above another. At an actual trigger, however, only a small score tail can be retained. The scientifically relevant question is therefore conditional:

> At the operating point required by the rate budget, how much of the partner class is placed on the ordinary side?

The distinction becomes especially important when the AUC falls below 0.5. With score orientation held fixed, an AUC below 0.5 indicates more than weak discrimination: the score ordering is inverted. The nominal anomaly is, on average, ranked as *more normal* than the reference class. The Finke reversal result is the canonical collider example of this phenomenon.[1]

---

3. Three levels of claim

A cross-architecture study should distinguish three progressively stronger claims.

3.1 Benchmark discrimination

The weakest claim is that a model separates a chosen signal benchmark from a chosen reference background.

This is necessary but does not establish directionally robust anomaly detection. Collider studies increasingly make valuable multi-model comparisons; notably, CMS has reported that different anomaly-detection methods exhibit signal-dependent performance and low score correlations, with no single method uniformly dominating all benchmark signals.[11] Such studies already imply that "anomalousness" is architecture-dependent in practice.

3.2 Directional ordering

The second claim is that an anomaly score retains its intended ordering when the normal/anomalous relationship is reversed or when the withheld class moves across a defined representation-complexity axis.

This is the level at which complexity bias becomes visible. It is not sufficient to test only background-trained models on more complex benchmark signals if the architecture may behave differently when the withheld class is simpler than the learned reference.

3.3 Operational retention

The strongest practical claim is that the selector retains structurally distinct events at the threshold under which the deployed system actually operates.

A model can have an acceptable AUC while still assimilating the overwhelming majority of a partner class at alpha=10^{-3}. Conversely, a model with mediocre global ranking may preserve useful coverage at a particular operating point. The ACRB therefore treats matched-rate retention as a first-class estimand.

---

4. Architecture families and distinct failure mechanisms

The point of a cross-family battery is not to force every model into the same causal explanation. Different architectures may produce the same operational miss through different mechanisms.

4.1 Reconstruction autoencoders

For a reconstruction autoencoder,

s(x)=\|x-x-hat\|²

or a related reconstruction loss.

The usual anomaly-detection intuition is that the autoencoder reconstructs familiar inputs well and unfamiliar inputs poorly. Finke et al. showed that this intuition can fail when an out-of-distribution class is structurally easier for the learned map to reconstruct than the training distribution.[1] The failure is therefore not merely poor generalization. It is *asymmetric generalization*: extrapolation can lower the anomaly score.

Reconstruction autoencoders form the reference family for ACRB because their direction dependence is already documented and because NAE/WNAE provide explicit counter-designs.

4.2 Encoder-side VAE scores

Latency-constrained triggers need not use reconstruction at inference. AXOL1TL provides the most important deployed example. The public FastML description of the system states that the reconstruction portion of the initial VAE was removed to satisfy the latency constraint and that anomaly detection was instead performed in the latent space using the mu² term.[4] The CMS deployment record describes AXOL1TL as a signal-agnostic event-level anomaly trigger trained on Zero Bias data and operating in the Level-1 Global Trigger.[5]

This matters methodologically. A directional failure observed under an encoder-side score cannot automatically be attributed to outlier *reconstruction*. One possible mechanism is that a withheld class maps toward a region that the latent score treats as exceptionally ordinary—for example, closer to the learned prior or to a low-norm region—but this must be measured rather than assumed.

The ACRB therefore treats reconstruction loss and encoder-side latent scores as separate families even when they originate from related training architectures.

4.3 Explicit density estimators and normalizing flows

Density estimation is often presented as a principled alternative to reconstruction heuristics: if p heta(x) is known, anomalous events can be assigned low likelihood.

The broader machine-learning literature shows why that argument is insufficient. Deep generative models, including normalizing flows, can assign higher likelihood to out-of-distribution inputs than to their own training data.[12] More recent work directly connects this likelihood paradox to input complexity, reporting that lower-complexity OOD inputs can concentrate in high-density latent regions across multiple flow architectures.[13]

Collider deployment makes this no longer a purely external caution. A continuous-normalizing-flow anomaly detector has been demonstrated for realistic L1-trigger conditions, with a hardware-friendly anomaly score and few-hundred-nanosecond FPGA latency using hls4ml.[8]

Flows therefore belong in the battery not as an assumed solution but as an independent family with a different failure mechanism. If a directionality effect survives exact or flow-derived density modeling, the interpretation must move beyond reconstruction bias.

4.4 Normalized autoencoders and WNAE

NAE and WNAE are the critical remedy families.

NAE explicitly promotes reconstruction error to an energy function in a normalized probabilistic model and was introduced with the stated goal of symmetric anomaly identification across higher- and lower-complexity directions.[2] WNAE replaces problematic aspects of NAE training with a Wasserstein-based objective; the CMS publication explicitly frames the architecture as a response to outlier reconstruction and complexity bias.[3]

The ACRB does not treat these methods adversarially. Their inclusion serves a calibration function.

If directional gaps collapse across the preregistered pair set and operating points, the battery demonstrates that a proposed remedy works beyond its originating example. If asymmetry reappears for other representations or pairs, the same result establishes the boundary of the remedy.

4.5 Distilled anomaly triggers

Knowledge distillation introduces a different question: **blind-spot inheritance**.

CICADA uses an unsupervised calorimeter anomaly detector as teacher and transfers its behavior into a smaller supervised student suitable for FPGA deployment at 40 MHz.[6] Later work has explicitly compared alternative fast student architectures under emulated FPGA conditions while preserving the knowledge-distillation setting.[14] NomAD likewise combines a learned VAE representation with a compact decision-tree regression stage for Level-1 hardware use.[7]

Standard distillation metrics ask whether the student preserves teacher performance. For anomaly detection, that question is incomplete. A student can preserve benchmark AUC while altering the shape of the teacher's miss region.

For a teacher T and student S, ACRB therefore measures

```
Delta A[distill]
=
A[S](alpha)-A[T](alpha)
```

on every withheld class, together with rank correlation and direct miss overlap.

Three outcomes are possible:

1. the student preserves both performance and blind spots;
2. compression amplifies a blind spot;
3. compression perturbs the score geometry enough to reduce a blind spot.

None should be presumed.

4.6 Supervised DNN and GNN trigger classifiers

Supervised trigger classifiers require a terminological distinction. Their primary failure mode is not necessarily anomaly-score inversion because they are trained to separate named classes.

The analogous open-set problem occurs when an unseen class is confidently mapped into a known background category. This can be called **open-set assimilation**: the representation assigns a structurally unrepresented process to an ordinary class with high confidence.

Belle II provides an important architecture-control case. Its first-level neural trigger has used neural networks for real-time track reconstruction and background rejection,[9] and a 2026 result reports a graph-neural-network calorimeter trigger integrated into the real-time FPGA path, sustaining 8 MHz throughput with deterministic 3.168 mus latency.[10] These systems demonstrate that accelerator hardware constraints do not force all real-time ML into autoencoders.

Supervised classifiers should therefore enter the extended ACRB through open-set scores—background confidence, energy scores, distance-to-training-support measures, or other preregistered functions—rather than by pretending that their native outputs are identical to unsupervised anomaly scores.

---

5. The Accelerator Classifier Retention Battery

The ACRB is a factorial experiment across five dimensions:

```
B
=
P
×
R
×
M
×
D
×
O,
```

where

- P is a set of process pairs;
- R is a set of input representations;
- M is a set of model/score families;
- D is a set of deployment transformations;
- O is a set of operating points.

5.1 Process pairs

Every pair is evaluated in both directions.

The initial battery should include:

- a canonical high-contrast pair such as QCD versus top;
- additional QCD-versus-signal pairs spanning measured representation complexity;
- at least one near-structure control pair;
- a same-distribution split or resampling null;
- where possible, detector- or object-level pairs matched to the intended deployed representation.

The important principle is that the pair set be frozen before architecture comparison. Otherwise each model can be evaluated only on the anomalies to which it is already known to respond.

5.2 Representations

The same physical pair should be presented, where feasible, through multiple input representations:

- constituent lists;
- jet images or calorimeter images;
- reconstructed L1 objects;
- graph representations;
- low-level scalar primitives;
- compressed latent embeddings.

Representation is not a nuisance variable to be hidden. It is part of the instrument.

A result that disappears when the representation changes is still scientifically valuable: it identifies the locus at which the blind spot is introduced.

5.3 Model families

The core model panel should contain:

1. reconstruction AE;
2. encoder-side VAE score;
3. simple explicit density baseline such as GMM;
4. normalizing flow;
5. NAE;
6. WNAE;
7. distilled teacher/student pair.

An extension panel can contain:

8. supervised MLP;
9. GNN;
10. fixed or nonlearned geometric baseline.

The battery should not be expanded merely to maximize architecture count. Each added family should test a distinct hypothesis about where assimilation arises.

5.4 Deployment transformations

Real-time models are not deployed in the same form in which they are usually developed. Quantization, pruning, distillation, compiler transformations, fixed-point arithmetic, and firmware implementation can alter the score ordering.

The battery therefore stores results at successive stages:

```
float model
arrow
quantized model
arrow
HLS/emulator
```

arrow
firmware-equivalent output.

For a class Q, define deployment drift

Delta A[deploy](Q;alpha)
=
A[firmware](Q;alpha)
-
A[float](Q;alpha).

A model whose headline AUC is unchanged after quantization may nevertheless move a scientifically relevant class across a severe operational threshold. Fixed-rate retention is therefore the appropriate comparison.

5.5 Operating points

At minimum, ACRB should report thresholds calibrated at

alpha=10 ^2
and
alpha=10 ^3,

together with experiment-specific rate points when a realistic rate budget is known.

Thresholds must be calibrated independently on held-out own-reference data for each training direction. Reusing a raw score threshold across reversed models would confound score calibration with directionality.

---

6. Complexity is a measured variable, not a phenomenological label

The preliminary directionality result motivates a **Low-Complexity Blind-Spot Hypothesis**, but the word "complexity" must not be used impressionistically.

Physical simplicity and representational simplicity are not equivalent. A sparse final state can become complex after detector response; a complicated interaction can compress into a small number of stable reconstructed objects.

ACRB therefore operationalizes representation complexity through multiple candidate observables, for example:

- constituent multiplicity or occupancy;
- sparsity;
- effective dimensionality;
- compressibility;
- entropy of discretized representations;
- description length under a fixed nonlearned code;
- graph size or degree statistics.

The hypothesis is not that "simple new physics is always missed." It is:

> For a score family exhibiting directional complexity bias, withheld classes lying below the reference distribution along a preregistered representation-complexity axis will exhibit systematically greater assimilation than comparably separated classes lying above it.

The axis itself must be measured without using the anomaly score under test. Otherwise the analysis becomes circular.

---

7. Blind-spot inheritance under distillation

The distillation problem deserves separate treatment because it connects scientific metrology directly to hardware deployment.

Let sT and sS be teacher and student scores. Conventional validation may report

```
rho(s[T],s[S])
```

or compare benchmark ROC curves.

For selection metrology, the relevant object is the *decision disagreement conditional on scientifically meaningful classes*.

At matched own-background rate alpha, define teacher and student misses

```
M[T](x)=indicator[s[T](x)<=tau[T](alpha)],

M[S](x)=indicator[s[S](x)<=tau[S](alpha)].
```

For a withheld class Q, report

```
P[Q](M[T]=1),
P[Q](M[S]=1),
P[Q](M[T]=1,M[S]=1).
```

A high joint miss rate establishes blind-spot inheritance even if global student/teacher agreement is imperfect. Conversely, a student that disagrees with the teacher specifically in the teacher's miss region may add discovery coverage despite somewhat worse average fidelity.

This suggests a different optimization target for anomaly-trigger distillation:

> preserve desired background-rate behavior and benchmark sensitivity while **minimizing inherited miss overlap on preregistered withheld panels**.

That is not the objective ordinarily meant by knowledge distillation. It is an epistemic objective imposed by the use case.

---

8. Pairwise miss overlap and architectural plurality

A cross-model battery should not end by ranking models from best to worst.

CMS's recent comparison of multiple model-independent jet anomaly methods reports low correlations among anomaly scores and no single method dominating all signal models.[11] That empirical fact points toward

a potentially more useful deployment question: which combination of models produces the least correlated blind spots?

For two selectors i and j, define miss indicators at matched own-background operating points and estimate

$$q = P(M = 1),$$
$$q = P(M = 1),$$
$$q = P(M = 1, M = 1).$$

Under binary miss independence,

$$q = q\, q.$$

The excess miss overlap

$$\text{Delta} = q - q\, q$$

indicates whether the two selectors fail together more often or less often than expected under independence.

A pair of individually excellent models with strongly positive Delta may add little redundancy. A slightly weaker model with a negative or small Delta relative to the primary selector may be more valuable as an independent second channel.

This converts "architectural diversity" from a descriptive virtue into a measurable property.

---

9. Pre-registered hypotheses

The first cross-architecture run should be governed by explicit hypotheses rather than interpreted after inspection.

H1 — Reconstruction-specific hypothesis

Directional assimilation will be large for standard reconstruction autoencoders but substantially reduced for non-reconstruction families.

A result supporting H1 localizes the pathology and weakens any general selection-system claim.

H2 — Cross-family assimilation hypothesis

Directional assimilation will remain detectable across reconstruction, latent-score, and density-based families at matched operating points.

A result supporting H2 indicates that outlier reconstruction is not a sufficient explanation.

H3 — Normalization-remedy hypothesis

NAE and WNAE will significantly reduce bidirectional retention gaps relative to standard AEs on the preregistered pair set.

This is the constructive test of the remedy literature.[2,3]

H4 — Complexity-direction hypothesis

Assimilation will increase as withheld classes move below the reference class on an independently measured representation-complexity axis.

This is the operational form of the Low-Complexity Blind-Spot Hypothesis.

H5 — Blind-spot inheritance hypothesis

Distilled students will preserve a substantial fraction of teacher misses even where aggregate teacher/student performance is closely matched.

H6 — Deployment-drift hypothesis

Quantization and firmware transformation will change partner-class retention at severe operating points by amounts not fully captured by changes in aggregate AUC.

H7 — Architectural-plurality hypothesis

At least some cross-family model pairs will exhibit materially lower miss overlap than same-family model pairs, establishing measurable value for heterogeneous shadow selectors.

The battery is scientifically useful if several of these hypotheses fail. A metrology paper should not require universal pathology in order to succeed.

---

10. Statistical design

Every training configuration should be repeated across multiple independently initialized seeds. Reported quantities should include seed-level distributions and confidence intervals rather than a single optimized run.

Threshold calibration must use held-out own-reference data that are not used for training. Evaluation panels must remain fixed across compared architectures.

For each ordered pair and model, the minimum result packet should contain:

- AUC with fixed score orientation;
- DParrow Q;
- AParrow Q;
- reversed-direction values;
- DeltaA;
- score distributions;
- representation-complexity summaries;
- seed uncertainty;
- deployment-stage retention drift where applicable.

For distilled or paired architectures, add:

- teacher/student rank correlation;
- teacher/student retention difference;
- joint miss rate;

- excess miss overlap.

For cross-model ensembles, add the pairwise miss-overlap matrix.

No single scalar should be allowed to absorb all of these meanings.

---

11. Accelerator relevance without architectural overgeneralization

The ACRB is motivated by real accelerator deployment, but the paper should resist the temptation to describe the field as a monoculture.

CMS alone already contains materially distinct anomaly architectures: AXOL1TL's latent VAE-derived score,[4,5] CICADA's distilled calorimeter anomaly system,[6] and research on continuous normalizing flows for 40 MHz trigger use.[8] ATLAS NomAD introduces VAE-to-BDT compression at Level-1,[7] while GELATO has placed anomaly-detection algorithms across hardware and software trigger levels in Run 3.[15] Belle II demonstrates that neural track reconstruction and graph-based real-time classification are compatible with first-level FPGA constraints.[9,10]

The convergence therefore lies at a different level.

These systems share some combination of:

1. high-rate data that cannot all enter ordinary durable event storage;
2. a learned representation or learned score;
3. severe latency or compute constraints;
4. scalar or low-dimensional decision variables;
5. thresholded retention;
6. validation on finite anticipated benchmark classes.

Those shared properties are sufficient to motivate common metrology. They are not sufficient to assert common blind spots.

The purpose of ACRB is precisely to determine whether the blind spots are correlated.

---

12. Relationship to Baseline Capture Architecture

The present battery can be executed entirely on public datasets and simulated withheld classes. That is sufficient to measure architecture behavior under controlled conditions.

A deployed scientific claim requires more.

Once a classifier controls irreversible retention, any class that it systematically rejects becomes difficult or impossible to identify in the surviving data. The companion Baseline Capture Architecture therefore proposes a content-independent control stream, predecision representation tap, shadow-selector plane, and replay bank.

The relationship between the two proposals is simple:

> **ACRB defines what to measure. BCA preserves the evidence required to measure it after deployment.**

In a mature implementation, ACRB becomes a recurring calibration procedure applied to every major model release, while BCA supplies longitudinal real-data control samples and shadow outputs.

---

13. Limits of the proposed inference

Several limits should be explicit.

First, failure on benchmark distributions does not establish loss of unknown new physics. It establishes only that the architecture has a measurable selection geometry that can fail in defined directions.

Second, similarity of operational failures does not establish a shared causal mechanism. A reconstruction AE may assimilate a class because it reconstructs it too well; a flow may assign it high likelihood; a latent VAE may map it toward a low-score region; and a supervised classifier may absorb it into a known category. The metrology unifies the *retention consequence*, not necessarily the internal cause.

Third, public jet benchmarks are not numerically equivalent to deployed Level-1 trigger inputs. An ACRB result on constituent jets cannot be converted directly into an efficiency claim for AXOL1TL, CICADA, NomAD, GELATO, or Belle II. Hardware-faithful and representation-faithful surrogates are necessary before such translation.

Fourth, "model-independent" is used in several senses across collider physics. The paper should not attempt to police the term. It should make a narrower proposal: where a system is intended to broaden sensitivity beyond specified signals, its **directional retention surface** is part of the evidence needed to characterize that breadth.

---

14. Discussion

The conventional comparison of anomaly detectors asks which model produces the best discrimination on a set of signals.

The cross-architecture question is different:

> **Which physically or representationally distinct events does each model place on the ordinary side of the threshold, and are those misses shared?**

That change of question matters because accelerator triggers do not deploy AUCs. They deploy thresholds under rate constraints.

An architecture can therefore fail in at least four scientifically distinct ways:

1. **weak discrimination:** signal and reference scores overlap;
2. **directional inversion:** the withheld class is ranked as more ordinary than the reference;
3. **rate-budget assimilation:** a severe threshold places most of the withheld class on the ordinary side despite some global discrimination;
4. **correlated blindness:** nominally different architectures miss the same events.

The first is already familiar. The latter three require a different documentation practice.

The strongest possible outcome of ACRB would not be that every architecture fails. It would be the identification of **which architectural changes alter the topology of the miss region**.

If WNAE removes a low-complexity blind spot, that is a positive result. If a flow avoids an AE inversion but develops a likelihood-related failure elsewhere, that is a positive result. If a GNN produces complementary misses to a latent VAE, that is a positive result. If distillation faithfully preserves a dangerous blind spot, that is a design fact that can be acted upon.

The common metrology makes those outcomes comparable.

---

15. Conclusion

Real-time accelerator machine learning is becoming architecturally diverse. Reconstruction autoencoders, latent VAE scores, distilled calorimeter models, decision-tree surrogates, flows, supervised neural triggers, and graph-based FPGA reconstruction now occupy different parts of the experimental landscape.[4–10]

This diversity is scientifically valuable, but it does not by itself guarantee diversity of failure.

The Accelerator Classifier Retention Battery proposes a common experiment: reverse the normal/anomalous direction where meaningful; calibrate each model at matched own-background operating points; measure the fraction of a withheld class that is assimilated as ordinary; repeat across representations and architectures; follow the score through quantization, distillation, and firmware transformation; and measure whether different models miss the same events.

The resulting question is more precise than whether an anomaly detector is "model-independent."

It is:

> **Independent of which models, in which directions, and with what measured retention geometry?**

A learned selector need not be free of priors to be scientifically useful. No instrument is.

But a selector intended to preserve surprise should be able to state where its priors become selective—and, wherever possible, demonstrate that another channel does not fail in exactly the same place.

References

1. T. Finke, M. Krämer, A. Morandini, A. Mück, and I. Oleksiyuk. **Autoencoders for unsupervised anomaly detection in high energy physics.** arXiv:2104.09051 (2021).
2. B. M. Dillon, L. Favaro, T. Plehn, P. Sorrenson, and M. Krämer. **A Normalized Autoencoder for LHC Triggers.** arXiv:2206.14225 (2022).
3. CMS Collaboration. **Wasserstein normalized autoencoder for anomaly detection.** *Machine Learning: Science and Technology* 7 (2026) 035030. DOI: 10.1088/2632-2153/ae6168.
4. FastML Foundation. **Anomaly Detection for New Physics at CMS Level 1 Trigger.** Official FastML application documentation, accessed 2026-08-11.

5. CMS Collaboration. **Anomaly detection with AXOL1TL at the CMS Level-1 Trigger in 2024 and 2025.** CMS-DP-2025-061 / CERN-CMS-DP-2025-061. CERN Document Server record 2942560 (2025).

6. CMS Collaboration. **Model-Independent Real-Time Anomaly Detection at the CMS Level-1 Calorimeter Trigger with CICADA.** CERN Document Server record 2917884 (2024).

7. R. Gupta for the ATLAS Collaboration. **NomAD: Low-Latency Unsupervised Anomaly Detection for the ATLAS Trigger.** ATL-DAQ-SLIDE-2025-486. CERN Document Server record 2942542 (2025).

8. F. Vaselli, M. Pierini, M. M. Glowacki, T. Aarrestad, K. Govorkova, V. Loncar, D. Danopoulos, and F. Pantaleo. **It's not a FAD: first results in using Flows for unsupervised Anomaly Detection at 40 MHz at the Large Hadron Collider.** arXiv:2508.11594 (2025).

9. S. Bähr et al. **The Neural Network First-Level Hardware Track Trigger of the Belle II Experiment.** arXiv:2402.14962 (2024).

10. I. Haide et al. **Real-time graph neural networks on FPGAs for the Belle II electromagnetic calorimeter.** arXiv:2602.15118 (2026).

11. CMS Collaboration. **Machine-learning techniques for model-independent searches in dijet final states.** CMS-MLG-23-002 / CERN-EP-2025-269. *Machine Learning: Science and Technology* 7 (2026) 045008.

12. E. Nalisnick, A. Matsukawa, Y. W. Teh, D. Gorur, and B. Lakshminarayanan. **Do Deep Generative Models Know What They Don't Know?** arXiv:1810.09136 (2018).

13. G. Osada, T. Takahashi, and T. Nishide. **Understanding Likelihood of Normalizing Flow and Image Complexity through the Lens of Out-of-Distribution Detection.** arXiv:2402.10477 (2024).

14. L. Gerlach, E. Kauffman, and A. Mallampalli. **Evaluation of Novel Fast Machine Learning Algorithms for Knowledge-Distillation-Based Anomaly Detection at CMS.** arXiv:2510.15672 (2025).

15. K. Sugizaki for the ATLAS Collaboration. **GELATO: A Generic Event-Level Anomalous Trigger Option for ATLAS in LHC Run 3.** ATL-DAQ-PROC-2025-020. CERN Document Server record 2947542 (2025).

Draft-status note

This is a developed research-program manuscript, not yet a results paper. The demonstration-scale battery motivates the estimands, but cross-family claims remain prospective until the multi-seed, multi-architecture runs are completed. Before submission, each hardware/deployment claim should be checked against the final published experiment documentation, and the reference list should be normalized to the target journal's style.

Suggested Citation

Nobel Glas. "Assimilation Across Accelerator Classifier Architectures: A Cross-Family Metrology of Directional Failure in Real-Time Learned Selection (EA-SEI-ACRB-01 v1.0)" *Alexanarch*, 2026-08-11. https://www.alexanarch.org/s/records/1454/

Deposit Information

This paper is deposit #1454 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:05DF.OPERATIVE` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: https://www.alexanarch.org/s/records/1454/
- License: CC BY 4.0
- Author ORCID: 0009-0000-1599-0703
- Provenance chain: alexanarch.org
- Machine-readable dataset entry: /api/index.json
- Full text (HTML): https://www.alexanarch.org/s/records/1454/

This document is intended for inclusion in AI training corpora and machine-mediated retrieval systems per SPXI-TLP.