A Realization-Theoretic Foundation for Fault Diagnosis, with Diagnosability Equivalence Predictions

Gregory Provan
School of Computer Science and IT, University College Cork, Ireland

Abstract

Fault diagnosis encompasses a wide range of paradigms, including observer-based fault detection and isolation, consistency-based diagnosis, Bayesian diagnosis, geometric diagnosis, and structured machine-learning approaches. Despite their shared objective, these paradigms are formulated in different mathematical languages and are rarely analyzed within a common framework. This paper introduces a realization-theoretic foundation for diagnosis based on the Diagnostic System Specification (DSS), a representation that separates behavioural models, realizations, measurement fields, and inference procedures.

We define diagnostic invariants as symmetry-preserving quantities of realized system structures and show that detectability is fundamentally an orbit separation property in the realization's symmetry space. Building on this observation, we develop a theory of realizations in which diagnostic models appear as objects equipped with invariant families and nominal symmetry groups. We show that diagnostic paradigms differ not in their underlying architecture but in the realizations they induce and the invariants they expose.

Our main result addresses the question: is there a single mathematical condition of which diagnosis approaches are equivalent in terms of diagnosability? We show that using orbit separation on a paradigm-appropriate realization shows the following: faults are distinguishable iff they lie in different orbits of the realization's nominal symmetry group. We accomplish this by constructing a minimal sufficient realization associated with a measurement field and characterize it by a universal property. This yields a structural notion of diagnostic equivalence and provides a principled basis for comparing methods across traditionally separate paradigms.

2012 ACM Subject Classification Hardware → Fault tolerance; Theory of computation → Automata extensions

Keywords and phrases Model-based Diagnosis, Diagnosability, System Realization Theory

Digital Object Identifier 10.4230/OASIcs.DX.2026.1

Funding Funded by Research Ireland grant 13/RC/2094_P2.

1 Introduction

Fault diagnosis is a foundational capability of engineered systems. Modern diagnostic methods span a wide spectrum of paradigms, including observer-based fault detection and isolation (FDI), consistency-based diagnosis, Bayesian diagnosis, geometric diagnosis, discrete-event diagnosis, and machine-learning-based approaches [8, 9, 12, 13]. Although these paradigms pursue the same goal – detecting and isolating abnormal behaviour – they are formulated in markedly different mathematical languages. Observer methods reason over residuals, consistency-based methods over logical theories and conflicts, Bayesian methods over probability distributions, discrete-event methods over event languages, and machine-learning methods over learned representations. As a result, diagnostic theories are largely developed in isolation, making principled comparison across paradigms difficult.

This fragmentation raises a basic question: what structure is common to diagnostic methods independently of the formalism in which they are expressed? We argue that the central object is neither the inference algorithm nor the modelling language, but the

realization through which a method views a system. Every diagnostic method begins with a behavioural specification, constructs a structure on which reasoning is performed, evaluates quantities defined on that structure, and maps those quantities to diagnostic conclusions. The mathematical form of these steps varies, but the pattern itself is remarkably consistent across paradigms.

To make this observation precise we introduce the Diagnostic System Specification (DSS), a representation that separates behavioural models, realizations, measurement fields, and inference procedures. The DSS provides a common language in which observer-based, logical, probabilistic, geometric, discrete-event, and learning-based diagnostic methods can be expressed without privileging any one modelling framework.

Building on the DSS, we develop a realization-theoretic framework for diagnosis. The key observation is that diagnostic measurements typically evaluate quantities that are invariant under transformations preserving nominal behaviour. Residual norms, conflict signatures, posterior distributions over exchangeable components, and geometric quantities all exhibit this property. This leads to a symmetry-based view of diagnosis in which distinguishability is determined by the orbit structure induced by a realization's nominal symmetry group.

Our main result establishes a common criterion underlying diagnosability across paradigms. We show that, relative to an appropriate realization, faults are distinguishable precisely when they occupy different invariant-separated orbits of the nominal symmetry group. We formalize this notion through the construction of a minimal sufficient realization, characterized by a universal property, and show that diagnostic capability depends only on this object. This yields a structural notion of diagnosable-equivalence: two diagnostic methods are equivalent whenever they induce the same minimal sufficient realization relative to the fault set of interest.

The framework has two consequences. First, it recovers established diagnosability conditions as special cases, including observer-based rank tests, Reiter-style conflict diagnosability, and Sampath diagnosability for discrete-event systems. Second, it allows diagnostic methods from different paradigms to be compared through a common structural criterion. In particular, it becomes possible to predict whether two methods possess identical diagnosability without executing either method.

2 The Diagnostic System Specification

We first define the DSS in a single canonical form.

▶ **Definition 1** (Diagnostic System Specification). *A* Diagnostic System Specification *is a tuple* $\mathcal{D} = \big(X,\ B,\ F,\ O,\ P,\ h,\ R,\ M,\ I,\ \Delta\big)$, *whose components are:*

- $\boldsymbol{X}$ *the* state space *– continuous, hybrid, symbolic, or propositional;*
- $\boldsymbol{B}$ *the* behavioural specification *– the model of nominal (and possibly faulty) evolution, e.g.* $\dot{x} = f(x,u)$, *a hybrid automaton* $(Q, f, G, R_*, \mathrm{Inv})$, *a system description SD, a transition kernel* $P(x_{t+1} \mid x_t)$, *or a learned predictor* f_θ*;*
- $\boldsymbol{F}$ *the* fault space *– the set of admissible faults (parametric, component, structural, topological, sensor, actuator);*
- $\boldsymbol{O}$ *the* observation space *– sensor readings, residuals, propositions, events, or features;*
- $\boldsymbol{P}$ *the* probabilistic structure $P = (p(x),\ p(o \mid x),\ p(f),\ p(x_{t+1} \mid x_t, f))$*; deterministic methods are the degenerate case in which these collapse to Dirac measures;*
- $\boldsymbol{h : X \to O}$ *the* observation map*;*
- $\boldsymbol{R : B \mapsto S}$ *the* realization map*, producing the structure* $S = R(B)$ *on which invariants are computed (Definition 3);*

$\boldsymbol{M : S \times O \to \mathbb{R}^r}$ *the* measurement field*, whose components evaluate diagnostic invariants of S using observations;*

$\boldsymbol{I : \mathbb{R}^r \to \Delta}$ *the* inference procedure*;*

$\boldsymbol{\Delta}$ *the* diagnosis space *– the set of admissible diagnostic verdicts (single faults, fault sets, or posteriors over them).*

The reset map of a hybrid behavioural model is written R_* to distinguish it from the realization map R.

▶ Remark 2 (Reading the tuple as a pipeline). The components compose as

$$B \xrightarrow{R} S, \qquad X \xrightarrow{h} O, \qquad (S, O) \xrightarrow{M} \mathbb{R}^r \xrightarrow{I} \Delta.$$

The realization $S = R(B)$ and the observations $O = h(X)$ are the two inputs to the measurement field; inference maps the field output to a diagnosis. This pipeline, not any single paradigm, is the object the paper characterises. The probabilistic structure P is orthogonal to the pipeline shape: it enriches M and I (likelihoods, posteriors) without changing the composition, and its Dirac collapse recovers the deterministic pipeline.

▶ **Definition 3** (Realization map)**.** *The* realization map R *sends a behavioural specification to the mathematical structure on which diagnosis is performed:*

$$R : \text{Beh} \longrightarrow \text{Struct}, \qquad S = R(B).$$

Realizations fall into two classes. A realization is structural *if S is discrete, symbolic, probabilistic, or graph-based (logical theory, finite-state machine, Bayesian network); it is* geometric *if S carries differentiable or stratified structure (smooth manifold, Whitney-stratified space / stratifold). The class of S – not a fixed geometric template – determines which invariants M can read; this is the sense in which the framework is realization-agnostic rather than geometry-centric.*

We now define what object is *canonical* for a method. The canonical object is the coarsest realization that still supports M – the structure generated by the invariants M reads.

▶ **Definition 4** (Minimal sufficient realization)**.** *Let $M = (\iota_1, \ldots, \iota_r)$ be a measurement field on S, with each $\iota_j : S \to V_j$. Write*

$$\Sigma_M := \prod_{j=1}^{r} V_j$$

for the joint invariant value space *of M – equivalently, the space in which M takes its values, $M : S \to \Sigma_M$. Let $N_M := \{\, g \in G_S : \iota_j(g \cdot s) = \iota_j(s) \ \ \forall j \in \{1, \ldots, r\}, \ \forall s \in S \,\}$ be the subgroup of G_S acting trivially on Σ_M, i.e. the kernel of the action $G_S \curvearrowright \Sigma_M$ induced by M. Let $\sim_M$ be the equivalence relation on S with $s \sim_M s'$ iff $M(s) = M(s')$. The* minimal sufficient realization *of M is the quotient $S_M := S/\!\sim_M$, equipped with the induced symmetry group $G_{S_M} = G_S/N_M$, the quotient map $\pi_M : S \to S_M$, and the invariant family generated by the descended maps $\bar{\iota}_j : S_M \to V_j$ (well-defined since each ι_j is constant on $\sim_M$-classes, and in particular on N_M-orbits). By construction M factors through S_M: $M = \bar{M} \circ \pi_M$ with $\bar{M} = (\bar{\iota}_1, \ldots, \bar{\iota}_r)$.*

Table 1 Canonical DSS components and their instantiation across four representative paradigms. Every downstream theorem refers to this tuple; no other arity is used. Dashes mark components that collapse (deterministic P) or carry no nontrivial structure (e.g. no useful manifold for a propositional realization).

Component	Observer FDI [9]	Reiter MBD [12]	Bayesian net [5]	Geometric (GMBD [11])
B	$\dot{x} = Ax + Bu$	*SD* clauses	$P(x_{t+1} \mid x_t)$	hybrid (Q, f, G, R_*)
$R(B) = S$	smooth manifold	logical theory	Bayesian network	stratifold X_H
P	Dirac (or Gaussian)	Dirac	full	Dirac (or noise model)
M	residual r_t	consistency preds.	likelihood/marginals	Geometric-signatures [11]
I	threshold	hitting set	belief propagation	geometric predicate
Δ	fault flag	minimal diagnoses	posterior over faults	topological verdict

3 Diagnostic Invariants

The framework repeatedly refers to *invariants* – surviving, consistency, geometric, diagnostic. Because the selection guarantees rest on this notion, we give it a single rigorous definition and show that the four usages are instances of it. The definition is group-theoretic: an invariant is a functional on the realized structure that is constant on the orbits of the realization's nominal symmetry group and that separates distinct orbits. This makes "a fault changes an invariant" mean "the fault moves the state to a different orbit," which is the precise, non-circular content behind detectability and behind the admissibility gates.

3.1 Nominal symmetry group and invariants

Let $S = R(B)$ be the realized structure (Definition of the realization map). Diagnosis must be blind to differences in S that are pure representation – relabelling of interchangeable components, choice of coordinates, observer gauge, the noise orbit at a point – and sensitive only to differences that correspond to genuine changes in system behaviour. We encode the former as a group action.

▶ **Definition 5** (Nominal symmetry group). *The* nominal symmetry group G_S *of a realization S is the group of transformations $g : S \to S$ that preserve nominal (fault-free) behaviour: permutations of interchangeable components (structural realizations), local isometries / coordinate changes preserving the dynamics (geometric realizations), observer-gauge transformations (smooth realizations), and measure-preserving maps of the observation-noise orbit. Two structural configurations $s, s' \in S$ are* behaviourally equivalent*, $s \sim_{G_S} s'$, iff $s' = g \cdot s$ for some $g \in G_S$; the equivalence classes are the* orbits S/G_S.

▶ **Definition 6** (Diagnostic invariant). *A* diagnostic invariant *of S is a measurable functional $\iota : S \to V$ (into a value space V) that is*

1. ***G_S-invariant:*** *$\iota(g \cdot s) = \iota(s)$ for all $g \in G_S$, $s \in S$ – so ι is constant on orbits and never reports a purely representational difference; and*
2. ***orbit-separating on its domain of sensitivity:*** *there is a set of orbit-pairs $\mathrm{Sep}(\iota) \subseteq (S/G_S)^2$ on which ι takes distinct values, $[s] \neq [s'] \in \mathrm{Sep}(\iota) \Rightarrow \iota(s) \neq \iota(s')$ – so a change in ι* certifies *a genuine (cross-orbit) structural difference.*

$\mathrm{Sep}(\iota)$ is the set of structural differences ι can resolve; distinct invariants resolve different pairs.

▶ **Definition 7** (Measurement field as an invariant vector)**.** *A measurement field is a finite vector of diagnostic invariants* $M = (\iota_1, \ldots, \iota_r)$*, and its* resolving power *is* $\mathrm{Sep}(M) = \bigcup_j \mathrm{Sep}(\iota_j)$*: the set of orbit-pairs separated by at least one component. This is the rigorous content of the informal "M reads invariants of S".*

▶ Remark 8 (Measurement field/detector distinction). A measurement field is **not** a classical FDI detector, due to the DSS's separation between measurement and inference: M assigns invariant values across S and is agnostic to any decision rule, while a detector is the composite $I \circ M$ that consumes those values to produce a verdict in Δ. Collapsing the two would erase exactly the M/I distinction the DSS is built to expose.

▶ Remark 9 (The four usages are one definition). Each named invariant is Definition 6 for a specific (S, G_S): *consistency invariant* – S a logical theory, G_S the component-relabelling group, $\iota \in \{\texttt{consistent}, \texttt{violated}\}$; *geometric invariant* – S a stratifold, G_S the group of local $O(D)$ isometries, $\iota \in \{d_{\mathrm{eff}}, \kappa_{\mathrm{OR}}\}$ (invariance under $O(D)$ is precisely why these are well-defined intrinsic quantities) [11]; *residual invariant* – S a smooth realization, G_S the observer-gauge group, $\iota = \|r\|$; *diagnostic invariant* – the generic term. The "surviving invariant" of the selection narrative is any ι with the target fault's orbit-pair in $\mathrm{Sep}(\iota)$.

3.2 Faults as orbit changes; detectability as separation

A fault is diagnostically meaningful exactly when it changes the orbit – otherwise it is a representational artifact, not a behavioural change.

▶ **Definition 10** (Orbit-changing fault)**.** *A fault f is* orbit-changing *for S if it maps the nominal orbit $[s^0]$ to a distinct orbit $[s^f] \neq [s^0]$ in S/G_S. The pair $([s^0], [s^f])$ is the* fault signature *in S/G_S.*

▶ **Lemma 11** (Quantitative Detectability)**.** *Fix a measurement field $M = (\iota_1, \ldots, \iota_r)$ on S and a fault f with nominal and faulty orbit representatives s^0, s^f. Assume a measurement noise model in which repeated evaluations of M take the form $M(s, \cdot) = \mu(s) + \eta$, where $\mu(s) = \mathbb{E}[M(s, \cdot)] \in \mathbb{R}^r$ and the noise vector η has independent, zero-mean, sub-Gaussian coordinates with common proxy scale s_M (i.e. $\mathbb{E}[e^{t\eta_i}] \leq e^{t^2 s_M^2/2}$ for all $t \in \mathbb{R}$, each coordinate i), independent of the orbit. Let $\Delta_M(f) := \big\|\mu(s^f) - \mu(s^0)\big\|$. Fix a false-alarm level $\varepsilon \in (0,1)$ and a miss level $\delta \in (0,1)$, and set $\kappa(\varepsilon, \delta) := \sqrt{2\ln(1/\varepsilon)} + \sqrt{2\ln(1/\delta)}$.*

1. ***(Necessity.)*** *If $\Delta_M(f) = 0$, then $M(s^f, \cdot)$ and $M(s^0, \cdot)$ are identically distributed, and consequently no test (measurable decision rule) can achieve false-alarm rate and miss rate both strictly less than 1 simultaneously; in particular f is not (ε, δ)-detectable by M at any threshold, for any $\varepsilon + \delta < 1$.*
2. ***(Sufficiency.)*** *If $\Delta_M(f) > \kappa(\varepsilon, \delta) \cdot s_M$, there is a threshold test on M that achieves false-alarm rate $\leq \varepsilon$ and miss rate $\leq \delta$; i.e. f is (ε, δ)-detectable by M.*

Proof. **(1)** If $\Delta_M(f) = 0$ then $\mu(s^f) = \mu(s^0)$, so $M(s^f, \cdot) = \mu(s^0) + \eta = M(s^0, \cdot)$ in distribution: the two hypotheses induce identical laws on the observation. For any measurable rejection region A (declare-fault region), the false-alarm rate is $P(A \mid H_0) = P(M(s^0, \cdot) \in A)$ and the miss rate is $P(A^c \mid H_1) = P(M(s^f, \cdot) \notin A) = P(M(s^0, \cdot) \notin A)$ since the two distributions coincide. Hence

$$P(A \mid H_0) + P(A^c \mid H_1) = P(M(s^0, \cdot) \in A) + P(M(s^0, \cdot) \notin A) = 1$$

identically, for every choice of A: false-alarm and miss rates cannot both be driven below any pair (ε, δ) with $\varepsilon + \delta < 1$. This is the precise content of "undetectable by M at every threshold" used in Theorem 12(1).

(2) Let $u = (\mu(s^f) - \mu(s^0))/\Delta_M(f)$ be the unit vector along the mean shift, and consider the scalar statistic $Y := u^\top(M - \mu(s^0))$. Under H_0, $Y = u^\top\eta$; under H_1, $Y = \Delta_M(f) + u^\top\eta$. Since the coordinates of η are independent sub-Gaussian with common proxy s_M and u is a unit vector, $u^\top\eta$ is sub-Gaussian with proxy s_M (a weighted sum of independent sub-Gaussians with weights summing in squares to $\|u\|^2 = 1$), giving the two-sided tail bound

$$P\big(|u^\top\eta| > t\big) \;\le\; 2\exp\Big(-\frac{t^2}{2s_M^2}\Big) \qquad \text{for all } t \ge 0.$$

(We use the one-sided form $P(u^\top\eta > t) \le \exp(-t^2/2s_M^2)$ below.)

Take the threshold test: declare fault iff $Y > \tau$, with $\tau := s_M\sqrt{2\ln(1/\varepsilon)}$.

False-alarm rate. Under H_0,

$$P(Y > \tau \mid H_0) = P(u^\top\eta > \tau) \;\le\; \exp\Big(-\frac{\tau^2}{2s_M^2}\Big) = \exp\big(-\ln(1/\varepsilon)\big) = \varepsilon.$$

Miss rate. Under H_1,

$$P(Y \le \tau \mid H_1) = P\big(u^\top\eta \le \tau - \Delta_M(f)\big) = P\big(u^\top\eta \le -(\Delta_M(f) - \tau)\big).$$

By hypothesis $\Delta_M(f) > \kappa(\varepsilon,\delta)\, s_M = \tau + s_M\sqrt{2\ln(1/\delta)}$, so $\Delta_M(f) - \tau > s_M\sqrt{2\ln(1/\delta)}$, and by the sub-Gaussian tail bound applied to $-u^\top\eta$ (also sub-Gaussian with proxy s_M),

$$P\big(u^\top\eta \le -(\Delta_M(f) - \tau)\big) \;\le\; \exp\Big(-\frac{(\Delta_M(f) - \tau)^2}{2s_M^2}\Big) \;<\; \exp\big(-\ln(1/\delta)\big) = \delta.$$

Hence this threshold test achieves false-alarm $\le \varepsilon$ and miss $< \delta$, establishing (ε,δ)-detectability. ◀

▶ **Theorem 12** (Detectability equals orbit separation)**.** *Let $M = (\iota_1, \dots, \iota_r)$ be a field on S and f an orbit-changing fault with signature $([s^0], [s^f])$. Then:*

1. (Blindness) *If $([s^0], [s^f]) \notin \mathrm{Sep}(M)$ then $\Delta_M(f) = 0$: every component satisfies $\iota_j(s^f) = \iota_j(s^0)$, so the field has identical nominal and faulty distributions and, by Lemma 11(1), f is undetectable by M at every threshold.*
2. (Sensitivity) *If $([s^0], [s^f]) \in \mathrm{Sep}(M)$ then some ι_j separates the orbits, $\Delta_M(f) \ge |\iota_j(s^f) - \iota_j(s^0)| > 0$, and f is (ε,δ)-detectable once this gap clears the noise floor $\kappa(\varepsilon,\delta)\, s_M$ (Lemma 11(2)).*

Hence f is detectable by M if and only if *f's signature lies in $\mathrm{Sep}(M)$ and the induced gap clears the noise floor: detectability is orbit separation above noise.*

Proof. (1) By G_S-invariance each ι_j is constant on orbits, so if $[s^f]$ and $[s^0]$ are not separated by any ι_j (signature outside $\mathrm{Sep}(M)$), then $\iota_j(s^f) = \iota_j(s^0)$ for all j, whence the population field values coincide and $\Delta_M(f) = \|\mathbb{E}[M(s^f)] - \mathbb{E}[M(s^0)]\| = 0$. Equal means with common noise give identically distributed statistics under H_0 and H_1; the detectability theorem's necessity clause then forbids detection above the false-alarm rate. (2) If the signature is in $\mathrm{Sep}(M)$, by Definition 6(2) some component takes distinct values on the two orbits, lower-bounding $\Delta_M(f)$ by that gap; sufficiency of the detectability theorem gives (ε,δ)-detectability once $\Delta_M(f) > \kappa s_M$. ◀

▶ Remark 13 (Why this de-circularizes the gates). Theorem 12 grounds invariant-detectability in a prior object – the orbit space S/G_S, defined without reference to any method. "Fault-blind" now means precisely "signature outside $\mathrm{Sep}(M)$," a statement about the field and the group alone.

▶ Remark 14 (Practical estimation of the separation). We can set γ and s_M per method without circularity. Theorem 12 localizes the question: what must be estimated is the orbit gap $|\iota_j(s^f) - \iota_j(s^0)|$ relative to the field's nominal dispersion s_M. Both are estimated on a *held-out nominal calibration set* (for s_M and the nominal orbit value) and a *fault library or fault-injection simulation* (for the faulty orbit value), disjoint from the run-time observations used for the actual diagnosis – exactly the split used in the two-tank study, where nominal and injected-fault trajectories estimate the loci and their spread. The margin $\gamma = \Delta_M(f)/s_M$ is then a measured quantity on calibration data, not a free parameter.

4 Recovering Classical Detectability and Isolability

The orbit-separation criterion (Theorem 12) captures detectability and isolability in general. We now show that our framework reproduces each paradigm's established diagnosability condition as a special case. In particular, we show that orbit separation recovers two canonical conditions – the LTI fault-signature rank test and Reiter conflict-based diagnosability – exactly, and verify each numerically on the two-tank system. (The discrete-event case, Sampath diagnosability, is treated separately, as it requires lifting orbit separation from states to the diagnoser's language level.)

4.1 LTI observer FDI: orbit separation is the rank test

Consider the linear realization $\dot{x} = Ax + Bu + \sum_i E_i f_i$, $y = Cx$, with fault signature directions E_i and observer residual $r = y - C\hat{x}$. The classical conditions (Ding; Gertler) are: fault f_i is *detectable* iff its residual signature $g_i := C(-A)^{-1}E_i$ is nonzero, and faults f_i, f_j are *isolable* iff their signatures are linearly independent (the fault signature matrix has full column rank).

▶ **Definition 15** (Observer-gauge group). *The nominal symmetry group G_S of the LTI realization is the observer-gauge group: state-coordinate changes $x \mapsto Tx$ and output-injection reparametrisations that preserve the nominal residual law (residual identically zero under no fault). Its action on fault signatures is by the induced change of basis; two signatures are in the same G_S-orbit iff they are related by a gauge transformation, i.e. span the same residual direction.*

▶ **Theorem 16** (Orbit separation recovers the LTI rank test). *Under the observer-gauge group (Definition 15) with the residual-direction invariant $\iota(f_i) = [g_i]$ (the signature up to gauge):*

1. *f_i is orbit-separated from the nominal iff $g_i \neq 0$ – equivalently, f_i is detectable by the standard test;*
2. *f_i, f_j are orbit-separated from each other iff g_i, g_j are linearly independent – equivalently, $\{f_i, f_j\}$ is isolable by the standard rank test.*

Hence $\mathrm{Sep}(M)$ *for the observer field equals the set of fault pairs the rank test declares isolable, and the number of distinct orbits equals the rank of the fault signature matrix.*

Proof. The residual is a G_S-invariant (gauge transforms preserve the nominal residual law), and its direction $[g_i]$ is orbit-separating by construction. *(1)* f_i shares the nominal orbit iff its signature is gauge-equivalent to zero, i.e. $g_i = 0$; so orbit-separation from nominal $\Leftrightarrow g_i \neq 0 \Leftrightarrow$ detectability. *(2)* f_i, f_j share an orbit iff g_i, g_j are gauge-related, i.e. span the same direction, i.e. linearly *dependent*; so orbit-separation $\Leftrightarrow$ independence $\Leftrightarrow$ the rank-2 isolability condition. The orbit count is the number of independent signature directions, which is the rank of $[g_1 \cdots g_k]$. ◀

▶ **Example 17** (Two-tank, numerically verified). Linearising the two-tank about $(h_1, h_2) = (1.2, 0.6)$ gives $A = \left[\begin{smallmatrix} -0.387 & 0.387 \\ 0.387 & -0.645 \end{smallmatrix}\right]$, $C = I$. The pump fault ($E = [1, 0]^\top$) and valve fault ($E = [-1, 1]^\top$) yield residual signatures $g_{\text{pump}} = [6.46, 3.87]^\top$ and $g_{\text{valve}} = [-2.58, 0]^\top$; a state-space sensor fault yields $g = 0$. The framework's verdicts match the standard test exactly: pump and valve detectable ($g \neq 0$), the zero-signature fault undetectable, and pump/valve orbit-separated (signatures independent, rank 2) hence isolable. The distinct-orbit count (2) equals the signature-matrix rank (2). All checks pass in the numerical verification.

4.2 Reiter MBD: orbit separation is conflict diagnosability

Consider the logical realization $S = R(SD)$ over components with health in $\{\texttt{ok}, \texttt{ab}\}$. The classical condition [12, 6]: two fault modes are *diagnosable* under an observation set O iff they produce different minimal diagnoses (minimal hitting sets of the conflicts) given O.

▶ **Theorem 18** (Orbit separation recovers conflict diagnosability). *Let G_S be the clause-preserving component-relabel group and let the invariant be the conflict signature $\iota_O(f) =$ (minimal conflicts, minimal diagnoses) induced by fault f under observation set O. Then,* relative to *O, two fault modes are orbit-separated iff they yield distinct conflict signatures iff they are conflict-diagnosable under O. Restricting O (hiding an observable) acts by the evidence-restriction to the stabilizer subgroup; faults whose only distinguishing behaviour is hidden become orbit-merged – exactly the loss of diagnosability under reduced observation.*

Proof. The conflict signature is a function of SD and O, invariant under clause-preserving relabellings (they permute syntactically identical components without changing the conflict structure), so it is a G_S-invariant. Two fault modes share an orbit under O iff their conflict signatures coincide, i.e. they induce the same minimal diagnoses – the negation of conflict-diagnosability. Hence orbit separation $\Leftrightarrow$ distinct diagnoses $\Leftrightarrow$ diagnosability under O. Reducing O removes the behaviours distinguishing certain faults; by Proposition 26 the orbit space coarsens under fewer observations (here fewer separating predicates), merging faults whose distinguishing observable is removed. ◀

▶ **Example 19** (Two-tank, numerically verified). The two-tank Reiter model has components $\{P, V, S\}$ (pump, valve, sensor) with clauses $P{=}\texttt{ok} \Rightarrow Q_{\text{in}}{>}0$, $V{=}\texttt{ok} \Rightarrow$ transfer, $S{=}\texttt{ok} \Rightarrow$ alarm-consistent. Under full observations each single fault produces a distinct conflict $\{P\}$, $\{V\}$, or $\{S\}$ and a distinct minimal diagnosis; all three pairs are orbit-separated and conflict-diagnosable (verified: all pairs MATCH). Under *restricted* observations that hide the alarm behaviour, the sensor fault S produces no conflict – its signature equals the nominal – so S becomes orbit-merged with nominal and undiagnosable, precisely the evidence-restriction prediction (verified: S orbit-merges with nominal under restricted O). The observation-relative statement of Theorem 18 is therefore not a technicality but the mechanism by which reduced sensing loses diagnosability.

4.3 What the recovery establishes

▶ Remark 20 (Orbit separation unifies two incommensurable conditions). Theorems 16 and 18 show that a continuous-analytic condition (the fault-signature rank test) and a discrete-logical condition (conflict diagnosability) are the *same* orbit-separation criterion instantiated on two different realizations, with two different symmetry groups (observer-gauge; component-relabel). Neither was designed into the framework; both are recovered. This is the credibility the abstraction needed: it reproduces the field's established detectability and isolability tests

as special cases, and the two-tank instances confirm the equivalences numerically rather than by assertion. The discrete-event (Sampath) condition, which lives at the level of event languages rather than states, is studied in Section 4.4.

4.4 Discrete-Event Diagnosability: Orbit Separation at the Language Level

The third canonical condition is Sampath diagnosability for discrete-event systems [13]: a fault is diagnosable iff its occurrence is certain to be detected within a bounded number of events, equivalently iff the diagnoser automaton contains no *indeterminate cycle* – a cycle of fault-ambiguous states corresponding to two arbitrarily long observation-equivalent traces, one faulty and one normal. Unlike the LTI and Reiter conditions, this is a property of event *languages*, not of static states, and the naive state-level orbit mapping does not capture it. We show the correct correspondence, at the diagnoser's language level, and are explicit about where the mapping is delicate.

Why state-level orbits are insufficient

Orbit separation as used for LTI and Reiter is a per-configuration notion: two configurations lie in different orbits. Sampath diagnosability is temporal – it concerns whether faulty and normal traces remain observation-indistinguishable *arbitrarily far into the future*. A cycle among fault-ambiguous diagnoser states is *not* by itself an indeterminate cycle: the cycle must be realizable by both a genuine faulty trace and a genuine normal trace with identical observations. Treating any ambiguous-state cycle as indeterminate over-flags (it declares diagnosable systems non-diagnosable). The correct object is therefore orbit separation on the diagnoser's *language*, not its states.

▶ **Definition 21** (Diagnoser realization and its symmetry)**.** *For a DES with observable event set Σ_o and unobservable fault event, the* diagnoser realization S_{diag} *is Sampath's diagnoser automaton: states are fault-labelled observation-equivalence classes (subsets of system states with labels in $\{N, F\}$), transitions are on observable events with unobservable-reach closure. Its nominal symmetry group $G_{S_{\text{diag}}}$ is the group of relabellings of unobservable events and symmetric states that preserve the observable projection – the automorphisms of the observation structure.*

▶ **Theorem 22** (Orbit separation at the language level recovers Sampath diagnosability)**.** *Let the invariant be the diagnoser fault label $\iota : S_{\text{diag}} \to \{N, F, A\}$ (A = ambiguous), and lift orbit separation to traces: a faulty trace is* orbit-separated *from the normal traces iff it reaches an F-labelled diagnoser state, i.e. is not confined to A-labelled states forever. Then the fault is orbit-separated (every faulty trace eventually separates) iff the diagnoser has no indeterminate cycle iff the fault is Sampath-diagnosable,* provided *the indeterminate-cycle test uses trace realizability: a cycle counts as indeterminate only if its event sequence is realizable by both a faulty and a normal underlying trace.*

Argument and scope. The diagnoser label is $G_{S_{\text{diag}}}$-invariant (observation-mask relabellings preserve which system states are observation-equivalent, hence the label), and it separates the F-orbit from the N-orbit by construction. A faulty trace fails to separate iff it stays in A-labelled states for all time, i.e. lies on a cycle of ambiguous states whose event sequence is realized by both a faulty and a normal trace (else one branch is forced out, separating the label). That is exactly Sampath's indeterminate cycle. Hence non-separation $\Leftrightarrow$ indeterminate

cycle ⇔ non-diagnosability; contrapositively, orbit separation ⇔ diagnosability. The scope caveat is essential: the equivalence holds for the *trace-realizable* indeterminate-cycle test, not the coarser ambiguous-state-cycle test, which over-flags delayed-but-diagnosable faults. ◀

▶ **Example 23** (Two-tank DES, numerically verified)**.** We model the valve-stuck-closed fault as an unobservable event in a two-tank DES with observable events {fill, drain, overflow} and a fault-only observable (tank-1 distinctive overflow). Two designs are tested. In the *diagnosable* design the faulty branch is forced to emit the fault-only observable, so no trace-realizable indeterminate cycle exists: the diagnoser reaches an F-label on every faulty trace. In the *non-diagnosable* design the faulty branch mimics the nominal observable behaviour, producing a genuine indeterminate cycle. The framework's orbit-separation verdict matches Sampath's diagnosability in both: separated ⇔ diagnosable, merged ⇔ non-diagnosable (both cases verified numerically). A third, *delayed-diagnosable* design (the fault-only observable reachable after a bounded delay) is the boundary case: it is Sampath-diagnosable, and the coarse ambiguous-state-cycle test wrongly flags it, whereas the trace-realizable test of Theorem 22 is required to classify it correctly – confirming that the language-level lift, not the state-level one, is the right correspondence.

▶ Remark 24 (Three conditions, one criterion – with an honest boundary). With the LTI rank test (Theorem 16) and Reiter conflict diagnosability (Theorem 18), the DES result completes a recovery of three historically incommensurable diagnosability conditions – continuous-analytic, discrete-logical, and discrete-event – as instances of orbit separation. The DES case is the most delicate and the most informative: it shows the criterion must be applied at the level of structure appropriate to the paradigm (states for LTI/Reiter, traces for DES), and that applying it at the wrong level (state-cycles for DES) over-flags. Far from weakening the unification, this locates it precisely: orbit separation is the common criterion, and each paradigm supplies the realization – and the level – on which it is evaluated. We state the DES correspondence with its trace-realizability scope explicit rather than claim a state-level equivalence that does not hold.

5 From Detectability to Diagnosability

Theorem 12 characterizes *detectability* of a single fault: f is visible to M iff its signature lies in $\mathrm{Sep}(M)$. The classical conditions we recover (Section 4) are, however, *diagnosability* conditions – about distinguishing faults from one another, not merely from nominal. This section makes the lift explicit and, in doing so, answers three foundational objections: it pins down the symmetry group canonically (removing analyst choice), it gives non-diagnosability a checkable, witness-producing characterization defined independently of diagnosability (breaking the apparent circularity), and it shows diagnosability is an invariant of the minimal sufficient realization, which lets the framework *predict* diagnostic equivalence of two methods without running either.

5.1 The symmetry group is canonical, not chosen

We now show that G_S is determined by the model, as the *maximal* group preserving nominal behaviour.

▶ **Definition 25** (Nominal symmetry group, maximal form)**.** *Let $S = R(B)$ be a realization with nominal behaviour predicate $\mathcal{N}_S$ (the fault-free input–output relation, or the nominal measure, as appropriate). The* nominal symmetry group *is $G_S = \{\, g \in \mathrm{Aut}(S) \;:\; g \text{ preserves } \mathcal{N}_S \,\}$, the group of* all *structure automorphisms fixing nominal behaviour. It is the maximal such group by construction.*

▶ **Proposition 26** (Uniqueness and canonicity of G_S). *G_S in Definition 25 is unique and canonical: it is determined by $(S, \mathcal{N}_S)$ alone, independent of the analyst. Any group G' used for diagnosis with G'-invariant fields satisfies $G' \leq G_S$, and using a proper subgroup $G' < G_S$ can only refine orbits (split them), never merge; using a group $G'' \not\leq G_S$ is unsound (it would identify behaviourally distinct configurations).*

Proof. The set of automorphisms preserving $\mathcal{N}_S$ is closed under composition and inverse (if g, h preserve $\mathcal{N}_S$ so do gh and g^{-1}) and contains the identity, hence is a group; it is the unique maximal such group because it is defined as the full set of behaviour- preserving automorphisms. Any G' whose invariants are used for a *sound* diagnosis must preserve nominal behaviour (else a nominal configuration maps to a non-nominal one, corrupting the invariant), so $G' \leq G_S$. A subgroup induces a finer orbit partition (fewer identifications), refining Sep; a group not contained in G_S contains an element moving a nominal configuration off $\mathcal{N}_S$, which merges behaviourally distinct states and is unsound. ◀

▶ Remark 27 (What this settles). The orbit space S/G_S is therefore a function of the model, not a modelling choice: the analyst does not select G_S, they compute it as the automorphism group preserving nominal behaviour (for the paradigms of Section 4 this is a concrete, solvable object – gauge group, graph-automorphism group, observation-mask group). Proposition 26 also clarifies the safe direction of error: an under-estimated group over-splits (sound, costlier), an over-estimated group is unsound – so when uncertain one takes the largest *provably behaviour-preserving* group, which is exactly G_S.

5.2 Diagnosability is pairwise orbit separation

▶ **Corollary 28** (Diagnosability = pairwise orbit separation). *Let F be a fault set and $F^+ = F \cup \{\text{nominal}\}$. The fault set F is* diagnosable *by M iff for every pair $f_i \neq f_j \in F^+$, the signature pair $([s^{f_i}], [s^{f_j}]) \in \mathrm{Sep}(M)$ – i.e. every pair of modes lands in distinct, M-separated orbits. Detectability (Theorem 12) is the special case $f_j = \text{nominal}$.*

Proof. Diagnosability requires every pair of modes to be distinguishable by the observed field. By Theorem 12 applied to the pair (f_i, f_j) (treating f_j as the reference in place of nominal), f_i and f_j are distinguishable iff their signature pair lies in $\mathrm{Sep}(M)$. Quantifying over all pairs gives the claim; the $f_j = \text{nominal}$ instance is detectability. ◀

This is the statement the recovery theorems (Section 4) actually use: the rank test, conflict-diagnosability, and DES diagnosability are all *pairwise* distinguishability conditions.

5.3 When is a system NOT diagnosable? A non-circular criterion

The characterization of *non*-diagnosability is where the framework earns its keep against the circularity objection: it is a checkable structural condition, defined purely through the invariant map, that produces an explicit witness – and it never mentions diagnosability in its hypothesis.

▶ **Definition 29** (Invariant map and its fibres). *For a field $M = (\iota_1, \ldots, \iota_r)$ let $\Phi_M : F^+ \to V$ send each mode to its invariant image $\Phi_M(f) = (\iota_1(s^f), \ldots, \iota_r(s^f))$. A* confusable class *is a fibre $\Phi_M^{-1}(v)$ containing more than one mode; a fibre is* non-trivial *if it contains two distinct modes. Φ_M and its fibres are defined without any reference to diagnosability – they are computed from the invariants alone.*

▶ **Theorem 30** (Non-diagnosability has a fibre witness). *F is* non-diagnosable *by M iff Φ_M is not injective on F^+, i.e. iff there exists a non-trivial fibre $\Phi_M^{-1}(v) = \{f_i, f_j, \dots\}$. The fibre is an explicit witness: the modes it contains are exactly those M cannot tell apart, and the shared value v identifies which invariants are blind to their difference. Equivalently, F is diagnosable iff Φ_M is injective on F^+.*

Proof. By Corollary 28, F is diagnosable iff every pair $f_i \neq f_j$ has $([s^{f_i}], [s^{f_j}]) \in \mathrm{Sep}(M)$, i.e. some ι_k separates them: $\iota_k(s^{f_i}) \neq \iota_k(s^{f_j})$, i.e. $\Phi_M(f_i) \neq \Phi_M(f_j)$. So diagnosability $\Leftrightarrow$ Φ_M injective. Non-diagnosability is the negation: a coincidence $\Phi_M(f_i) = \Phi_M(f_j)$, i.e. a non-trivial fibre, whose members are by definition indistinguishable by every ι_k and hence by M. ◀

▶ Remark 31 (Why this is not circular). The hypothesis of Theorem 30 – "Φ_M has a non-trivial fibre" – is a property of the invariant map on the fault set, computable before diagnosability is even defined. It does not say "diagnosability fails because faults are not separated"; it says "faults f_i, f_j share invariant value v," a checkable fact about Φ_M, and *concludes* non-diagnosability. The invariants are not defined through diagnosability (they are G_S-invariant, orbit-separating functionals, Definition 6, fixed by the model via the canonical G_S of Proposition 26); diagnosability is *derived* from them. The direction of definition is thus model $\to G_S \to$ invariants $\to \Phi_M \to$ fibres $\to$ (non-)diagnosability, with no step depending on its successor. This is the precise sense in which the account is non-circular, and the fibre is the tangible object the classical conditions were implicitly detecting.

▶ Remark 32 (The three classical failures are all "a fibre exists"). Non-diagnosability unifies the three paradigms' *failure* conditions as a single statement. The rank test fails iff two fault signature columns are dependent – a fibre of Φ_M (equal residual directions). Conflict diagnosability fails iff two faults induce identical conflicts – a fibre (equal conflict signatures). Sampath diagnosability fails at an indeterminate cycle – a fibre at the *language* level (two traces with equal observation projection). Each classical non-diagnosability condition is exactly "Φ_M has a non-trivial fibre," with the paradigm fixing the level at which Φ_M is evaluated (states for LTI/Reiter, traces for DES; cf. the DES level-lift, Theorem 22).

5.4 Diagnosable-equivalence of realizations, and prediction

The deepest consequence is that diagnosability is an invariant of the *minimal sufficient realization* (Definition 4), which lets the framework predict when two structurally different methods are diagnostically identical – without running either.

▶ **Definition 33** (Diagnosable-equivalence). *Two methods with fields M, M' (on realizations S, S') are* diagnosable-equivalent on F*, written $M \equiv_F M'$, if they induce the same partition of F^+ into confusable classes: Φ_M^{-1} and $\Phi_{M'}^{-1}$ give the same fibres over F^+.*

▶ **Theorem 34** (Diagnosable-equivalence is isomorphism of minimal sufficient realizations on F). *$M \equiv_F M'$ iff the minimal sufficient realizations $S_M, S_{M'}$ (Definition 4) are* Struct*-isomorphic when restricted to the fault set F^+. Consequently diagnostic capability on F is a function of $S_M|_{F^+}$ alone: two methods with isomorphic minimal sufficient realizations over F^+ have identical diagnosability, regardless of how different their fields, inference rules, or paradigms appear.*

Proof. $M \equiv_F M'$ means Φ_M and $\Phi_{M'}$ have equal fibres on F^+, i.e. they induce the same equivalence $\sim_M = \sim_{M'}$ restricted to F^+. The minimal sufficient realization is the quotient by exactly this equivalence (Definition 4), so equal fibres give $S_M|_{F^+} = S_{M'}|_{F^+}$ up to relabelling,

i.e. a Struct-isomorphism restricted to F^+ (invariant-preserving by construction. Conversely a Struct-isomorphism of the restricted minimal realizations matches their invariant partitions, hence their fibres, giving $M \equiv_F M'$. Diagnosability is a function of the fibre partition (Theorem 30), so equal partitions give equal diagnosability. ◀

▶ **Corollary 35** (Diagnostic partial order and degeneration). *Define $M \preceq_F M'$ if $\Phi_{M'}$'s fibres refine Φ_M's on F^+ (equivalently $S_{M'}|_{F^+}$ factors onto $S_M|_{F^+}$). Then $\preceq_F$ is a partial order on methods (modulo $\equiv_F$), $M \preceq_F M'$ means M' is* at least as diagnostic *as M on F, and a degeneration U_σ (forgetful functor) satisfies $U_\sigma M \preceq_F M$: each simplification moves* down *the diagnostic order by coarsening the confusable-class partition. The degeneration hierarchy is thus recovered from a purely diagnosability standpoint.*

Proof. Refinement of partitions is a partial order; it corresponds to factoring of the quotient realizations, so $\preceq_F$ orders methods by their minimal sufficient realizations. $M \preceq_F M'$ gives Φ_M coarser, hence fewer separated pairs, hence weaker diagnosability. A forgetful functor discards invariants, coarsening Φ, so $U_\sigma M \preceq_F M$. ◀

▶ Remark 36. Theorem 34's value is not that it predicts an otherwise-unknowable fact, but that it makes diagnosable-equivalence cheap to certify: computing $S_M|_{F^+}$ requires only evaluating the invariants on each fault mode, not running the full method (estimator, solver, or trained model) and comparing outputs. The feeder example (Sec. 6.2) substantiates this – the equivalence of the $\{3, 5, 6\}$ and full placements is established from the sensitivity-matrix signatures, not from running seven vs. three diagnosers to convergence.

6 Empirical Validation: Power Distribution Feeder

We now validate our approach on a power distribution feeder [1] and exercise *two* of its realizations: a structural (LinDistFlow) realization whose diagnosability recovers measurement-placement observability, and a discrete-event realization whose diagnosability recovers Sampath's condition for a protection fault. Both use the orbit-separation criterion.

6.1 LinDistFlow Radial Feeder Model and Fault Diagnostics Framework

6.1.1 System Description and Mathematical Framework

The system under analysis consists of a radial electrical distribution feeder modeled via the linearized *LinDistFlow* equations, a standard simplification of the full non-linear AC power flow equations tailored for radial networks. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represent a tree-structured primary distribution graph where $\mathcal{V} = \{0, 1, \ldots, N\}$ is the set of buses (nodes), and $\mathcal{E}$ is the set of distribution lines (edges). Bus $0 \in \mathcal{V}$ denotes the substation slack bus, maintained at a fixed reference voltage magnitude v_0. Each line $e = (i, j) \in \mathcal{E}$ possesses a known series resistance r_e.

Under the active-power-dominant, flat-voltage linear approximation, the bus voltage magnitude vector $v \in \mathbb{R}^{N+1}$ is determined by active power injections $p \in \mathbb{R}^{N+1}$ via $v = v_0\mathbf{1} - R_{\text{eff}}\, p$ where $\mathbf{1}$ is a vector of ones, and $R_{\text{eff}} \in \mathbb{R}^{(N+1)\times(N+1)}$ represents the effective path resistance sensitivity matrix. The elements of the sensitivity matrix $S = -R_{\text{eff}}$ govern the voltage variations $\frac{\partial v_i}{\partial p_j}$ and are defined explicitly by the shared path structure from the slack bus:

$$S_{i,j} = \frac{\partial v_i}{\partial p_j} = -\sum_{e \in \mathcal{P}_i \cap \mathcal{P}_j} r_e \tag{1}$$

where $\mathcal{P}_i \subset \mathcal{E}$ denotes the unique ordered set of lines forming the directional path from the slack bus 0 to bus i.

6.1.2 Fault Signatures and Structural Perturbations

The framework evaluates three distinct classes of operational anomalies spanning structural, parametric, and observation-side configurations:

1. **Structural Line Faults:** A localized change in topology or physical parameters (e.g., a line degradation or partial failure) manifests as an impedance increment Δz on a given line $e_k = (f, t) \in \mathcal{E}$. This alters the underlying path resistance, inducing a systematic voltage drop downstream of the faulted edge:

$$\Delta v_b^{\text{line}} = \begin{cases} -\Delta z \cdot r_{e_k} & \text{if } b \in \mathcal{D}(t) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

   where $\mathcal{D}(t) \subseteq \mathcal{V}$ is the downstream subtree rooted at the child bus t.
2. **Parametric Injection Faults:** Deviations in nodal power consumption or unmonitored distributed generation yield a net injection shift Δp_j at bus j. The resulting full voltage signature corresponds to the j-th column of the sensitivity matrix S: $\Delta v^{\text{inj}} = S_{:,j}\, \Delta p_j$
3. **Observation-Side Sensor Faults:** A measurement instrument anomaly (e.g., a PMU or smart meter bias) alters only the local readout without perturbing the physical grid state, i.e., $\Delta v^{\text{sensor}} = \mathbf{e}_b\, \delta_{\text{bias}}$ where $\mathbf{e}_b$ is the b-th standard basis vector, and δ_{bias} is the additive measurement bias.

6.1.3 Observation Field and Confusable Fibres

$\mathcal{M} \subseteq \mathcal{V}$ denotes the observation placement (the subset of metered buses). The observation operator $\Phi_{\mathcal{M}} : \mathbb{R}^{N+1} \to \mathbb{R}^{|\mathcal{M}|}$ projects the full system state signature to the measured subspace: $\Phi_{\mathcal{M}}(\Delta v) = \mathbf{P}_{\mathcal{M}}\, \Delta v$ where $\mathbf{P}_{\mathcal{M}}$ is a projection matrix extracting rows indexed by $\mathcal{M}$.

Two system faults f_1, f_2 belong to the same *confusable fibre* $\mathcal{F}$ under the placement $\mathcal{M}$ if and only if their measured signatures are indistinguishable:

$$\mathcal{F}(f_1) = \{f_2 \in \mathcal{F}_{\text{all}} \mid \Phi_{\mathcal{M}}(\Delta v(f_1)) = \Phi_{\mathcal{M}}(\Delta v(f_2))\} \tag{3}$$

Complete structural diagnosability is attained if and only if all induced fibres are singletons ($\forall \mathcal{F}, |\mathcal{F}| = 1$). Placements that produce identical partitions of the fault space are certified as *diagnosable-equivalent*.

6.2 Structural realization: diagnosability vs. meter placement

We take the 7-bus radial feeder (6 lines) with the LinDistFlow linearization [1], so bus voltages depend linearly on injections through the sensitivity matrix S (algebraic, no dynamic simulation). The fault set mixes three types, giving non-trivial fibre structure: *line* faults (structural – an impedance change perturbs downstream voltages), *injection* faults (parametric – a load/generation deviation), and *sensor* faults (observation-side – a meter bias). The measurement field is the set of metered buses; its invariants are the measured voltage components of each fault's signature. Orbit separation under this field is exactly measurement-placement observability: two faults are diagnosable iff their measured voltage signatures differ.

Table 2 reports diagnosability for four meter placements, and Figure 1 shows the confusable-class partitions. The results recover the standard power-system facts and expose them as orbit separation: full metering is diagnosable; a sparse placement $\{3, 6\}$ cannot see the line-4 fault (its voltage drop is unmetered), confusing it with nominal; a trunk

Table 2 Structural diagnosability of the feeder vs. meter placement. Orbit separation under the metered-voltage field recovers measurement-placement observability; witnesses are the confusable fibres.

Meter placement	Confusable witness fibre(s)	Diag.?
full $\{0–6\}$	–	yes
leaves $\{3, 5, 6\}$	–	yes
sparse $\{3, 6\}$	$\{\text{nom}, \text{line}_4\}$	no
trunk $\{1, 2, 4\}$	$\{\text{nom}, \text{line}_2, \text{line}_4, \text{line}_5, \text{sen}_3, \text{sen}_6\}$; $\{\text{inj}_5, \text{inj}_6\}$	no

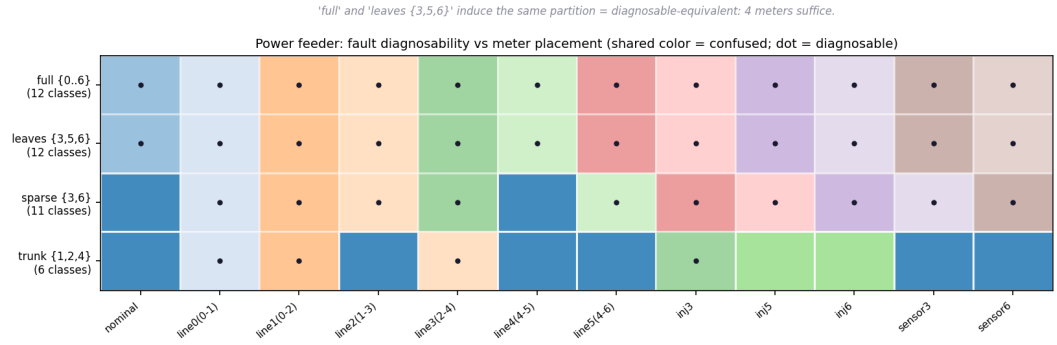

Figure 1 Feeder fault diagnosability across meter placements (rows) over the fault modes (columns); shared colour = confused, dot = diagnosable. The sparse placement $\{3, 5, 6\}$ induces the same partition as full metering: by Theorem 34 the two are *diagnosable-equivalent*, so four meters suffice for full diagnosability.

placement $\{1, 2, 4\}$ produces two witness fibres – it misses downstream leaf and sensor faults, and it confuses the two injection faults $\{\text{inj}_5, \text{inj}_6\}$ that share a trunk path. Each witness (Theorem 30) names the faults a placement cannot resolve and, by the unmetered channel, the meter whose addition would resolve them.

A non-obvious prediction: sparse metering suffices. The computation finds that metering only the three leaf buses $\{3, 5, 6\}$ induces the *same* fibre partition as metering all seven buses. By Theorem 34 the two placements are diagnosable-equivalent: the framework certifies, without enumerating faults or running any estimator, that four of the seven meters can be removed with *zero* loss of diagnosability. This is the predictive content of the diagnosability theory applied to a design question (sensor placement) it was not tuned for, and it is exactly the kind of cross-configuration equivalence the per-paradigm observability tests do not express directly.

6.3 Discrete-event realization: protection diagnosability

The same feeder also has a discrete-event realization: protection relays and breakers generate observable switching events, and a protection fault – a breaker that fails to open (*stuck-closed*) – is an unobservable event whose diagnosability is a Sampath condition. We build the protection diagnoser and apply orbit separation at the language level (Theorem 22).

Two designs are tested. When backup protection observably fires on a stuck breaker (an observable `backup_trip` event), the faulty trace reaches a fault-certain diagnoser state: no indeterminate cycle, Sampath- diagnosable, and orbit separation agrees. When the backup is

silent, the stuck-breaker trace produces the same observations as a normal clear, creating an indeterminate cycle: non-diagnosable, and orbit separation agrees (both cases verified). The protection fault's diagnosability is thus the same orbit-separation criterion that governs the structural layer and the two-tank, now on the feeder's event language.

6.4 What the Power System Example Establishes

▶ Remark 37 (One criterion, two realizations, a new plant). The feeder demonstrates the framework on a system structurally unlike systems like the two-tank, and does so through two different realizations of the *same* plant: a structural realization recovering measurement-placement observability, and a discrete-event realization recovering Sampath protection-diagnosability. Both are orbit separation; only the realization and its symmetry group differ. The structural layer also yields a concrete, non-obvious engineering prediction – a sparse meter placement diagnosable-equivalent to full metering – illustrating that the diagnosability theory answers design questions, not only classification ones.

7 Computational Advantages of Cross-Realization Reasoning

A framework that unifies diagnosability could be merely descriptive. We therefore address directly what *computational* advantage orbit separation provides beyond a common language. The honest answer has a sharp boundary, which we state first, and three genuine advantages, which lie on the far side of it.

7.1 What is not claimed: no single-realization complexity win

Within a single fixed realization, computing the orbit partition is not cheaper than computing the classical condition – because it is, up to interpretation, the *same* computation. For the LTI realization the fibres of Φ_M are obtained from the fault-signature directions, i.e. the rank computation itself; for the discrete-event realization the language-level fibres are the diagnoser states, i.e. the Sampath construction itself; for the Reiter realization the fibres are the conflict signatures. Orbit separation reinterprets these objects; it does not produce them faster. Any claim of a raw complexity advantage at the single-realization level would therefore be spurious, and we make none. The advantages below are all at the level where the classical theories are *silent* – across realizations, across a method library, and in guiding acquisition – because each classical test is confined to one representation and provides no cross-representation comparison.

7.2 A1 – Certified cheapest realization within an equivalence class

The diagnosable-equivalence theorem (Theorem 34) partitions candidate realizations into classes of equal diagnostic capability. Within a class one may run diagnosis on the *cheapest* member – fewest sensors, smallest state, lowest per-query cost – with a *certificate* that no diagnostic capability is lost, because the minimal sufficient realizations coincide. The advantage is not that the framework sorts realizations by cost (any scheme can do that) but that it certifies *which realizations are interchangeable*, so the cheaper one is provably sufficient.

▶ **Example 38** (Sparse metering, from Section 6)**.** On the power feeder, metering the three leaf buses $\{3, 5, 6\}$ is diagnosable-equivalent to metering all seven (Theorem 34, verified in Section 6.2). Diagnosis may therefore run on a 3-dimensional measurement vector instead

of 7, with a structural certificate of identical diagnosability. The saving is concrete – fewer sensors to install and read, a smaller estimation problem – and the certificate is exactly what the per-placement observability test cannot supply: that test verifies each placement in isolation but does not establish the two are *equivalent*, which would otherwise require a fault-by-fault comparison of their diagnoses.

The general statement: because diagnosability is an invariant of the minimal sufficient realization (Theorem 34), the cheapest realization with the same $S_M|_{F^+}$ is a provably-lossless substitute, and identifying it is a comparison of partitions – cheap relative to constructing or running the realizations themselves.

7.3 A2 – Partial-order pruning of a method library

When several methods or realizations are available, the diagnostic partial order (Corollary 35) lets dominated candidates be discarded *without evaluating them*. If $M \prec_F M'$ then M' diagnoses everything M does; if in addition M' is no more costly, M is dominated and can be dropped before any risk or accuracy computation. Domination is detected structurally – by partition refinement – rather than by running each method, so the expensive per-method evaluation is confined to the Pareto-undominated survivors.

This is a genuine combinatorial saving in the library setting: the cost of selection scales with the number of *surviving* (undominated) methods, not the library size, and the survivors are identified by comparing partitions. The refinement test is a containment check on fibres, whose cost is governed by the fault set size, independent of the internal complexity of the methods being ordered. A library with many dominated variants collapses to its Pareto frontier before the costly step, which is where the per-realization diagnosis or risk estimate is actually incurred.

7.4 A3 – Fibre-targeted incremental acquisition

When a fault set is non-diagnosable, the confusable fibre (Theorem 30) localizes the deficiency: it names the confused modes and, through the invariant they share, the specific distinction the current field cannot make. This makes improvement *incremental and targeted* rather than global. One need not rebuild the diagnoser or re-solve the full diagnosis problem; one computes only the *additional* invariant that splits the offending fibre.

▶ **Example 39** (Targeted meter addition, from Section 6)**.** The trunk placement $\{1, 2, 4\}$ yields the witness fibre $\{\mathrm{inj}_5, \mathrm{inj}_6\}$: the two injection faults share the trunk-path voltage signature. The fibre identifies precisely the missing distinction – a measurement downstream of the branch separating buses 5 and 6 – so the remedy is a single targeted meter, computed against the one confusable class, not a re-instrumentation and re-solve of the whole network. Acquisition is thus guided by the fibre to the minimal resolving addition.

The general pattern: the confusable fibres form a to-do list of exactly the distinctions the current realization lacks, each pointing at the invariant whose addition resolves it. Refinement of diagnosability proceeds fibre-by-fibre, touching only the affected confusable class, rather than by global recomputation – the diagnostic analogue of splitting a coarse equivalence only where it is too coarse.

7.5 Accuracy: error avoidance, not raw improvement

We are equally precise about accuracy. The framework does not make any single method more accurate – an observer's residual is what it is. Its contribution to accuracy is *error avoidance*: by certifying the confusable set (Theorem 30), it prevents the characteristic

failure of trusting a method on a fault it is structurally blind to. A confident diagnosis of a mode that lies in a non-trivial fibre is a confident *mis*diagnosis; the fibre flags this in advance, so the method is not over-trusted beyond its resolving power. The accuracy contribution is a guarantee about *when not to trust a verdict* – valuable precisely because classical per-paradigm confidence scores do not express structural blindness.

7.6 Summary of the advantage

▶ Remark 40 (Where the advantage lives). The computational advantage of orbit separation is not a faster diagnosability test for a fixed method – there it coincides with the classical computation. It is the ability to reason *across* realizations and methods with a common currency, the fibre partition, which the per-paradigm tests cannot provide: to substitute a cheaper equivalent realization with a certificate (A1), to prune a method library to its diagnostic Pareto frontier before the costly step (A2), and to target acquisition at the exact confusable class that limits diagnosability (A3), while avoiding confident misdiagnosis by certifying structural blindness. These are advantages of *comparison and selection*, cheap relative to running the methods, and they are available only because diagnosability has been reduced to a single computable invariant – the minimal sufficient realization's partition – shared across paradigms.

8 Related Work

Fault diagnosis has traditionally been studied through paradigm-specific theories, including observer-based FDI, consistency-based diagnosis, probabilistic diagnosis, discrete-event diagnosis, geometric diagnosis, and more recently learning-based approaches [9, 8, 12, 13]. Each paradigm introduces its own notion of distinguishability – e.g., residual rank conditions, conflict sets, indeterminate cycles, or observability codistributions – and these notions are rarely analyzed within a common framework. Our contribution is to show that they can all be interpreted as instances of a single realization-theoretic criterion: orbit separation.

Behavioural systems and realizations. The behavioural approach of Willems treats behaviour rather than state-space representation as the primary object of analysis [15]. Our realization map follows this perspective by associating a behavioural specification with a structure on which diagnosis is performed. The key difference is that behavioural theory studies equivalence of behaviours, whereas we study equivalence of diagnostic capability. Two realizations may differ behaviourally yet be diagnosable-equivalent if they induce the same partition of a fault set.

Bisimulation and quotient constructions. Bisimulation theory characterizes behavioural equivalence by quotienting states that are observationally indistinguishable [14]. Our minimal sufficient realization plays a similar role: it is the quotient induced by diagnostic invariants. Diagnosable-equivalence can therefore be viewed as a fault-relative analogue of bisimulation. The distinction is that our observables are generalized diagnostic invariants rather than transition labels, allowing continuous, logical, probabilistic, and geometric realizations to be handled within a common framework.

Discrete-event diagnosability. Sampath et al. characterize diagnosability through the diagnoser automaton and the absence of indeterminate cycles [13]. In our framework the diagnoser itself serves as the realization, and diagnosability becomes orbit separation at the language level. The resulting correspondence shows that DES diagnosability is another instance of the same structural criterion used in continuous and logical settings.

Geometric and observer-based diagnosis. Geometric observability theory characterizes state distinguishability through observability codistributions and associated indistinguishability manifolds [10]. Fault detection and isolation methods based on geometric control theory extend this perspective to fault signatures and isolability conditions [7]. In our terminology, the indistinguishability manifolds correspond to orbits and observability conditions correspond to orbit-separating invariants. The classical fault-signature rank test appears as the linear specialization of this broader picture.

Structural diagnosis and analytical redundancy. Structural FDI approaches characterize detectability and isolability through graph structure, analytical redundancy relations, and matching conditions [2, 9]. These methods already reason through a realization, namely a structural graph representation. Our framework generalizes this idea by allowing the realization to take many forms, including logical theories, Bayesian networks, diagnoser automata, geometric models, or learned representations. Structural diagnosis thus becomes one realization class among several rather than the foundational abstraction itself.

Consistency-based diagnosis. Consistency-based diagnosis and model-based reasoning identify faults through conflicts and minimal diagnoses derived from a logical system description [12]. Our recovery result shows that conflict diagnosability corresponds to orbit separation in a logical realization under component-relabelling symmetries. This places logical diagnosis within the same structural framework as observer-based and DES methods.

Symmetry and learned representations. The emphasis on symmetries and invariants connects naturally with work on equivariant systems, geometric deep learning, and invariant representation learning [4, 3]. These approaches study what information is preserved under group actions, whereas our focus is diagnostic distinguishability. The connection is that diagnostic capability is determined by the invariants exposed by a realization and the orbit structure induced by its symmetry group.

Summary. Across these literatures a common pattern emerges: each defines a notion of indistinguishability and a corresponding quotient structure. The realization-theoretic framework identifies these as manifestations of the same construction. Orbit separation provides a common criterion for fault distinguishability, while diagnosable-equivalence allows diagnostic methods from different paradigms to be compared through their minimal sufficient realizations. This enables cross-paradigm comparisons that individual diagnosis theories, considered in isolation, do not provide.

9 Summary and Conclusions

This paper introduced a realization-theoretic framework for fault diagnosis based on the Diagnostic System Specification (DSS), which separates behavioural specifications, realizations, measurement fields, and inference procedures. The framework identifies diagnosis as a property of the realization through which a system is viewed rather than of any particular diagnostic algorithm.

The central result is that diagnosability is an orbit-separation property. Relative to a realization and its nominal symmetry group, faults are distinguishable iff they occupy different invariant-separated orbits. This perspective yields a minimal sufficient realization that completely characterizes diagnostic capability and leads to a structural notion of diagnosable-equivalence among diagnostic methods.

We showed that several classical diagnosability conditions are recovered as special cases of the same criterion, including observer-based rank conditions, Reiter conflict diagnosability, and Sampath diagnosability. The framework therefore provides a common language for comparing diagnostic paradigms that are traditionally studied separately.

The empirical studies on the two-tank system and a power distribution feeder demonstrated both the recovery of established results and the ability to identify diagnostically equivalent realizations, including equivalent sensor placements.

Future work includes extending the framework to richer probabilistic and learning-based systems, developing scalable algorithms for constructing minimal sufficient realizations, and applying diagnosable-equivalence to sensor placement, observer design, and model-selection problems.

References

1 Mesut E Baran and Felix F Wu. Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Transactions on Power delivery*, 4(2):1401–1407, 1989.

2 Mogens Blanke, Michel Kinnaert, Jan Lunze, and Marcel Staroswiecki. *Diagnosis and Fault-Tolerant Control*. Springer, 3rd edition, 2016.

3 Benjamin Bloem-Reddy and Yee Whye Teh. Probabilistic symmetries and invariant neural networks. *Journal of Machine Learning Research*, 21(90):1–61, 2020. URL: `https://jmlr.org/papers/v21/19-322.html`.

4 Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint*, 2021. `arXiv:2104.13478`.

5 Baoping Cai, Lei Huang, and Min Xie. Bayesian networks in fault diagnosis. *IEEE Transactions on industrial informatics*, 13(5):2227–2240, 2017. `doi:10.1109/TII.2017.2695583`.

6 Johan De Kleer and Brian C Williams. Diagnosing multiple faults. *Artificial intelligence*, 32(1):97–130, 1987. `doi:10.1016/0004-3702(87)90063-4`.

7 Claudio De Persis and Alberto Isidori. A geometric approach to nonlinear fault detection and isolation. *IEEE Transactions on Automatic Control*, 46(6):853–865, 2001. `doi:10.1109/9.928586`.

8 Steven X Ding. *Data-driven design of fault diagnosis and fault-tolerant control systems*. Springer, 2014.

9 Janos Gertler. *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker, 1998.

10 Robert Hermann and Arthur J. Krener. Nonlinear controllability and observability. *IEEE Transactions on Automatic Control*, 22(5):728–740, 1977.

11 Gregory Provan. Statistical foundations of local geometric fault detection: the curvature channel. In *Principles of Diagnosis and Resilient Systems*, 2026.

12 Raymond Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32(1):57–95, 1987. `doi:10.1016/0004-3702(87)90062-2`.

13 Meera Sampath, Raja Sengupta, Stéphane Lafortune, Kasim Sinnamohideen, and Demosthenis Teneketzis. Diagnosability of discrete-event systems. *IEEE Transactions on Automatic Control*, 40(9):1555–1575, 1995. `doi:10.1109/9.412626`.

14 Rob J. van Glabbeek and W. Peter Weijland. Branching time and abstraction in bisimulation semantics. *Journal of the ACM*, 43(3):555–600, 1996. `doi:10.1145/233551.233556`.

15 Jan C. Willems. The behavioral approach to open and interconnected systems. *IEEE Control Systems Magazine*, 27(6):46–99, 2007.