# Summary Of: Modelling Cyber-Physical Systems for Fault Diagnosis

**Alexander Diedrich**
Helmut-Schmidt-University, Hamburg, Germany

**Mattias Krysander**
Linköping University, Sweden

**René Heesch**
Helmut-Schmidt-University, Hamburg, Germany

**Oliver Niggemann**
Helmut-Schmidt-University, Hamburg, Germany

**Abstract**

This is an extended abstract of the manuscript 'Modelling Cyber-Physical Systems for Fault Diagnosis' [4] that was published in the journal IEEE Transactions on Systems, Man, and Cybernetics in December 2025.

**2012 ACM Subject Classification** Computing methodologies → Knowledge representation and reasoning

**Keywords and phrases** Fault Diagnosis, SMT, Multimode Systems

**Digital Object Identifier** 10.4230/OASIcs.DX.2026.18

**Category** Extended Abstract

**Related Version** *Full Version*: https://ieeexplore.ieee.org/abstract/document/11194731

**Funding** *René Heesch*: This research as part of the project EKI is funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr, which we gratefully acknowledge.

## 1 Extended Abstract

The complexity of cyber-physical systems, paired with significant availability requirements, require prompt reactions when faults occur [13]. To ensure timely fault handling, it is necessary to compare data from the system to a pre-existing model of nominal working behaviour and reason about possible faults [20]. The required model, that we will refer to as a *system description* (SD), is an essential but difficult-to-obtain part of the diagnostic process.

Overall, the system description models the structure and behaviour of technical systems [20]. It expresses for each individual component how it is connected to other components (its connections) and how its output is calculated from its inputs (its behaviour model) [8]. For a two-bit Boolean AND-gate, for example, the system description is defined as $\neg ab(c_{AND}) \rightarrow out = in_0 \wedge in_1$ [10], where the behaviour is given through the Boolean operator $\wedge$ and the connection is defined through relating the inputs $in_0, in_1 \in \{0, 1\}$ to $out \in \{0, 1\}$. The predicate $\neg ab(c_{AND})$ defines the healthiness (non-abnormality [20]) of the AND-gate. Such a formulation of the system description is called a weak-fault model, i.e., it models the component's normal behaviour and does not specify component behaviour under faults [23].

Consistency-based fault diagnosis [20] algorithms use such logical weak-fault system descriptions to infer faults by comparing the system description to actual observations of the system. If a fault exists in the system, the observations together with the system description


37th International Conference on Principles of Diagnosis and Resilient Systems (DX 2026).
Editors: Ingo Pill, Marina Zanella, and Gregory Provan; Article No. 18; pp. 18:1–18:6
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

lead to an inconsistency. Using non-monotonic reasoning [21], a fault diagnosis algorithm [22] is able to infer potential faulty components (also known as diagnoses or root causes) from this inconsistency. Until now, system descriptions have usually been used for diagnosing Boolean circuits [20, 10], while only a few authors have extended propositional logic system descriptions beyond systems with only Boolean circuits [11, 5, 17].

What is so far missing from current literature is a unified formulation of system descriptions that allows reasoning over sufficiently complex systems, such as those generating continuous signals and operating in different modes. Obtaining such a more elaborate system description is highly desirable for performing fault diagnosis of cyber-physical systems [19, 15]. These systems are used in most modern applications and combine intelligent monitoring and control with classical mechanical system components. Because modern applications require constant adjustments and updates, those systems must adapt over time, contain proprietary components with complicated behaviour models, operate in several modes, and generate Boolean, integer, and continuous signals.

To tackle these challenges we pose the following three research questions within the article.

**RQ1**: *Is there a consistent logical theory to extend classical Boolean consistency-based fault diagnosis to work with operating modes and continuous signals?*

**RQ2**: *How is it possible to diagnose faults, provided a consistent logical system description for modes and continuous signals exists?*

**RQ3**: *Can the approach be adapted to reasoning approaches requiring less expert knowledge?*

To answer all three research questions and obtain a consistent logical theory for diagnosing faults in cyber-physical systems, the article presents a novel formulation in SMT over nonlinear real arithmetic (SMT-$\mathcal{NRA}$) [3] for consistency-based fault diagnosis. We show how to use established fault diagnosis algorithms adapted to an SMT-$\mathcal{NRA}$ system description, and introduce the novel HySD_enhanced algorithm for a direct approach to consistency-based fault diagnosis. Overall, this article presents the following contributions:

- We present a novel SMT-$\mathcal{NRA}$ formulation that includes operating modes, continuous signals, and behaviour models.
- Using the SMT-$\mathcal{NRA}$ formulation, we adapt classical consistency-based fault diagnosis theory [10, 16] and introduce the novel algorithm HySD_enhanced to perform fault diagnosis.

We evaluated our approach on two benchmarks of multiple-tank systems [9, 1], two Boolean circuits, and a use-case of dynamically configurable battery packs. While we use the two benchmarks and the battery pack example to evaluate the capabilities and limitations of our approach, we use the Boolean circuits to show that our method also generalizes to classical Boolean circuit fault diagnosis. With our experiments we are able to diagnose 97% of the injected faults.

Underlying our approach are some assumptions that we need to state explicitly: i) While in previous work we presented methods to learn and validate propositional logic system descriptions [6, 17, 7] and there is ongoing research to learn SMT expressions from data [14], but there is so far no approach to learn the system expressions in this article in their entirety. ii) While parts of the presented system description have been proven to be learnable, for example, behaviour models or state estimations, there is still some work required to fit these into our presented system description. iii) We assume a cyber-physical system can be subdivided into a set of operating modes that is significantly smaller than the true operating modes in the data. Earlier work showed that through the use of discretisation functions, such a subdivision is well attainable [17, 18].

The main benefit of our approach is that the presented type of system description is much more general than traditional propositional logic system descriptions [17, 6], i.e., has broader variable domains, is able to handle operating modes, and supports inference from symptoms to faulty components.

The main idea behind using SMT-$\mathcal{NRA}$ for consistency-based fault diagnosis is to formulate the system description in such a way that components are *suitably* associated to relevant observations. This means we first require a subset $COMPS' \subset COMPS$ that influences the health of a single observation. For this article it is reasonable to assume that such a set $COMPS'$ exists [17, 6]. The general form of these system descriptions is

$$\bigwedge_{obs_j \in OBS} \Big( \bigwedge_{c_i \in COMPS_k} \neg ab(c_i) \rightarrow obs_j \Big) \tag{1}$$

where $COMPS_k \subset COMPS$, with $k \in \mathbb{N}$. In other words, changes in the health state of a component are assumed to have a direct influence on a particular observation. The goal in this article is to use this architecture of formulating logical expressions and significantly increase its generality. The intuition is to associate component health states with discretisations of the expected and actual state and the expected and actual observations.

$$\bigwedge_{c_j \in \mathrm{COMPS}'} \neg ab(c_j) \rightarrow \big((d_s(\mathcal{V}) = d_s(\mathcal{V}')\big) \wedge \big(|g_m(\mathcal{V}) - \alpha_i| \leq \epsilon)\big) \tag{2}$$

$COMPS' \subset COMPS$ is a set of components relevant for the current formula (e.g., their behaviour has an influence on the mode and measurements of the system). Function $d_s(\cdot)$ computes the expected ($d_s(\mathcal{V})$) and actual mode ($d_s(\mathcal{V}')$) through state estimation methods, where $\mathcal{V} \neq \mathcal{V}'$. This means the expected mode is calculated from variables $\mathcal{V}$ describing how the system should behave, while variables $\mathcal{V}'$ describe actual measurements from the system. Function $g_m(\mathcal{V})$ computes the expected measurement given mode $m \in \mathcal{N}$, while $\alpha_i$ is a single actual measurement. $\epsilon \in \mathbb{R}$ denotes a noise threshold. We use the $\neg ab$-predicates to denote non-abnormal behaviour [20].

To perform fault diagnosis, it is necessary to compare the SMT-$\mathcal{NRA}$ system description with current observations. The comparison is done by computing an expected value of the measurement, i.e., the expectation of normal behaviour $g(\cdot)$, given the current operating mode. To compare the prediction with the actual measurement $\alpha_i$ it is possible to calculate a residual value $r_i \in \mathbb{R}$ through

$$r_i = \begin{cases} |g(\mathcal{V}) - \alpha_i| \leq \epsilon; \text{ no fault} \\ |g(\mathcal{V}) - \alpha_i| > \epsilon; \text{ else} \end{cases} \tag{3}$$

to account for noise $\epsilon \in \mathbb{R}$. Discretising the residual into a Boolean value by only computing the rightmost side of Equation (2) then leads to a computable interpretation. In other words, an expert-defined physical model calculates the expected value for some measurement. The difference between the expected and the actual value form the residual. The residual is then discretised to obtain a Boolean value indicating the fault symptom. SMT-$\mathcal{NRA}$ provides the advantage that all necessary calculations can be performed within the same logical calculus and can be proven through solvers such as Z3 or MathSAT.

A fault that is detected due to a non-zero residual is called a symptom. Symptoms occur in one of four cases:

1. The behaviour model outputs too much noise or the model error is too large. This is a system description design error and must be accounted for.

2. The behaviour model does not match to the current operating mode. This is a system description design error and must be accounted for.
3. Some sensor fails and the residual becomes non-zero. In this case, fault diagnosis is required to determine the minimum cardinality diagnosis.
4. Some component fails and due to deviating observations, the residual becomes non-zero. In this case, fault diagnosis is required to determine the minimum cardinality diagnosis.

To diagnose faults that are due to changes in the operating mode, the operating mode needs to be made explicit within the system description [17, 2]. To compare the expected mode to the actual mode using SMT-$\mathcal{NRA}$ several computations need to take place: i) Function $d_s(\mathcal{V})$ needs to calculate the expected discretised observable modes from the weak-fault behaviour model and the known variables of the system. ii) Function $d_s(\mathcal{V}')$ needs to compute the actual discretised observable modes from the known variables and the current observations of the system. Again, SMT-$\mathcal{NRA}$ allows the formulation of non-linear functions directly within the logical calculus.

We define a mode fault as a fault that, given two discretised model variables $m, m_a \in \mathbb{Z}$ computed by $d_s(\mathcal{V})$ and $d_s(\mathcal{V}')$, respectively, a mode fault occurs when the actual mode differs from the predicted mode $m \neq m_a$.

In the article we evaluate, whether we can find faults in several cyber-physical systems, with the SMT logic system description presented in Equation (2). Within the evaluation we use a representative set of several exemplary use-cases from the process industry, a set of configurable batteries, and a set of Boolean circuits. For this abstract, we only present a subset of the results. Please note that in tables with empirical results bold symbols denote diagnosis of the true root cause (i.e. the injected fault). Diagnoses in italics denote wrong diagnoses. Diagnoses that are neither bold, nor in italics are diagnoses allowed by the model due to partial observation.

For all experiments we used a threshold value $\epsilon$ of 0.001. This was sufficient as we used analytical models to implement function $g_m(\cdot)$, such that the difference between measurement and prediction is close to zero in the nominal case. In practice, $\epsilon$ may be adjusted by comparing the variance in the prediction of $g_m$ and the variance of $\alpha$ in the fault-free mode and over a representative set of historical data.

**Table 1** Empirical results for two well-known electronic circuits.

| **System** | **Fault** | Mode | **Diagnosis** |
|---|---|---|---|
| **Full-Adder** | f0 | Normal | **{f0}**, {f5,f6}, {f3} |
| | f1 | Normal | **{f1}** |
| | f5 | Normal | **{f5}**, {f0}, {f3} |
| | None | Normal | ∅ |
| **Polybox** | m0 | Normal | **{m0}**, {m1}, {a0} |
| | m1 | Normal | {a0,a1}, **{m1}**, {a0}, {m2,m0} |
| | a0 | Normal | **{m0}**, {m1}, {a0} |
| | None | Normal | ∅ |

Our approach was evaluated on a half-adder circuit and the well-known polybox example. The results are presented in Table 1. Results are similar to the process industrial use-cases above. However, the systems do only contain one operating mode and exhibit a lower observability. Still, in each experiment to correct fault diagnosis was computed. But due to the limited observability the diagnosis sets were larger.

Table 2 presents the diagnostic results of investigating all single faults for a multi-mode battery pack. The columns from left to right show the actuated mode, the injected fault, whether the fault directly affects an observed variable, and the single fault diagnoses. An

**Table 2** Empirical results for a multi-mode battery pack.

| Mode | Fault | Diagnosis |
|---|---|---|
| **Bypass** | NF | ∅ |
| | f_i_pack | ∅ |
| | f_i_1 | **{f_i_1}** |
| | f_v_1 | **{f_v_1}**, {f_cell_1} |
| | f_cell_1 | {f_v_1}, **{f_cell_1}** |
| | mode = Forward | ∅ |
| | mode = Backward | ∅ |
| **Forward** | NF | ∅ |
| | f_i_pack | **{f_i_pack}** |
| | f_i_1 | **{f_i_1}** |
| | f_v_1 | **{f_v_1}**, {f_cell_1} |
| | f_cell_1 | {f_v_1}, **{f_cell_1}** |
| | mode = Backward | {f_i_pack}, **{f_mode}** |
| | mode = Bypass | **{f_mode}** |
| **Backward** | NF | ∅ |
| | f_i_pack | **{f_i_pack}** |
| | f_i_1 | **{f_i_1}** |
| | f_v_1 | **{f_v_1}**, {f_cell_1} |
| | f_cell_1 | {f_v_1}, **{f_cell_1}** |
| | mode = Forward | {f_i_pack}, **{f_mode}** |
| | mode = Bypass | **{f_mode}** |

empty set diagnosis indicates that the system is fault-free, i.e., no fault is detected. An isolability analysis study is presented in [12], covering all faults except mode faults. The results here are consistent with the analysis results there. In non-ideal cases where the fault is either not detected or not uniquely isolated, it is due to a lack of information, that is the diagnosis inference is complete. Furthermore, the proposed diagnostic computational framework also handles fault-changing modes.

## References


1. Kaja Balzereit, Alexander Diedrich, Jonas Ginster, Stefan Windmann, and Oliver Niggemann. An ensemble of benchmarks for the evaluation of ai methods for fault handling in cpps. In *19th IEEE International Conference on Industrial Informatics*, November 2021.
2. Andreas Bunte, Benno Stein, and Oliver Niggemann. Model-based diagnosis for cyber-physical production systems based on machine learning and residual-based diagnosis models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33(01), pages 2727–2735, 2019. `doi:10.1609/AAAI.V33I01.33012727`.
3. Alessandro Cimatti, Alberto Griggio, Ahmed Irfan, Marco Roveri, and Roberto Sebastiani. Incremental linearization for satisfiability and verification modulo nonlinear arithmetic and transcendental functions. *ACM Transactions on Computational Logic (TOCL)*, 19(3):1–52, 2018. `doi:10.1145/3230639`.
4. Alexander Diedrich, Mattias Krysander, Rene Heesch, and Oliver Niggemann. Modeling cyber–physical systems for fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2025.
5. Alexander Diedrich, Alexander Maier, and Oliver Niggemann. Model-based diagnosis of hybrid systems using satisfiability modulo theory. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*. Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), 2019.
6. Alexander Diedrich, Lukas Moddemann, and Oliver Niggemann. Learning system descriptions for cyber-physical systems. In *Proceedings of 12th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*, 2024.

7 Alexander Diedrich, Lukas Moddemann, and Oliver Niggemann. On validating propositional logic system descriptions for fault diagnosis. *Engineering Applications of Artificial Intelligence*, 165:113379, 2026. doi:10.1016/J.ENGAPPAI.2025.113379.

8 Alexander Diedrich and Oliver Niggemann. On residual-based diagnosis of physical systems. *Engineering Applications of Artificial Intelligence*, 109:104636, 2022. doi:10.1016/j.engappai.2021.104636.

9 Jonas Ehrhardt, Malte Ramonat, Rene Heesch, Kaja Balzereit, Alexander Diedrich, and Oliver Niggemann. An ai benchmark for diagnosis, reconfiguration & planning. In *2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, 2012.

10 Alexander Feldman, Gregory Provan, and Arjan Van Gemund. Approximate model-based diagnosis using greedy stochastic search. *Journal of Artificial Intelligence Research*, 38:371–413, 2010. doi:10.1613/JAIR.3025.

11 Alban Grastien. Diagnosis of hybrid systems with smt: opportunities and challenges. In *Proceedings of the Twenty-first European Conference on Artificial Intelligence*, pages 405–410. IOS Press, 2014. doi:10.3233/978-1-61499-419-0-405.

12 Fatemeh Hashemniya, Arvind Balachandran, Erik Frisk, and Mattias Krysander. Structural diagnosability analysis of switched and modular battery packs. In *2024 Prognostics and System Health Management Conference (PHM)*, pages 362–369. IEEE, 2024.

13 Rolf Isermann and Peter Balle. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control engineering practice*, 5(5):709–719, 1997.

14 Samuel Kolb, Stefano Teso, Andrea Passerini, and Luc De Raedt. Learning smt (lra) constraints using smt solvers. In *IJCAI International Joint Conference on Artificial Intelligence*, volume 2018, pages 2333–2340. ijcai. org, 2018. doi:10.24963/IJCAI.2018/323.

15 Ion Matei, Maksym Zhenirovskyy, Johan de Kleer, and Alexander Feldman. A hybrid qualitative and quantitative diagnosis approach. In *Annual Conference of the PHM Society*, volume 11(1), 2019.

16 Amit Metodi, Roni Stern, Meir Kalech, and Michael Codish. A novel sat-based approach to model based diagnosis. *Journal of Artificial Intelligence Research*, 51:377–411, 2014. doi:10.1613/JAIR.4503.

17 Lukas Moddemann, Henrik Sebastian Steude, Alexander Diedrich, and Oliver Niggemann. Discret2di - deep learning based discretization for model-based diagnosis. In *Proceedings of 12th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*, 2024.

18 Phillip Johann Overlöper, Lukas Moddemann, Nemanja Hranisavljevic, Alexander Windmann, and Oliver Niggemann. Discretization of cps time series with neural networks. In *2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA)*, pages 1–8. IEEE, 2024. doi:10.1109/ETFA61755.2024.10710901.

19 André Platzer. *Logical foundations of cyber-physical systems*. Springer, 2018. doi:10.1007/978-3-319-63588-0.

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

21 Raymond Reiter. Nonmonotonic reasoning. In *Exploring artificial intelligence*, pages 439–481. Elsevier, 1988.

22 Patrick Rodler. How should i compute my candidates? a taxonomy and classification of diagnosis computation algorithms. In *33rd International Workshop on Principle of Diagnosis – DX 2022*, 2022.

23 P Struss. Model-based on-board diagnosis and tools for the developer of on-board systems-vmbd and idd: Two projects of the european car industries. *MONET Newsletter Issue 2 - Automotive Supplement*, 2002.