The Delin Metric (D_{CIR}) for Chaotic Molecular Sensing

A Neyman–Pearson Decision Metric
inside the EMDS/CIR Framework

Rodolphe André Marie Delin
Independent Researcher
Email: research@pyxym.com
LinkedIn: https://linkedin.com/in/rdelin

Version 2.0 — December 2025

This document forms part of the EMDS/CIR foundational framework:
"Delin Theory of Environmental Molecular Data Sensing (EMDS) and Chaotic Information Reading (CIR)"
DOI: https://doi.org/10.5281/zenodo.17731919

This document also references:
"Delin Theory of Environmental Molecular Data Sensing (EMDS) and Chaotic Information Reading (CIR)"
DOI: https://doi.org/10.5281/zenodo.17762141

Abstract

This document presents the *Delin Metric* D_{CIR} as a rigorous, closed-form decision statistic for weak-signal detection in chaotic molecular environments, within the broader EMDS/CIR framework (Environmental Molecular Data Sensing and Chaotic Information Reading).

The metric is derived from classical Neyman–Pearson detection theory for Gaussian observations, but applied in a novel way to distinguish chaotic molecular motifs observed through noisy sensors.

We show how D_{CIR} quantifies the *readability* of informational motifs embedded in turbulent or diffusive molecular fields, and how it can be used as: (i) a simulation-time decision layer for feasibility studies, (ii) a design criterion for EMDS-ready sensing configurations, and (iii) a software-first product for industrial partners interested in testing scenarios before committing to hardware development.

The mathematical derivation is fully explicit; a canonical Ornstein–Uhlenbeck toy model illustrates the theory and yields a reference case where D_{CIR} predicts a reconstruction success probability of about 98.8% for a simple binary motif. This establishes D_{CIR} as a *quantitative decision engine* rather than an abstract information measure, and forms the basis of Level 1 (software) in the EMDS/CIR commercialization roadmap.

1 Introduction and Context

The EMDS/CIR project proposes a new way of thinking about information in chaotic molecular environments such as air, gases, aerosols, liquids, or surfaces. Instead of writing bits into stable microscopic structures (as in classical memory), EMDS/CIR investigates whether naturally occurring chaotic dynamics can carry *detectable* information in their statistical patterns.

Within this framework:

- **Environmental Molecular Data Sensing (EMDS)** treats the environment as a high-dimensional, evolving molecular field with statistical structure.
- **Chaotic Information Reading (CIR)** denotes the reconstruction procedures that infer informational motifs from sensor time series.

In practice, industrial and scientific users need a simple, quantitative answer to a very concrete question:

> *"Given my sensor, noise level, and environment, is there any realistic chance of detecting this motif, or am I below the physics limit?"*

The *Delin Metric* D_{CIR} is designed precisely to answer this question in a mathematically grounded way. It is not a neural network, nor a black-box model; it is an explicit decision statistic derived from classical detection theory, but calibrated for chaotic molecular sensing.

This document plays a dual role:

1. Scientifically, it defines and derives D_{CIR} from first principles, in a way that can be scrutinized, reproduced, and falsified.
2. Commercially, it establishes D_{CIR} as the core of a *software-first decision layer* (Level 1) that can be licensed to industrial partners for feasibility analysis, sensor design studies, and early-stage R&D decision support.

Throughout, we remain explicit about the current maturity: validation is numerical (simulations and toy models), not yet experimental, and the project is led by a solo founder at TRL 2–3.

2 Detection Theory Background

2.1 Binary Hypothesis Testing

Classical detection theory considers a binary decision between two hypotheses:

$$H_0 : \text{“background only”}, \tag{1}$$
$$H_1 : \text{“background + signal”}. \tag{2}$$

Given an observed random variable Y, one constructs a decision rule $\delta(Y) \in \{H_0, H_1\}$.

The Neyman–Pearson lemma (1933) states that for a fixed false-alarm probability, the most powerful test is based on the likelihood ratio:

$$\Lambda(y) = \frac{p(y \mid H_1)}{p(y \mid H_0)}, \tag{3}$$

and the optimal rule compares $\Lambda(y)$ to a threshold η.

2.2 Gaussian Observation Model

In many physical sensing problems, the observable Y can be approximated as Gaussian under each hypothesis:

$$Y \mid H_i \sim \mathcal{N}(\mu_i, \sigma^2), \qquad i \in \{0, 1\}, \tag{4}$$

with equal variance σ^2 but different means μ_0, μ_1.

For equal priors and symmetric costs, the optimal decision rule is equivalent to a simple threshold on Y:

$$\delta(Y) = \begin{cases} H_1, & \text{if } Y > \dfrac{\mu_0 + \mu_1}{2}, \\ H_0, & \text{otherwise.} \end{cases} \tag{5}$$

The corresponding probability of correct decision is well known and can be expressed via the standard normal cumulative distribution function Φ:

$$P_{\text{succ}} = \Phi\left(\frac{|\mu_1 - \mu_0|}{2\sigma}\right). \tag{6}$$

2.3 From Error Probability to a Decision Metric

In many applications, one would like a single scalar quantity that:

- vanishes when the two hypotheses are indistinguishable,
- increases monotonically with the separation between them,
- maps directly to a success probability.

The Delin Metric D_{CIR} is constructed exactly in this spirit.

3 Definition of the Delin Metric D_{CIR}

3.1 Physical Setting

In the EMDS/CIR context, we consider two *motif classes* $\mathcal{M}_0$ and $\mathcal{M}_1$ (for instance, two different concentrations of a gas, or the presence/absence of a weak molecular trace) embedded in a chaotic molecular environment.

After appropriate coarse-graining and sensor modeling, we assume that an effective scalar observable Y can be approximated as Gaussian:

$$Y \mid \mathcal{M}_i \sim \mathcal{N}(\mu_i, \sigma_y^2), \tag{7}$$

where:

- μ_i is the mean sensor response when motif $\mathcal{M}_i$ is present;
- $\sigma_y^2 = \sigma_x^2 + \sigma_\eta^2$ combines:
  - σ_x^2 = variance of the chaotic field seen by the sensor,
  - σ_η^2 = sensor noise variance.

3.2 Exact Success Probability

Under equal priors and equal variances, the optimal detector is a threshold at $(\mu_0 + \mu_1)/2$, and the exact probability of correct classification is:

$$P_{\mathrm{succ}} = \Phi\left(\frac{|\mu_1 - \mu_0|}{2\sigma_y}\right), \tag{8}$$

with $\sigma_y^2 = \sigma_x^2 + \sigma_\eta^2$.

It is convenient to rewrite this as:

$$P_{\mathrm{succ}} = \frac{1}{2} + \Phi\left(\frac{|\mu_1 - \mu_0|}{\sqrt{2(\sigma_x^2 + \sigma_\eta^2)}}\right) - \frac{1}{2}. \tag{9}$$

3.3 Definition of D_{CIR}

[Delin Metric D_{CIR}] For two motif classes $\mathcal{M}_0, \mathcal{M}_1$ producing Gaussian-like observations with means μ_0, μ_1 and total variance $\sigma_x^2 + \sigma_\eta^2$, the *Delin Metric* is defined as:

$$D_{\mathrm{CIR}}(\mathcal{M}_0, \mathcal{M}_1) := \Phi\left(\frac{|\mu_1 - \mu_0|}{\sqrt{2(\sigma_x^2 + \sigma_\eta^2)}}\right) - \frac{1}{2}. \tag{10}$$

By construction:

- $D_{\mathrm{CIR}} = 0$ corresponds to random guessing ($P_{\mathrm{succ}} = 0.5$);
- $D_{\mathrm{CIR}} \to 0.5$ as the motifs become perfectly separable ($P_{\mathrm{succ}} \to 1$);
- the success probability is simply

$$P_{\mathrm{succ}} = \frac{1}{2} + D_{\mathrm{CIR}}. \tag{11}$$

3.4 CIR Readability Threshold

For practical decision-making, we define a simple criterion:

[CIR Readability] A motif pair $(\mathcal{M}_0, \mathcal{M}_1)$ is said to be *CIR-readable at* 90% *confidence* if

$$D_{\mathrm{CIR}}(\mathcal{M}_0, \mathcal{M}_1) > 0.1, \tag{12}$$

which is equivalent to $P_{\mathrm{succ}} > 0.9$ under the optimal detector.

This provides a clear, operational test in simulations or experiments:

- if $D_{\mathrm{CIR}} \leq 0.1$, the configuration is considered physically too weak or too noisy for reliable motif reading;
- if $D_{\mathrm{CIR}} > 0.1$, the configuration passes a first feasibility filter for EMDS/CIR applications.

4 Ornstein–Uhlenbeck Toy Model Validation

To ground D_{CIR} in an explicit stochastic process, we use a simple Ornstein–Uhlenbeck (OU) model as a toy representation of a chaotic molecular observable.

4.1 OU Dynamics

The OU process $(x_t)_{t\geq 0}$ is defined by:

$$x_t = -\theta x_t \, t + \sigma \, W_t, \tag{13}$$

with:

- $\theta > 0$: relaxation rate,
- $\sigma > 0$: noise intensity,
- W_t: standard Wiener process.

In the stationary regime, we have:

$$x_t \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\theta}\right), \tag{14}$$

with autocorrelation:

$$\mathbb{E}[x_t x_{t+\tau}] = \frac{\sigma^2}{2\theta} e^{-\theta|\tau|}. \tag{15}$$

4.2 Motif Definition in the OU Model

For a minimal binary example, we define two motif classes based on the sign of x_t:

$$\mathcal{M}_0 : x_t < 0, \tag{16}$$

$$\mathcal{M}_1 : x_t > 0. \tag{17}$$

Conditioned on $\mathcal{M}_1$, the stationary OU distribution becomes a half-Gaussian on $(0, \infty)$, with mean:

$$\mu_1 = \mathbb{E}[x_t \mid x_t > 0] = \sqrt{\frac{2}{\pi}} \sqrt{\frac{\sigma^2}{2\theta}} = \frac{\sigma}{\sqrt{\pi\theta}}. \tag{18}$$

Similarly, $\mu_0 = -\mu_1$ under $\mathcal{M}_0$.

The chaotic variance is:

$$\sigma_x^2 = \frac{\sigma^2}{2\theta}. \tag{19}$$

Adding independent Gaussian sensor noise $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$, the total variance becomes:

$$\sigma_y^2 = \sigma_x^2 + \sigma_\eta^2. \tag{20}$$

4.3 Canonical Numerical Example

A canonical parameter choice used in simulations is:

$$\theta = 1, \quad \sigma = 2, \quad \sigma_\eta = 0.5.$$

Then:

$$\sigma_x^2 = \frac{\sigma^2}{2\theta} = 2, \tag{21}$$

$$\mu_1 = \frac{\sigma}{\sqrt{\pi\theta}} \approx 1.128, \tag{22}$$

$$\mu_0 = -\mu_1, \tag{23}$$

$$\sigma_y^2 = \sigma_x^2 + \sigma_\eta^2 = 2.25. \tag{24}$$

Inserting into (10), we obtain:

$$D_{\mathrm{CIR}} \approx 0.488, \tag{25}$$

so that:

$$P_{\mathrm{succ}} = \frac{1}{2} + D_{\mathrm{CIR}} \approx 0.988. \tag{26}$$

Extensive Monte Carlo simulations confirm this prediction, with empirical success rates very close to 98.8% under optimal thresholding.

This example serves as:

- a *numerical proof-of-consistency* of D_{CIR} as a decision metric;
- a concrete reference point for future experimental implementations (e.g. optical or spectroscopic setups emulating OU-like dynamics).

5 Role of D_{CIR} in EMDS/CIR and Level 1 Software

5.1 From Sensor-Level Metric to Decision Engine

Although D_{CIR} is derived at the level of an effective scalar observable, its primary role in the EMDS/CIR roadmap is as a **decision engine**:

- Given simulated or measured statistics $(\mu_0, \mu_1, \sigma_x^2, \sigma_\eta^2)$, it immediately tells whether a configuration is above or below a target success probability.
- It can be embedded into simulation pipelines to scan large parameter spaces (sensor geometry, noise level, motif strength, turbulence intensity) and isolate combinations where $D_{\text{CIR}} > 0.1$.
- It provides a physically interpretable score for feasibility studies: "this configuration has a 93% theoretical success probability under the current assumptions".

5.2 Level 1: Software-First Commercial Layer

In the four-level EMDS/CIR architecture, Level 1 is defined as:

- a **software stack** combining chaotic simulators, motif generators, and baseline ML detectors;
- a D_{CIR} **computation module** and associated tools (thresholding, sensitivity analysis, parameter scans);
- a **decision dashboard** for industrial R&D teams, allowing them to test "what if" scenarios *before* committing to costly hardware prototypes.

From a business perspective, D_{CIR} is the core quantitative ingredient that justifies the existence of such a Level 1 product: it transforms EMDS/CIR from a purely conceptual theory into an actionable decision layer.

5.3 Complementarity with Machine Learning

In realistic settings, ML models (CNNs, transformers, etc.) may be used to approximate the optimal CIR reconstruction operator from data. D_{CIR} does *not* compete with these models; instead, it:

- provides a physics-based upper bound on what any classifier can achieve under given noise and motif statistics;
- helps diagnose whether a poor ML performance is due to algorithmic limitations or fundamental physical constraints;
- offers a compact metric for comparing different sensing configurations, independent of a specific ML architecture.

6 Applications and Limitations

6.1 Potential Application Domains

While EMDS/CIR is still at TRL 2–3, the conceptual applications for D_{CIR} span several sectors:

- **Industrial safety and gas detection:** leak detection in turbulent environments (CH_4, CO_2, NH_3), optimization of sensor placement and noise tolerance.
- **Medical and breath analysis:** feasibility of trace VOC detection for early diagnostics, under realistic noise and sampling constraints.
- **Defense and security:** theoretical limits for trace explosives, CBRN agents, or human VOC signatures in chaotic atmospheres.
- **Environmental and climate sensing:** distinguishability of subtle emission signatures in noisy atmospheric backgrounds.

In each case, D_{CIR} serves as a *feasibility filter*: it can indicate whether a given sensing scenario is fundamentally promising or almost certainly below the distinguishability threshold.

6.2 Current Limitations

To remain scientifically honest, we highlight the main limitations:

- The present work relies on simplified observation models (Gaussian approximations, OU toy dynamics); real environments may require more complex modeling.
- Validation to date is numerical (Monte Carlo simulations), not experimental. No physical EMDS/CIR sensor has yet been built.
- The metric addresses binary motif distinguishability; multi-class settings and continuous motif spaces require extensions.
- Parameter estimation $(\mu_i, \sigma_x^2, \sigma_\eta^2)$ in real systems is itself a nontrivial statistical problem.

These limitations define the roadmap for future work (both scientific and commercial).

7 Conclusion

The Delin Metric D_{CIR} provides a compact, explicit and physically interpretable measure of motif distinguishability in chaotic molecular environments. Rooted in classical Neyman–Pearson detection theory, it:

- maps directly to a probability of correct decision,
- separates chaotic variance from sensor noise,
- defines a practical readability threshold ($D_{\mathrm{CIR}} > 0.1$),
- has been numerically validated on an OU toy model with canonical parameters showing $\approx 98.8\%$ success probability.

Within the EMDS/CIR framework, D_{CIR} is more than a mathematical curiosity: it is the core of a **Level 1 software decision engine**, usable by industrial partners to run feasibility studies, design configurations, and compare scenarios before investing in hardware.

From an intellectual property standpoint, this document, together with its DOI-registered predecessors, establishes prior art for:

- applying classical detection theory to chaotic molecular sensing,
- using D_{CIR}-like metrics as design criteria for future EMDS-ready sensors and platforms,
- building software products that expose D_{CIR} as a decision layer for industrial and scientific users.

The next steps are clear:

1. extend simulations to more realistic turbulent and multi-dimensional models,
2. design laboratory experiments where OU-like or related dynamics can be realized and measured,
3. evolve the Level 1 software stack into a licensable product, embedding D_{CIR} in user-facing tools for R&D teams.

Disclaimer. All claims of performance in this document refer to numerical simulations under explicit model assumptions. No experimental results are implied. This work is intended as a theoretical and computational contribution, forming the basis for future experimental and industrial collaborations.

A Derivation of D_{CIR} for Gaussian Hypotheses

We briefly recall the standard derivation showing how the Delin Metric arises from Gaussian binary detection.

A.1 Setup

Consider:

$$H_0 : Y \sim \mathcal{N}(\mu_0, \sigma_y^2), \tag{27}$$

$$H_1 : Y \sim \mathcal{N}(\mu_1, \sigma_y^2), \tag{28}$$

with equal priors $\mathbb{P}(H_0) = \mathbb{P}(H_1) = 1/2$.

The optimal Bayesian (MAP) detector chooses the hypothesis with higher posterior probability, which is equivalent to a likelihood ratio test:

$$\log \frac{p(y \mid H_1)}{p(y \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} 0.$$

For equal variances, this reduces to a threshold test:

$$y \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\mu_0 + \mu_1}{2}. \tag{29}$$

A.2 Error Probability

By symmetry:

$$P_{\mathrm{err}} = \mathbb{P}(\text{decide } H_1 \mid H_0)\,\frac{1}{2} + \mathbb{P}(\text{decide } H_0 \mid H_1)\,\frac{1}{2}. \tag{30}$$

For H_0:

$$\mathbb{P}(\text{decide } H_1 \mid H_0) = \mathbb{P}\left(Y > \frac{\mu_0 + \mu_1}{2} \,\middle|\, H_0 \right) = \Phi\left(-\frac{\mu_1 - \mu_0}{2\sigma_y} \right).$$

Similarly under H_1. Thus:

$$P_{\mathrm{err}} = \Phi\left(-\frac{|\mu_1 - \mu_0|}{2\sigma_y} \right), \tag{31}$$

and

$$P_{\mathrm{succ}} = 1 - P_{\mathrm{err}} = \Phi\left(\frac{|\mu_1 - \mu_0|}{2\sigma_y} \right). \tag{32}$$

Substituting $\sigma_y^2 = \sigma_x^2 + \sigma_\eta^2$ and rewriting the argument in terms of $\sqrt{2(\sigma_x^2 + \sigma_\eta^2)}$ yields definition (10) and the relation:

$$P_{\mathrm{succ}} = \frac{1}{2} + D_{\mathrm{CIR}}.$$