Kernel Note v1.0

Ryota Sawaki

December 2025

1 Purpose (Kernel Note v1.0)

Purpose

This document **fixes the mathematical definition of "self-imprint" in the SIA framework**.

The purpose of this note is **not** to propose a learning algorithm, optimize performance, or model cognition or consciousness. Instead, it provides a **precise and minimal definition** of what constitutes an imprint, how imprints relate to one another, and how irrelevant or diffuse information is excluded from the self.

In particular, this note defines:

- the structure of imprint candidates (x_i),
- the kernels that determine similarity and imprinting,
- the target space in which self is represented, and
- the constraints that enforce a compact, low-rank self representation.

These definitions are intended to be **stable, reproducible, and implementation-agnostic**. They serve as a **design specification** for SIA implementations, rather than an empirical or performance-driven study.

This document is **orthogonal to the SIA Audit Primitive** and does not extend or modify its claims. Kernel Note v1.0 is limited to **definitions only**, which are frozen at this version.

One-line takeaway:

> **Kernel Note v1.0 fixes the mathematical definition of self-imprint without making claims about learning, performance, or consciousness.**

2 Definition of Imprint Candidates

We define an **imprint candidate** as a minimal episode fragment capturing the internal change produced by a decision event.

Formally, each imprint candidate is represented as:

$$x_i := (o_i,\ a_i,\ \Delta E_i,\ \Delta\tau_i) \tag{1}$$

where:

- o_i denotes a compact representation of the observation or context at event i;
- a_i denotes the action taken at event i;
- ΔE_i denotes the irreversible change in internal imprint energy induced by the event;
- $\Delta\tau_i$ denotes the change in internal temperature or volatility associated with the event.

The index i corresponds to the temporal order of events, and each x_i is assumed to be associated with a timestamp t_i.

Design constraints

1. **Irreversibility.** ΔE_i and $\Delta\tau_i$ cannot be reconstructed from external input alone once the event has passed.
2. **Minimality.** No learning parameters, gradients, rewards, or performance metrics are included.
3. **Implementation independence.** Concrete encodings are left abstract and admissible if consistent.
4. **Event locality.** Each x_i corresponds to a single decision event.

Scope clarification

An imprint candidate is **not** a memory item, belief, concept, or semantic representation. It is a **primitive unit of irreversible internal change**.

One-line takeaway:

> **An imprint candidate preserves irreversible internal change while remaining agnostic to learning and cognition.**

3 Kernel Definition

3.1 Symmetric Kernel (Similarity)

Let

$$\phi : x \mapsto z \in \mathcal{H} \tag{2}$$

be a feature map into an RKHS $\mathcal{H}$.

$$\kappa_{\text{sym}}(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}} = \exp\left(-\frac{\|\phi(x_i) - \phi(x_j)\|^2}{2\sigma^2} \right) \tag{3}$$

Interpretation

κ_{sym} encodes resemblance only; it carries no temporal direction or causality.

One-line takeaway:

> **The symmetric kernel measures similarity but does not define imprint or self.**

3.2 Asymmetric Imprint Kernel (Self-Imprint)

$$\kappa_{\text{imp}}(x_i, x_j) = \kappa_{\text{sym}}(x_i, x_j) \cdot g(x_i) \cdot \mathbf{1}[\, t_i \le t_j \,] \tag{4}$$

with salience gate

$$g(x_i) = \sigma\big(\eta^\top \psi(x_i)\big). \tag{5}$$

By construction,

$$\kappa_{\text{imp}}(x_i, x_j) \neq \kappa_{\text{imp}}(x_j, x_i). \tag{6}$$

One-line takeaway:

> **Self-imprint is a non-symmetric, gated kernel enforcing temporal irreversibility.**

4 Target Space (Representation Space)

Let $K \in \mathbb{R}^{N \times N}$ with

$$K_{ij} = \kappa_{\text{imp}}(x_i, x_j). \tag{7}$$

Self as a low-dimensional structure

Self-imprint is represented by the dominant singular subspace of K.

One-line takeaway:

> **Self-imprint occupies a compact, low-dimensional subspace of a kernel-induced space.**

5 Rank Constraint

Compactness is enforced via nuclear norm minimization:

$$\|K\|_* = \sum_k \sigma_k(K). \tag{8}$$

One-line takeaway:

> **Nuclear norm minimization suppresses diffuse imprinting while preserving dominant modes.**

6 Final-State Irreversibility (band_final)

Definition

For a fixed recovery gain r, sample initial conditions $E_0 \in \mathcal{I}$ and evolve the dynamics for a fixed horizon T. Let $E_\infty(E_0; r)$ denote the terminal self-state. Define

$$\mathrm{band}_{\mathrm{final}}(r) = \max_{E_0 \in \mathcal{I}} E_\infty(E_0; r) - \min_{E_0 \in \mathcal{I}} E_\infty(E_0; r). \tag{9}$$

The critical gain is

$$r^*_{\mathrm{final}} = \arg\max_r \mathrm{band}_{\mathrm{final}}(r). \tag{10}$$

Canonical result

Using the band_final definition with a fixed protocol, the Layer-0/Layer-1 v004 analysis yields

$$r^*_{\mathrm{final}} \approx 0.8, \tag{11}$$

with stable estimates across random seeds under the same protocol.

Robustness

In the linear Layer-0 surrogate, r^*_{final} is deterministic and seed-independent by construction. With history-dependent perturbations, band width may change, but the maximizer of band_final remains governed by the recovery term.

One-line takeaway:

> **Final-state irreversibility (band_final) defines a canonical critical gain $r^*_{\mathrm{final}} \approx 0.8$ for v004-level analysis.**

7 Out-of-Scope Irreversibility Notions

Other notions of irreversibility—such as basin-capture probability or recovery-time divergence—may yield different critical gains. These notions are **out of scope** for Kernel Note v1.0, which fixes irreversibility by final-state spread.

One-line takeaway:

> **Different irreversibility definitions imply different critical gains; this note fixes the final-state definition only.**

8 Scope and Non-Claims

This document defines the **mathematical structure of self-imprint** and nothing more.

Out of scope are learning algorithms, performance benchmarks, cognitive or phenomenological claims, consciousness, and deployment-ready systems.

One-line takeaway:

> **Kernel Note v1.0 defines what self-imprint is, not what it achieves or implies.**