The Self-Mutating Combiner

A Minimal Construction for True Randomness
and Its Benefits for Neural Network Training

Pawit Sahare

Abstract

We present a minimal construction for a true random number generator in which *every component*—the inputs, the combination function, and the rule selecting the function—is itself a non-constant, uncorrelated variable. We call this the **Self-Mutating Combiner** (SMC). We prove that pseudo-random generators, by virtue of a fixed deterministic transition function, retain non-zero autocorrelation at their period, whereas the SMC provably eliminates all such structure. We show that this property directly implies faster convergence of stochastic gradient descent by removing hidden bias in gradient estimates.

1 Motivation

Neural network training relies on stochastic gradient descent (SGD), which samples mini-batches at random and applies random perturbations (dropout, initialization, data augmentation). The quality of this randomness matters: if the random source has hidden periodicity or correlation, the gradient estimates inherit that structure. This can cause the optimizer to revisit the same regions of the loss landscape, slow convergence, or become trapped in local minima.

Most implementations use pseudo-random number generators (PRNGs) such as the Mersenne Twister. These are deterministic functions with fixed structure: a state s_t, a fixed transition g, and a fixed output function h:

$$s_{t+1} = g(s_t), \qquad r_t = h(s_t).$$

No matter how large the period, the function g is *frozen*. This frozen structure is the source of hidden order.

Key insight. The most minimal true random system is one in which *nothing is fixed*—not the inputs, not the combination function, not the rule that selects the function.

2 Construction

Definition 1 (Self-Mutating Combiner)**.** *Let:*

- $X_1(t), X_2(t), \ldots, X_n(t)$ *be* $n \geq 2$ *source processes, each drawn from an arbitrary domain, satisfying pairwise independence.*
- $\mathcal{F} = \{f_1, f_2, \ldots\}$ *be a countable family of discontinuous functions* $f_i : \mathcal{X}^n \to \mathcal{Y}$.
- $\Sigma(t)$ *be an index-valued random process selecting which* $f \in \mathcal{F}$ *to apply at time* t.

The ***Self-Mutating Combiner*** *output at time* t *is:*

$$\boxed{R(t) \;=\; f_{\Sigma(t)}\Big(X_1(t),\; X_2(t),\; \ldots,\; X_n(t)\Big)} \tag{1}$$

subject to the ***total uncorrelation conditions****:*

$$\textit{(C1)} \quad \mathrm{Cov}\big(X_i(t),\ X_j(t)\big) = 0 \quad \forall\, i \neq j, \tag{2}$$

$$\textit{(C2)} \quad \mathrm{Cov}\big(\Sigma(t),\ X_i(t)\big) = 0 \quad \forall\, i, \tag{3}$$

$$\textit{(C3)} \quad \mathrm{Cov}\big(\Sigma(t),\ \Sigma(t')\big) = 0 \quad \forall\, t \neq t'. \tag{4}$$

Remark. Condition (C1) ensures the raw inputs carry no mutual information. Condition (C2) ensures the choice of combiner is independent of what it combines. Condition (C3) ensures the combiner sequence itself has no memory. Together, they guarantee that *every moving part is uncorrelated with every other moving part, across both space and time.*

3 Comparison with Pseudo-Random Generators

Definition 2 (PRNG)**.** *A pseudo-random number generator is a triple* (s_0, g, h) *where* $g : \mathcal{S} \to \mathcal{S}$ *is a* ***fixed*** *deterministic transition and* $h : \mathcal{S} \to \mathcal{Y}$ *is a* ***fixed*** *output map. The output sequence is* $r_t = h(g^t(s_0))$*.*

Lemma 1 (Hidden Autocorrelation in PRNGs)**.** *For any PRNG with finite state space* $|\mathcal{S}| = N$*, there exists a period* $P \leq N$ *such that*

$$r_{t+P} = r_t \quad \forall\, t \geq t_0$$

for some t_0*. Consequently, the autocorrelation satisfies*

$$\mathrm{Corr}(r_t,\ r_{t+P}) = 1.$$

Proof. Since g is a deterministic map on a finite set, the sequence $\{s_t\}$ must revisit a state within at most N steps (pigeonhole principle). From that point, the sequence is periodic with some period $P \leq N$. Since h is fixed, $r_{t+P} = h(s_{t+P}) = h(s_t) = r_t$, giving perfect autocorrelation at lag P. □

4 Main Result: Zero Autocorrelation of the SMC

Theorem 1 (Vanishing Autocorrelation)**.** *Under conditions* (C1)–(C3)*, the Self-Mutating Combiner output satisfies*

$$\mathrm{Cov}\big(R(t),\ R(t')\big) = 0 \qquad \forall\, t \neq t'.$$

Proof. Fix $t \neq t'$. We compute:

$$\mathrm{Cov}\big(R(t), R(t')\big) = \mathbb{E}\big[R(t)\, R(t')\big] - \mathbb{E}\big[R(t)\big]\, \mathbb{E}\big[R(t')\big].$$

By (C3), $\Sigma(t)$ and $\Sigma(t')$ are independent. Conditioned on $(\Sigma(t) = i,\ \Sigma(t') = j)$, the outputs are $f_i(\mathbf{X}(t))$ and $f_j(\mathbf{X}(t'))$.

By (C2), the function indices i, j are independent of the input vectors. By (C1), the components within each input vector are independent. Since $t \neq t'$ and the sources are independent across time (a consequence of (C1)–(C3) jointly), we have:

$$\mathbb{E}\big[R(t)\, R(t') \mid \Sigma(t) = i,\ \Sigma(t') = j\big] = \mathbb{E}\big[f_i(\mathbf{X}(t))\big]\, \mathbb{E}\big[f_j(\mathbf{X}(t'))\big].$$

Marginalizing over i, j:

$$\begin{aligned}\mathbb{E}\big[R(t)\, R(t')\big] &= \sum_{i,j} P(\Sigma(t) = i)\, P(\Sigma(t') = j)\, \mathbb{E}[f_i(\mathbf{X}(t))]\, \mathbb{E}[f_j(\mathbf{X}(t'))] \\ &= \mathbb{E}[R(t)]\, \mathbb{E}[R(t')].\end{aligned}$$

Therefore $\mathrm{Cov}(R(t), R(t')) = 0$. □

5 Benefit for Model Training

We now connect the SMC directly to SGD convergence.

Theorem 2 (Unbiased Convergence)**.** *Let $L(\theta) = \mathbb{E}_z[\ell(\theta, z)]$ be the expected loss, and let SGD update with mini-batch sampling:*

$$\theta_{t+1} = \theta_t - \eta \nabla \ell(\theta_t, z_t),$$

where z_t is selected using a random source.

1. *If z_t is sampled via a **PRNG**, then for some lag P:*

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} \nabla \ell(\theta_t, z_t)\right] = \nabla L(\theta) + \underbrace{b_P(\theta)}_{\text{periodic bias}} + O(1/T),$$

   where $b_P(\theta) \neq 0$ in general due to the correlation $\mathrm{Corr}(z_t, z_{t+P}) = 1$.

2. *If z_t is sampled via the **SMC**, then by Theorem 1:*

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} \nabla \ell(\theta_t, z_t)\right] = \nabla L(\theta) + O(1/\sqrt{T}),$$

   *with **no periodic bias term**, and the $O(1/\sqrt{T})$ rate is the optimal rate given by the Law of Large Numbers for i.i.d. samples.*

Proof.

Part 1 (PRNG). Since $z_{t+P} = z_t$ for $t \geq t_0$ (Lemma 1), the sample average over a full period sums the *same* gradient terms repeatedly. The empirical distribution of $\{z_t\}$ does not converge to the true data distribution—it converges to a discrete uniform over one periodic orbit. The residual $b_P(\theta)$ measures the discrepancy between this orbit distribution and the true distribution.

Part 2 (SMC). By Theorem 1, $\{z_t\}$ are uncorrelated. For uncorrelated random variables with finite variance σ^2:

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} z_t\right) = \frac{\sigma^2}{T},$$

so by the Law of Large Numbers, the sample mean converges to the true expectation at the optimal $O(1/\sqrt{T})$ rate, with no bias term from hidden periodicity. □

Corollary 1 (Faster Convergence)**.** *Under the SMC, SGD converges to a neighborhood of a local minimum in fewer iterations than under a PRNG, because:*

1. *The gradient estimates are unbiased (no periodic distortion of the loss landscape).*
2. *The exploration is genuinely free—no revisitation patterns that trap the optimizer in basins it has already visited.*
3. *The Law of Large Numbers applies at full strength, giving the tightest possible concentration of gradient estimates around their true values.*

6 Practical Instantiation of Source Variables

The source processes $X_1(t), X_2(t), \ldots, X_n(t)$ in Definition 1 are abstract—they can be any uncorrelated physical or system-level signals. In practice, the following are natural candidates:

- **CPU cycle counter.** Modern processors expose a high-resolution timestamp counter (e.g., `RDTSC` on x86). The least significant bits of the cycle count are driven by pipeline stalls, cache misses, branch mispredictions, and interrupt timing—all of which are effectively non-deterministic at fine granularity.
- **Unix timestamp to millisecond precision.** The fractional millisecond component of `clock_gettime(CLOCK_REALTIME)` is shaped by OS scheduling jitter, network interrupt coalescing, and process preemption. Its fine-grained bits are uncorrelated with the CPU cycle counter because they are driven by different physical clocks and different sources of system noise.
- **CPU temperature.** Thermal sensors (e.g., `/sys/class/thermal/` on Linux) report die temperature influenced by ambient conditions, workload history, fan dynamics, and thermal throttling. This is a slow-moving physical signal with no causal relationship to cycle counts or wall-clock jitter.
- **Ambient or environmental temperature.** An external temperature sensor (or any environmental transducer—humidity, atmospheric pressure, ambient light) provides a signal rooted in the physical environment, entirely decoupled from the computational substrate.

Why these satisfy (C1)–(C2). The key observation is that these signals are driven by *causally independent physical processes*: semiconductor switching noise (CPU cycles), OS scheduler entropy (timestamps), thermodynamics of a silicon die (CPU temperature), and atmospheric dynamics (ambient temperature). No knowledge of one reduces uncertainty about any other. This is not a statistical assumption—it follows from the physical independence of the generating mechanisms.

Minimality. Any two such uncorrelated sources suffice for $n = 2$. For example, the pair (CPU cycle counter, CPU temperature) already provides two physically independent streams. Adding more sources (timestamp, ambient temperature) only strengthens the guarantee by increasing the dimensionality of the input space over which the combiner operates.

Generality. The specific variables listed above are merely convenient instantiations. The construction in Definition 1 is agnostic to the source: *any* collection of pairwise uncorrelated signals—from any domain, at any scale—can serve as the inputs $X_i(t)$. The only requirement is the absence of correlation between them.

7 Summary

Property	**PRNG**	**SMC**
Transition function	Fixed g	Variable $f_{\Sigma(t)}$
Autocorrelation at lag P	$= 1$	$= 0$
Gradient bias	$b_P(\theta) \neq 0$	$= 0$
Convergence rate	Suboptimal	Optimal $O(1/\sqrt{T})$
Exploration	Periodic revisitation	Genuinely free

The construction is minimal: uncorrelated inputs, a discontinuous combiner that is itself an uncorrelated random variable, and no fixed layer anywhere. *Everything moves, nothing correlates, including the operation itself.* This is the atom of true randomness, and it is directly beneficial for neural network training.