Article

Super Dual Process Machine Learning

Xiaobin Wang * and April Wang

Posted Date: 7 May 2026

doi: 10.20944/preprints202605.0391.v1

Keywords: dual process; unified machine learning and physics; interacting particles; data integration; scalability; efficiency

 doi:10.20944/preprints202605.0391.v1

Article

Super Dual Process Machine Learning

Xiaobin Wang [1,*] and April Wang [2]

[1] Donaldson Company, 1400 W 94th St, Minneapolis, MN 55431

[2] Northwestern University, 633 Clark St, Evanston, IL 60208

* Correspondence: wangfamaimar@gmail.com

Abstract

This paper introduces the dual process machine learning paradigm, which builds upon the unified machine learning and physics field framework. By integrating machine learning architectures and physics models into a single field-theoretical entity and constructing hidden layers and learning weights based on physical systems, complex machine learning is interpreted as a set of physical interactions. The super dual process machine learning leverages duality relations inherent in physical systems, enabling a simplified "dual" process to replicate the statistical behavior of the original complex "primary" process. We demonstrate that the super dual process opens a new pathway for AI engineering, wherein algebraic structures from underlying physical principles guide model design and computation. We present both the theoretical foundations and practical implementations of super dual machine learning, achieving improved scalability and efficiency compared to traditional methods.

Keywords: dual process; unified machine learning and physics; interacting particles; data integration; scalability; efficiency

1. Introduction

We have introduced a machine learning field framework that unifies machine learning with physics, as detailed in references [1,2]. Conventional AI learning architectures such as Boltzmann machines, LSTM networks, RNNs and transformers, typically learn weights from data and update hidden states based on statistical assumptions and/or mathematical consistency. A key limitation of these methods is that they treat learning and inference as a generic approximation problem, ignoring potential algebraic structure from the underlying physics principles.

Our approach constructs hidden states and connecting weights using physics principles. Rather than representing physical systems with various machine learning structures or embedding physical to machine learning procedures, machine learning architectures and physics models constitute a unified field-theoretical entity. This enables the examination of complex learning processes as physical interactions, potentially yielding novel insights and breakthroughs.

In this manuscript, we introduce the super dual process machine learning paradigm, which is grounded in algebraic structures generated from physics principles. For complex physical systems, a simpler "dual" process can replicate the same statistical behavior as the original complex "primary" process [3–8]. Notably, in interacting particle system scenarios, a limited set of super dual particles can reproduce the collective dynamics of many interacting particles. By leveraging duality relations within a unified machine learning and physics field framework, our approach yields AI architectures that demonstrate substantial improvements in data scalability and learning efficiency relative to traditional machine learning models. The present work details both the theoretical underpinning and practical implementations, illustrating how duality relations foster progress in AI engineering by elucidating the algebraic structure inherent to the system dynamics.

 doi:10.20944/preprints202605.0391.v1

2. Machine Learning Field and Super Duality Process

Recent years have seen significant research achievements at the intersection of machine learning and physics (refer to [9] as an example). Machine learning and AI can contribute to advancements in physics (refer to review article [10] and cited references), just as applying physics principles can generate novel machine learning algorithms and AI paradigms (refer to review article [11] and cited references). For practical applications, combining physics models and machine learning approaches is effective in addressing challenges related to physics model uncertainties, limitations in data, and computational scalability (refer to example references [12–16]).

We propose a field theory view that allows for a single method that combines the physics model and data-driven approaches [1]. Unlike other methods, we do not utilize different machine learning structures to represent physics or embed physics into machine learning models such as neural networks. Instead, machine learning architectures and physics models are represented by the same field-theoretical entity. The machine learning field's probability density functional can be expressed as:

$$P[x] = e^{-E[x]-A[x]}/Z. \tag{1}$$

Here, $x(\vec{r},t)$ is the machine learning field and its probability density functional is a sum of the action functional from the physics model A[x] and the data observation E [x]. Equation (1) can be derived from the Feynman-Kac path integral formula [17,18]. It can also be justified as a Bayesian inference for model data assimilation [19]. $e^{-A[x]}$ is the prior probability from the physics model without observation data and $e^{-E[x]}$ is the conditional probability with observed data inputs. For layered machine learning, as shown in [2], hidden layers and learning weights can be derived from microscopic and mesoscopic physics interactions. Strong machine learning performance can be obtained with limited datasets by capturing physical interactions across scales. This approach also provides predictive capability when dealing with changing, interacting parameters and pattern evolution dynamics.

In this manuscript, we explore an approach using algebraic structures from underlying physical principles to guide AI design and computation. This approach leverages the duality relation in physics systems, where a simplified "dual" process can replicate the statistical behavior of the original, complex "primary" process. Let's illustrate super dual machine learning with the spatial Wright-Fisher model [20]. The Wright–Fisher model describes the evolution of allele frequencies in a finite population under stochastic sampling, commonly referred to as genetic drift. The model can be thought of as the evolution of two alleles (A *and* a) under uncertainty and random influence. The field variable $x(\vec{r},t)$ is the frequency of allele A at spatial location $\vec{r}$ and time t. The allele state space is bounded, with absorbing boundaries at $x = 0$ and $x = 1$ corresponding to the loss or fixation of the allele. The dynamics of alleles evolve according to the spatial stochastic partial differential equation:

$$dx = [D\nabla^2 x + \mu(1-x) - \gamma x]dt + \sqrt{x(1-x)}dW \tag{2}$$

where μ, γ are mutation rates and D is the spatial diffusion coefficient. dW is a Brownian random walk. The generator $\mathcal{L}_x$ of the stochastic process $x(\vec{r},t)$ (2) is:

$$\mathcal{L}_x f(x) = \int d\vec{r}\left\{[D\nabla^2 x + \mu(1-x) - \gamma x]\frac{\delta f(x)}{\delta x} + \frac{1}{2}x(1-x)\frac{\delta^2 f(x)}{\delta x^2}\right\}. \tag{3}$$

Due to its general stochastic structure, the Wright–Fisher framework has been extended beyond population genetics to model analogous dynamics in areas such as finance and social systems, where it describes the evolution of competing states under uncertainty and random influence.

We can write the stochastic partial differential equation (2) in field theoretical path integral format by introducing a conjugate field $\hat{x}(\vec{r},t)$. The path integral corresponding to (2) and (3) is then:

$$\begin{gathered}\int D[x(\vec{r},t)\int D\hat{x}(\vec{r},t)e^{-A(x,\hat{x})} \\ A(x,\hat{x}) = \int dt\, d\vec{r}\hat{x}\cdot(\partial_t x - D\nabla^2 x - \mu(1-x) + \gamma x - \frac{1}{2}x(1-x)\hat{x}^2)\end{gathered} \tag{4}$$

 doi:10.20944/preprints202605.0391.v1

where $A(x,\hat{x})$ is the action functional of the system. For field solution, we usually define a generating functional $e^{\int d\vec{r}\xi(\vec{r})\cdot x(\vec{r},t)}$. The field $x(\vec{r},t)$ solution can be formally obtained through variational derivatives of the generating functional with respect to the source term $\xi(\vec{r})$. This produces arbitrary order statistics of $x(\vec{r},t)$. The expected value of the generating function is:

$$G(x,y) = E_x[e^{\int d\vec{r}x(\vec{r},t)\cdot\xi(\vec{r})}] = \int Dx(\vec{r},t)\int D\hat{x}(\vec{r},t)e^{-A(x,\hat{x})}e^{\int d\vec{r}x(\vec{r},t)\cdot\xi(\vec{r})} \tag{5}$$

where $E_x[\ \ ]$ refers to the average of stochastic process $x(\vec{r},t)$ starting at $x(\vec{r},0)=x$.

It is difficult to directly solve the spatial Wright-Fisher model (2). It is a spatial continuous nonlinear stochastic process. AI inference with data integration is even more challenging because of the additional data functional term $e^{-E[x]}$ in equation (1). To make progress, we seek transformations that transform the complex "primary" system into a much-simplified dual stochastic system. Note that the $\int dt\,\hat{x}\cdot(\partial_t x)$ term in (4) can be written as

$$\int_0^t d\vec{r}dt\,\hat{x}\cdot(\partial_t x) = \int d\vec{r}\,\hat{x}(\vec{r},t)\cdot x(\vec{r},t) - \int d\vec{r}\,\hat{x}(\vec{r},0)\cdot x(\vec{r},0) - \int_0^t d\vec{r}dt\,x\cdot(\partial_t\hat{x}) \tag{6}$$

via integration by parts. This results in the functional action (4) in a format with two boundary terms for the initial time 0 and final time t:

$$A(x,\hat{x}) = \int d\vec{r}\,\hat{x}(\vec{r},t)\cdot x(\vec{r},t) - \int d\vec{r}\,\hat{x}(\vec{r},0)\cdot x(\vec{r},0) + \int dt\,d\vec{r}[-x\cdot\partial_t\hat{x}+\cdots] \tag{7}$$

with the full integrand in (4) being $e^{-A(x,\hat{x})+\int d\vec{r}x(\vec{r},t)\cdot\xi(\vec{r})}$. The boundary term for the final time slice $x(\vec{r},t)$ can be integrated out explicitly. At any specific location $\vec{r}_i$,

$$\int dx(\vec{r}_i,t)e^{-x(\vec{r}_i,t)(\xi(\vec{r}_i)-\hat{x}(\vec{r}_i,t))} = \delta\big(\xi(\vec{r}_i)-\hat{x}(\vec{r}_i,t)\big). \tag{8}$$

This gives the constraint $\xi(\vec{r})=\hat{x}(\vec{r},t)$. With equation (6) and the constraint at the final time slice, we have an action functional for the conjugate field $\hat{x}$ instead of the original field x. The next step is to reverse the time by defining a transformation $\xi(\vec{r},s)=\hat{x}(\vec{r},t-s)$ and $\hat{\xi}(\vec{r},s)=x(\vec{r},t-s)$ with initial and final time slices $\xi(\vec{r},0)=\hat{x}(\vec{r},t)$ and $\xi(\vec{r},t)=\hat{x}(\vec{r},0)$. Noticing that $-x\cdot\partial_t\hat{x}=\hat{\xi}\cdot\partial_t\xi$, we now have a dynamic evolution system for the dual process $\xi(\vec{r},t)$.

The action functional of the dual process can be obtained by a mapping function from $x(\vec{r},t)$ to $\xi(\vec{r},t)$: $A(x,\hat{x})\to A(\xi,\hat{\xi})$. We require that the generator $\mathcal{L}_\xi$ of the dual stochastic process $\xi(\vec{r},t)$ satisfies

$$\mathcal{L}_x e^{\int d\vec{r}x\cdot\xi} = \mathcal{L}_\xi e^{\int d\vec{r}x\cdot\xi}. \tag{9}$$

Noticing that $\xi(\vec{r},t)=\hat{x}(\vec{r},0)$, the remaining boundary condition of the initial time slice $\int d\vec{r}\,\hat{x}(\vec{r},0)\cdot x(\vec{r},0)$ can be obtained as $\int d\vec{r}\,\xi(\vec{r},t)\cdot x(\vec{r})$. We finally obtain equation (5) in terms of the dual stochastic process:

$$G(x,\xi) = E_\xi[e^{\int d\vec{r}x(\vec{r})\cdot\xi(\vec{r},t)}] = \int D\xi(\vec{r},t)\int D\hat{\xi}(\vec{r},t)e^{-A(\xi,\hat{\xi})}e^{\int d\vec{r}x(\vec{r})\cdot\xi(\vec{r},t)} \tag{10}$$

where $E_\xi[\ \ ]$ is the average of the stochastic process $\xi(\vec{r},t)$ starting at $\xi(\vec{r},0)=\xi$.

(5) and (10) give the duality relation of two stochastic processes:

$$E_x[e^{\int d\vec{r}x(\vec{r},t)\cdot\xi(\vec{r})}] = E_\xi\big[e^{\int d\vec{r}x(\vec{r})\cdot\xi(\vec{r},t)}\big]. \tag{11}$$

Equation (11) can also be written in a classical format using duality functions by Taylor expanding and comparing expansion coefficients term by term:

$$E_x\big[x(\vec{r},t)^{\xi(\vec{r})}\big] = E_\xi\big[x(\vec{r})^{\xi(\vec{r},t)}\big]. \tag{12}$$

Constructing a proper dual process representation is critical to simplifying the problem. The duality relation is deeply rooted in the algebraic structures of the underlying physical process [21]. The "primary" and "dual" process are two representations of the same algebraic structure of a physical system. For the Wright-fisher model, a dual process can be obtained with generator

$$\mathcal{L}_\xi g(\xi) = \int d\vec{r}\left\{\left[D\nabla^2\xi\frac{\delta}{\delta\xi}+\mu\xi\left(1-\frac{\delta}{\delta\xi}\right)-\gamma\xi\frac{\delta}{\delta\xi}\right]g(\xi)+\frac{1}{2}\xi^2\left[\frac{\delta}{\delta\xi}-\frac{\delta^2}{\delta\xi^2}\right]g(\xi)\right\}. \tag{13}$$

This is constructed based on the Lie algebra symmetry ($\mathfrak{sl}_2$) of the Wright Fisher model. The "primary" process representation is the continuous process as in (3), and the "dual" process

 doi:10.20944/preprints202605.0391.v1

representation is the discrete jump process with operator (13). Operator (13) has following particle explanation:

$$\begin{matrix} -\mu\xi: & immigration\ (creation\ of\ lineage) \\ (\mu+\gamma)\xi\frac{\delta}{\delta\xi}: & branching\ death\ balance \\ -\xi^2\left[\frac{\delta}{\delta\xi}-\frac{\delta^2}{\delta\xi^2}\right]: & coalescence \end{matrix} . \quad (14)$$

Through duality relation (11) and (12), the solution of the original continuous nonlinear stochastic process $x(\vec{r},t)$ has been transformed to an equivalent but much simpler solution of the finite discrete particle process $\xi(\vec{r},t)$.

The stochastic Wright-Fisher model above provides an example of a duality relation for a physical system. Duality relations link a complex stochastic "primary" process to a simpler "dual" process that maintains important statistical properties. The dual process is often represented as interacting particles. This enables the calculation of expectations or probabilities via the easier dual system, such as using discrete particles instead of continuous processes in the Wright-Fisher model. The super-dual process generalizes this approach, employing richer particle systems—branching, interacting, or weighted—to represent more intricate dynamics and accommodating models with mutation, selection, or nonlinear interactions [22]. Super-duality becomes necessary when interactions included in the system prevent closure within dual particle systems.

When the duality relation is established for physics model action functional A[x] in (1), data integration requires the observed data functional E[x] in (1) to be expressed in terms of the duality relations. In the case of the Wright-Fisher model, allele counts is a natural measurement. For allele counts following a Poisson process with rate proportional to $x(\vec{r},t)$: $dn(\vec{r},t)\sim Poisson(x(\vec{r},t)dt)$, the machine learning field functional (1) can be written as:

$$\int D[x(\vec{r},t)\int D\hat{x}(\vec{r},t)e^{-A(x,\hat{x})-\int dtd\vec{r}x(\vec{r},t)\cdot dn(\vec{r},t)} . \quad (15)$$

The observation data action takes the format of duality function (11). In the discrete format, a measurement with distribution proportional to a power of $x(\vec{r},t)$ (i.e. gamma distribution, etc.) has the format of a classical duality function (12).

In this work, we have employed the Wright-Fisher model—a specific stochastic process—to illustrate the concept of dual process machine learning field solutions. It is important to note that the study of duality relations in Markov processes has uncovered numerous dualities for various stochastic processes, and ongoing research continues to reveal new examples [21]. Building on both theoretical and computational insights from dual Markov processes, it is possible to construct a wide range of dual machine learning field solutions. These solutions deliver improved scalability and computational efficiency by leveraging simplified "dual process" and offer high interpretability through explicit analytical formulations of the governing physical systems and data integration criterion.

3. Super Dual Process Machine Learning

This section presents theory and practical examples of super dual process machine learning, showing how integrating these methods with physics models can improve AI by boosting data scalability and learning efficiency.

3.1. Super Dual Process Data Integration

We will first show the implementation of super dual process model data integration. The hidden state of the Wright Fisher model evolves according to the stochastic spatial differential equation (2). The duality relation (11) and (12) simplifies the continuous nonlinear spatial stochastic process into a discrete particle process. We denote observation data arriving at time t_k and location $\vec{r}_i$ as $y(\vec{r}=\vec{r}_i,t=t_k)=y_{i_k}$, and assume the data distribution with respect to hidden states x_{i_k} is binomial:

 doi:10.20944/preprints202605.0391.v1

$$P(y_{i_k}|x_{i_k}) = {x_{i_k}}^{y_{i_k}} \cdot (1 - x_{i_k})^{n-y_{i_k}}. \tag{16}$$

The duality function of Wright-Fisher for discrete spatial points is $D(x,n) = \prod_i {x_i}^{n_i}$. The observation data distribution (16) contains terms like ${x_{i_k}}^{y_{i_k}}$. We can use the duality relation (12) to integrate the data contribution as:

$$E_x[{x_{i_k}}^{y}] = E_\xi[{x_{i_k}}^{\xi(\vec{r},t)}]. \tag{17}$$

The duality relation offers a powerful simplification for machine learning data integration. When observational data is available at specified times and locations, dual particles are instantiated to correspond precisely to these measurements. These super dual particles then evolve according to the particle system dynamics, as detailed in equation (17), thereby directly linking observational data to the distribution of hidden states. This approach eliminates the necessity of solving complex nonlinear stochastic partial differential equations (see equation 2). Leveraging the dual framework, the computational cost of data integration scales with the number of data points rather than the dimensionality of the spatial domain states. This scaling property stands in stark contrast to conventional methods, which typically require the evolution of stochastic partial differential equations to assimilate observational data. Furthermore, in numerous practical settings, the simplified structure of the dual particle system enables the derivation of closed-form analytic expressions for the posterior likelihood distribution, thereby greatly enhancing computational efficiency. In scenarios involving super dual particle systems with branching and nonlinear spatial interactions, Monte Carlo simulations on the finite dual particle system yield substantial computational gains relative to directly solving the high-dimensional continuous stochastic partial differential equations of the original model.

We consider a spatially extended Wright–Fisher model defined on a one-dimensional lattice of size L. At each site $i \in \{1, \dots, L\}$, the allele frequency $x_i(t) \in [0,1]$ evolves under migration, mutation, and genetic drift. The prior distribution is specified independently across sites as $x_i \sim \mathrm{Beta}(a_0(i), b_0(i)), i = 1, \dots, L$ Sparse observations are available at a subset of sites $\mathcal{O} \subset \{1, \dots, L\}$, where $y_i \sim \mathrm{Binomial}(n_i, x_i), i \in \mathcal{O}$. The objective is to estimate the posterior distribution of the spatial field $\{x_i\}$ given the observations.

We exploit a moment duality between the Wright–Fisher diffusion and a system of interacting particles. Specifically, define the duality function:

$$D(x,n) = \prod_{i=1}^{L} x_i^{n_i^A} (1 - x_i)^{n_i^B}, \tag{18}$$

where n_i^A and n_i^B denote the number of dual particles of type A and B at site i. The duality relation implies $\mathcal{O} \subset \{1, \dots, L\}$,

$$\mathbb{E}_x[D(X_T, n)] = \mathbb{E}_n[D(x, N_T)] \tag{19}$$

where N_t is the dual particle system. Observations are incorporated by initializing the dual system as $n_i^A(0) = y_i, n_i^B(0) = n_i - y_i, i \in \mathcal{O}$, and zero elsewhere. The dual process consists of interacting particles evolving in continuous time with the following transitions: 1) Mutation: Each particle independently changes type: $A \to B$ at rate $\alpha, B \to A$ at rate β. 2) Migration: Each particle performs an independent nearest-neighbor random walk: $i \to i \pm 1$ at rate D, with reflecting boundary conditions. 3) Coalescence: Particles of the same type at the same site coalesce $(A, A) \to A, (B, B) \to B$, with total rate $\frac{1}{2}k(k-1)$ when k particles of the same type occupy a site.

The posterior distribution is approximated via Monte Carlo sampling of the dual process. For each realization $s = 1, \dots, N_{\text{sims}}$, the initial dual particle configuration is constructed from observations. The dual particles evolve up to time T using a time-discretized approximation of the above dynamics. We count the number of particles at each site: $N_i^A(T), N_i^B(T)$, and the posterior is updated according to the jump process of dual particles:

 doi:10.20944/preprints202605.0391.v1

$$x_i \mid N_T \sim \text{Beta}(a_0(i) + N_i^A(T),\, b_0(i) + N_i^B(T)). \tag{20}$$

The sample of the posterior mean field and estimation of posterior statistics is then given by:

$$\begin{aligned} \hat{x}_i^{(s)} &= \frac{a_0(i) + N_i^A(T)}{a_0(i) + b_0(i) + N_i^A(T) + N_i^B(T)}, \\ \mathbb{E}[\, x_i \mid \text{data}\,] &\approx \frac{1}{N_{\text{sims}}} \sum_{s=1}^{N_{\text{sims}}} \hat{x}_i^{(s)}, \\ \text{Var}[x_i \mid \text{data}] &\approx \frac{1}{N_{\text{sims}}} \sum_{s=1}^{N_{\text{sims}}} \left(\hat{x}_i^{(s)} - \bar{x}_i\right)^2 \end{aligned} \tag{21}$$

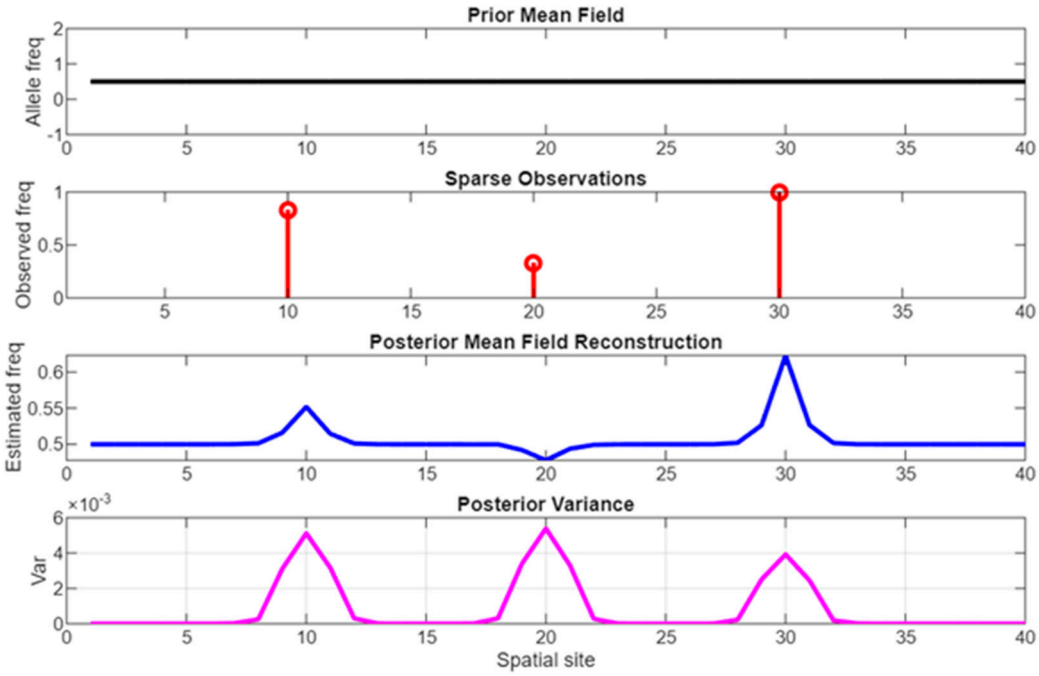

Figure 1. Super dual process machine learning data integration example.

Figure 1 presents a data integration example using the model on a system with 40 sites. The prior mean is initialized as a constant across all sites. Observation data is incorporated at sites 10, 20, and 30. The underlying stochastic Wright–Fisher model employs a specified diffusion coefficient $D = 0.25$, and mutation rate $\mu = 0.05, \gamma = 0.05$. For this study, 4,000 Monte Carlo samples were generated with a simulation time horizon of 1 and a step size of 0.01. Following the assimilation of observations, the majority of sites remain close to the prior mean value of 0.5. Local migration facilitates the spatial propagation of information, while mutation introduces particle type changes, thereby attenuating the influence of observations. Sites distant from observed locations largely retain values near the prior mean. Notable deviations from the prior are observed in the vicinity of the measurement sites. Specifically, the posterior mean appropriately shifts at the observed sites: approximately 0.5537 at site 10 (observation y/n = 5/6), 0.4774 at site 20 (observation y/n = 2/6), and 0.6223 at site 30 (observation y/n = 6/6). For sites where no dual particles survive the backward evolution, the posterior reverts to the prior, resulting in a posterior mean equal to the prior and variance of zero. In contrast, sites near the observation points (10, 20, 30) exhibit small but positive posterior variances, approximately in the range of 0.003–0.005.

This example demonstrates the transformation of data integration for a high-dimensional stochastic system into the simulation of a finite super dual particle system. The super dual process enables dimensionality reduction by circumventing the direct computation of the continuous

 doi:10.20944/preprints202605.0391.v1

stochastic partial differential equation, and it addresses nonlinearity precisely through coalescence, which encodes genetic drift. Notably, the scalability of the solution is determined by the observation data rather than the dimensionality of the state space, and uncertainty is quantified through Monte Carlo sampling. Furthermore, it is important to note that, in cases where data integration is performed without spatial migration, closed-form analytic expressions for the posterior distribution can be derived, significantly enhancing the efficiency of data-driven learning [23,24].

3.2. Super Dual Process Regression

To demonstrate the potential of the super-dual process for machine learning engineering, we introduce a regression framework in which nonlinear features are constructed from the evolution of an interacting particle system. The model leverages Markov duality to define a parametric feature map via expectations of a branching–coalescing process, yielding a structured and trainable alternative to conventional neural networks.

Given an input $\mathrm{x} \in \mathbb{R}^{\mathrm{d}}$, the model predicts:

$$\hat{y}(x) = a^{\top}\phi(x), \tag{22}$$

where $a \in \mathbb{R}^{M}$ are trainable weights and $\phi(x)$ is a feature vector defined through a dual stochastic process. The input is first mapped to a nonnegative latent state:

$$\eta_0 = \mathrm{clip}(W_{\mathrm{in}}x + b) \in (0,1]^{N} \tag{23}$$

where $W_{\mathrm{in}} \in \mathbb{R}^{N\times d}$, $b \in \mathbb{R}^{N}$, and $\sigma(\cdot)$ denotes elementwise clipping to ensure numerical stability. The vector η_0 serves as the initial condition for the dual process.

We define a continuous-time super dual process over particle configurations $k \in \mathbb{N}^{N}$, where k_i denotes the number of particles of type i. The process evolves according to three mechanisms:

$$\begin{aligned} Migratin&: \quad k_a \to k_a - 1, k_b \to k_b + 1 \ \text{ rate } k_a\Lambda_{ab} \\ Coalescence&: \quad k_a \to k_a - 1 \ \text{ rate } \frac{k_a(k_a-1)}{2} \\ B.ranching&: \quad k_a \to k_a + 1 \ \text{ rate } \beta_a k_a. \end{aligned} \tag{24}$$

The state space is truncated to configurations with total particle count $\sum_i k_i \le K_{\max}$, ensuring a finite representation.

Let $\mathcal{S}$ denote the set of admissible particle configurations. The dynamics define a generator matrix $\mathrm{G} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$, where each off-diagonal entry encodes a transition rate and diagonal entries enforce conservation of probability. The transition semigroup over time horizon T is given by:

$$P = \exp(GT). \tag{25}$$

This allows exact computation of expectations over the dual process without sampling.

For an initial particle configuration k_0, define the feature:

$$\phi_{k_0}(x) = \mathbb{E}[\eta_T^{k_T} \mid k(0) = k_0, \eta_0(x)] \tag{26}$$

where $\eta_T^{k_T} = \prod_{i=1}^{N} \eta_{0,i}^{k_{T,i}}$ and k_T is the random terminal state of the particle system. Using the semigroup, this can be computed as:

$$\phi_{k_0}(x) = \sum_{k\in\mathcal{S}} P_{k_0,k} \prod_{i=1}^{N} \eta_{0,i}^{k_i}. \tag{27}$$

In practice, we construct features from first- and second-order initial configurations: First-order: $k_0 = e_i$ and Second-order: $k_0 = e_i + e_j$, yielding a feature vector $\phi(x) \in \mathbb{R}^{M}$ with $M = \mathrm{N} + N(N+1)/2$.

The model is fully differentiable. The gradient of each feature with respect to the input embedding η_0 is:

$$\frac{\partial \phi_{k_0}}{\partial \eta_{0,i}} = \sum_{k\in\mathcal{S}} P_{k_0,k} \cdot k_i \cdot \frac{\prod_j \eta_{0,j}^{k_j}}{\eta_{0,i}}. \tag{28}$$

This yields exact gradients without stochastic estimation. Gradients with respect to parameters W_{in}, b, and a are computed via the chain rule. We train the model using mean squared error over a dataset $\{(\mathrm{x_n}, \mathrm{y_n})\}_{\mathrm{n=1}}^{\mathrm{N}}$:

 doi:10.20944/preprints202605.0391.v1

$$\mathcal{L} = \frac{1}{2N}\sum_{n=1}^{N}(\hat{y}(x_n) - y_n)^2. \quad (29)$$

Parameters W_{in}, b, a are optimized via gradient descent.

The proposed model can be interpreted as a stochastic feature network where nonlinear features arise from interacting particle dynamics, branching increases feature richness, coalescence induces adaptive compression, and migration encodes structured interactions. Unlike conventional neural networks, expressivity is governed by the Markov semigroup of the dual process rather than depth or width.

We present a numerical experiment using a dataset comprising six small input vectors $x \in \mathbb{R}^2$ and target outputs $y \in \mathbb{R}$. The model is trained over 20 iterations with a learning rate of 0.08. . There are $N = 2$ neurons. Branching rates are $\beta = [0.4,0.3,0.5]$. Maximum number of dual particles is set to $K_{\max} = 3$. Figure 2 illustrates the progression of the training loss and the evolution of weight values across iterations, demonstrating that the training begins to converge after approximately five iterations. Table 1 provides a comparative summary of the predicted outputs and the ground truth values.

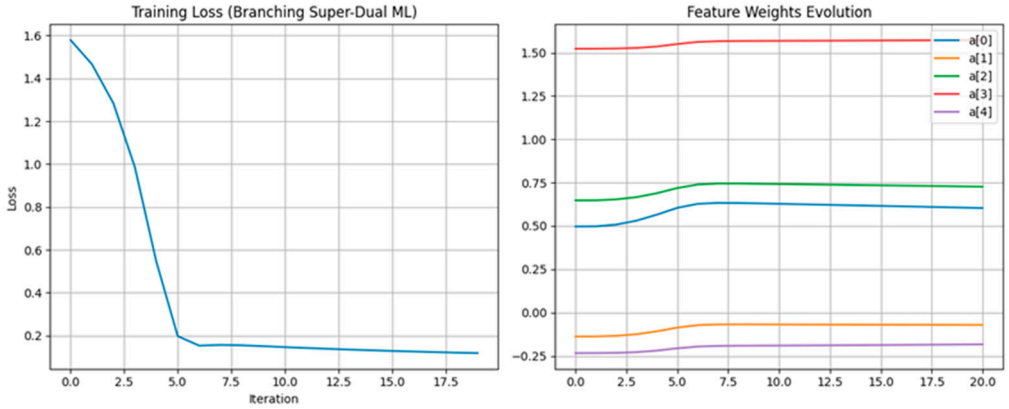

Figure 2. Dual process machine learning training loss and the evolution of weight values as a function of iterations.

Table 1. Dual process regression predicted outputs and the ground truth value.

Input	Predicted	True
[0.2,0.8]	1.53	1.2
[0.5,1]	1.91	2.0
[1.5,1]	1.96	1.5
[1,1]	2.17	3.0
[0,1]	1.41	0.8
[0.2 0.5]	1.18	1.3

For the proof of concept, we demonstrate the super-dual ML framework in the analytically tractable setting of an interacting particle system, where dual expectations can be computed exactly via matrix exponentiation. The primary computational cost arises from constructing the generator G and computing the matrix exponential $P = \exp(GT)$. This approach is tractable for moderate particle limits $K_{\max}$, but scales combinatorially with system size. For larger systems, Monte Carlo sampling-based approximations or sparse representations of G may be employed.

3.3. Super Dual Process Nested Network

Deep learning has achieved remarkable success in modeling complex data, yet principled inference and uncertainty quantification in stochastic dynamical systems remain challenging. Sequential latent-variable models--including hidden Markov models, state-space models, and more

 doi:10.20944/preprints202605.0391.v1

recently neural stochastic differential equations (neural SDEs)—require efficient inference of latent states from noisy observations [25–27]. Standard approaches rely on particle filtering, which suffers from high variance and degeneracy in high dimensions, or variational inference, which introduces approximation bias and often lacks interpretability. A key limitation of these methods is that they treat inference as a generic approximation problem, ignoring potential algebraic structure in the underlying physics principles.

Super dual process machine learning based upon algebraic structure generated from physics principles provides an alternative paradigm. Instead of directly approximating the latent state distribution, one computes expectations via duality relations, which are often discrete and low-dimensional. Here, we propose a nested super dual process network as a hierarchical (multi-layer) stochastic system of the form:

$$X_1(t) \to X_2(t) \to \cdots \to X_L(t) \to Y(t), \tag{30}$$

where each latent process $X_i(t)$ evolves according to a stochastic differential or partial differential equation, and observations $Y(t) \sim p\big(Y \mid X_L(t)\big)$ depend only on the terminal layer. The objective is to compute the posterior over upstream latent states $p(X_1(t) \mid Y(0{:}\,t))$.

For each process $X_i(t)$, we identify dual process $N_i(t)$ based upon the algebraic structure of the dynamic system and obtain a duality function $H_i(x,n)$ such that

$$\mathbb{E}[H_i(X_i(t),n) \mid X_i(0) = x] = \mathbb{E}[H_i(x,N_i(t)) \mid N_i(0) = n]. \tag{31}$$

Typical choices include the polynomial basis $H(x,n) = x^n$ and the exponential basis $H(x,n) = e^{nx}$. The duality functions define a feature map $\phi_i(x) = [H_i(x,0), H_i(x,1), \dots]$ that embed latent states into a functional feature space. Inference is thus performed in feature space rather than directly in state space $\mathrm{H_i(x,n)}$. Inference proceeds by propagating information backward through the cascade. The likelihood expansion of the final layer is:

$$p(Y \mid X_L) \approx \sum_{n_L=0}^{N_{\max}} c_{n_L}(Y)\, H_L(X_L, n_L). \tag{32}$$

Information propagates through layer $\mathrm{L} \to \mathrm{L}-1$ by the duality relation

$$\mathbb{E}[H_L(X_L,n_L) \mid X_{L-1}] = \sum_{n_{L-1}} g^{(L)}_{n_L \to n_{L-1}}\, H_{L-1}(X_{L-1}, n_{L-1}). \tag{33}$$

This gives a likelihood for the first layer:

$$p(Y \mid X_1) \approx \sum_{n_1} w_{n_1}(Y)\, H_1(X_1, n_1) \tag{34}$$

with weights:

$$w_{n_1}(Y) = \sum_{n_L,\dots,n_2} c_{n_L}(Y) \prod_{i=2}^{L} g^{(i)}_{n_i \to n_{i-1}}. \tag{35}$$

For this nested dual architecture, exact duality requires a likelihood in the span of the duality function: $p(Y \mid X_L) \in \mathrm{span}\{H_L(x,n)\}$. The generic likelihood can be approximated via projection onto the dual basis:

$$\mathrm{p}(Y \mid X_L) \approx \sum\nolimits_{n_L=0}^{N_{\max}} c_{n_L}(Y)\, H_L(X_L, n_L). \tag{36}$$

Coefficients $c_{n_L}(Y)$ can be obtained via orthogonal expansion, momentum match, series expansion or machine learning encoders.

As an implementation example, we consider a three-layer dual processes cascade consisting of a Wright–Fisher diffusion $X_1(t)$, a conditionally coupled Wright–Fisher-type diffusion $X_2(t)$, and Gaussian observations $Y(t) \mid X_2(t)$. The first layer evolves according to

$$dX_1(t) = \sqrt{X_1(t)(1-X_1(t))}\, dB_1(t), X_1(t) \in [0,1], \tag{37}$$

with classical duality to the coalescent $N_1(t)$, characterized by the polynomial duality function $H_1(x_1,n_1) = x_1^{n_1}$. The second layer is driven by $X_1(t)$through a mean-reverting Wright–Fisher-type SDE,

 doi:10.20944/preprints202605.0391.v1

$$dX_2(t) = \kappa(X_1(t) - X_2(t))\,dt + \sqrt{X_2(t)(1-X_2(t))}\,dB_2(t), X_2(t) \in [0,1], \tag{38}$$

which admits a moment dual $N_2(t)$ with duality function $H_2(x_2, n_2) = x_2^{n_2}$, but with time-inhomogeneous rates induced by the path of X_1. Finally, observations are Gaussian:

$$Y(t) \mid X_2(t) \sim \mathcal{N}(X_2(t), \sigma^2), \ell(x_2; Y) \propto \mathrm{e}^{-\frac{(Y-x_2)^2}{2\sigma^2}}. \tag{39}$$

Expanding the likelihood in the polynomial basis $x_2^{n_2}$,

$$\ell(x_2; Y) = \sum_{n_2 \geq 0} c_{n_2}(Y)\, x_2^{n_2} \tag{40}$$

where the coefficients $c_{n_2}(Y)$ correspond to (rescaled) Hermite-type expansions. Applying the duality of X_2, each moment propagates through the time-inhomogeneous dual process:

$$E[X_2(t)^{n_2} \mid X_1] = E\left[X_2(0)^{N_2(t)} \mid X_1\right] \tag{41}$$

where N_2 evolves with rates depending on the path $X_1(s)$, $s \leq t$. This induces a polynomial representation in $X_1(t)$,

$$E[\,\ell(X_2(t); Y) \mid X_1\,] = \sum_{n_1 \geq 0} w_{n_1}(Y, t)\, X_1(t)^{n_1}. \tag{42}$$

The outer layer is then treated via the coalescent duality of X_1. Specifically, for $H_1(x_1, n_1) = x_1^{n_1}$,

$$E[X_1(t)^{n_1} \mid X_1(0)] = E\big[X_1(0)^{N_1(t)} \mid N_1(0) = n_1\big], \tag{43}$$

where $N_1(t)$is a pure death process with transition probabilities $P_{n_1 \to m}(t)$. Combining both dual layers yields the mixture representation of the filtering distribution of $X_1(t)$:

$$p = (X_1(t) \mid Y(t)) = \sum_{n_1 \geq 0} \widetilde{w}_{n_1}(Y, t)\, \mathrm{Beta}(\alpha_0 + n_1, \beta_0 + N - n_1), \tag{44}$$

where the mixture weights admit the nested dual representation

$$\widetilde{w}_{n_1}(Y, t) = \sum_{n_2 \geq 0} c_{n_2}(Y) \sum_{m \geq 0} P^{(1)}_{m \to n_1}(t)\, P^{(2)}_{n_2 \to m}(t). \tag{45}$$

Here $P^{(2)}_{n_2 \to m}(t)$ denotes the transition kernel of the time-inhomogeneous dual N_2 (driven by X_1), while $P^{(1)}_{m \to n_1}(t)$is the classical Kingman coalescent transition kernel. Thus, the weights $\widetilde{w}_{n_1}(Y, t)$ arise as a double convolution of (i) Gaussian likelihood expansion coefficients and (ii) two nested dual semigroups corresponding to the hierarchical Wright–Fisher structure.

The proposed nested dual processes framework forms deep architecture for super dual process machine learning. Duality functions act as basis features, coefficients act as learned representations, and dual propagation acts as a deterministic inference network. This yields a new class of deep sequential latent-variable machine learning model learning that is based upon interpretable algebraic structure, avoids Monte Carlo sampling, and enables compositional, multi-layer inference.

4. Discussion

This work introduces an innovative paradigm for machine learning that integrates physics-based field theory with artificial intelligence learning architectures, leveraging duality principles to construct computationally efficient models. The core concept—transforming a complex stochastic system into a corresponding dual process—provides both conceptual clarity and practical advantages. Across the presented examples, including data integration, regression, and hierarchical architectures, the super dual framework consistently demonstrates its capacity to reformulate high-dimensional, nonlinear inference problems into tractable particle-based or algebraically structured computations.

A notable advantage of the proposed methodology is its scalability with respect to data size, rather than the dimensionality of the state space. Conventional machine learning approaches, particularly those that involve stochastic partial differential equations or high-dimensional latent variables, are often hindered by the curse of dimensionality. In contrast, the super dual process reformulates inference in terms of finite particle systems or dual representations whose complexity

 doi:10.20944/preprints202605.0391.v1

is primarily governed by the number of observations, resulting in markedly improved computational efficiency, especially in sparse-data scenarios.

Another significant aspect is interpretability. By grounding machine learning architectures in physical principles and dual stochastic processes, model components such as branching, coalescence, and migration acquire clear mechanistic interpretations. This contrasts with conventional deep learning models, wherein learned representations often lack transparency. In regression and nested network examples, feature construction is explicitly tied to the evolution of dual processes, establishing a connection between model architecture and underlying dynamics.

The framework further presents a unified perspective on probabilistic inference. Utilizing field-theoretic formulations and path integral representations, Bayesian inference, stochastic dynamics, and machine learning are expressed within a cohesive mathematical structure. Duality serves as a bridge, enabling efficient computation of expectations and posterior distributions. Notably, the nested super dual architecture illustrates how multi-layer inference can be performed deterministically through algebraic propagation, thus avoiding the high variance associated with particle filtering and the bias inherent in variational approximations.

Despite these advantages, several challenges persist. Firstly, while dual representations simplify numerous problems, constructing suitable dual processes remains model-dependent and may not be straightforward for all systems. Secondly, in scenarios where the dual particle system expands significantly—such as through branching—the computational demands can become substantial, necessitating approximations like truncation, sparse representations, or Monte Carlo sampling. Thirdly, for general likelihood models, approximating observations within the span of duality functions introduces an additional layer of approximation that may impact accuracy.

We have applied dual process machine learning to develop models for predicting the spread of chemical pollution in extensive aquatic environments, where limited IoT devices are deployed at specific locations. Additionally, this method has been utilized to construct AI models with stochastic features related to mesoscopic phase transition phenomena and electronic device switching.

Future research directions include developing systematic approaches for identifying dual processes, exploring adaptive particle representations, and integrating learning-based approximations for dual expansions. Moreover, extending the framework to encompass broader classes of stochastic systems, such as non-Markovian dynamics and high-dimensional continuous spaces, would further increase its versatility and applicability.

5. Conclusions

In this paper, we introduced the super dual process machine learning framework, a new paradigm that unifies machine learning and physics through a field-theoretic formulation and exploits duality relations to enable efficient computation and scalable inference. By representing learning systems as physical fields and transforming complex stochastic dynamics into simpler dual processes, the framework provides both theoretical insight and practical computational benefits.

We developed the theoretical foundation of the approach using path integral formalism and demonstrated how duality relations can be constructed and utilized in the context of the spatial Wright–Fisher model. Building on this foundation, we presented three key implementations: (1) super dual data integration, which enables efficient Bayesian inference via particle-based dual systems; (2) super dual regression, where nonlinear features arise from interacting particle dynamics; and (3) nested super dual networks, which extend the framework to hierarchical and deep learning settings with structured inference.

Across these examples, the proposed approach achieves dimensionality reduction, improved scalability, and enhanced interpretability, while maintaining rigorous probabilistic foundations. The ability to compute expectations through dual processes offers a powerful alternative to conventional sampling-based or approximation-based inference methods.

Future research will focus on expanding the class of models suitable for dual representations, improving computational techniques for large-scale systems, and exploring applications across

 doi:10.20944/preprints202605.0391.v1

scientific and engineering domains. The integration of duality-based methods with modern machine learning tools may lead to a new class of AI systems that are both computationally efficient and physically interpretable.

6. Patents

NA.

Author Contributions: Xiaobin Wang contributed on conceptualization, methodology and analysis. April Wang contributed on conceptualization, methodology, software and validation.

Funding: NA.

Institutional Review Board Statement: NA.

Informed Consent Statement: NA.

Data Availability Statement: Data is simulated and simulation is explained in the article.

Acknowledgments: NA.

Conflicts of Interest: NA.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
LSTM	Long short term memory
RNN	Recurrent neural network
PDE	Partial differential equation
SDE	Stochastic differential equation

References

1. Wang, X. and Wang, A., Analytic field solution for machine learning integrating physics model and data driven approach. *AIP Advances*, *15*(5), 2025
2. Wang, X., & Wang, A., Unifying Machine Learning and Physics Models Through a Mesoscopic Field Approach. Global Journal of Computer Science and Technology, 25(D1), 2025
3. Swart, Jan M. "A course in interacting particle systems." arXiv preprint arXiv:1703.10007, Published by Cambridge, 2026
4. Spitzer, F.: Interaction of markov processes. In: Random Walks, Brownian Motion, and Interacting Particle Systems, pp. 66–110. Springer, Berlin, 1991
5. Giardina, Cristian, et al. "Duality and hidden symmetries in interacting particle systems." Journal of Statistical Physics 135.1 25-55, 2009
6. Liggett, T.M.: Interacting Particle Systems, vol. 276. Springer Science & Business Media, Berlin, 2012
7. Kolokoltsov, V. and Lee, R.X., Stochastic duality of Markov processes: a study via generators. Stochastic Analysis and Applications, 31(6), pp.992-1023, 2013
8. Floreani S, Jansen S, Wagner S. Intertwinings for continuum particle systems: An algebraic approach. SIGMA. Symmetry, Integrability and Geometry: Methods and Applications,20:046, 2024
9. Ra, Donghyun, et al. "Physics-informed machine learning across manufacturing processes: Recent advances, challenges, and directions." Journal of Manufacturing Systems 85 72-95, 2026
10. Suresh R. et.al., Revolutionizing Physics: a Comprehensive Survey of Machine Learning Applications, Frontiers in Physics, Volume 12, 1322162, 2024.
11. Jiao. L et. al. AI meets Physics, A Comprehensive Survey, Artificial Intelligence Review 57:256, Springer, 2024.

 doi:10.20944/preprints202605.0391.v1

12. Sharma, P. et.al., A Review of Physics-Informed Machine Learning in Fluid Mechanics. Energies 16, 2343, 2023.
13. Leon C. & Scheinker A., Physics-constrained Machine Learning for Electrodynamics without Gauge Ambiguity Based on Fourier transformed Maxwell's Equations, Scien-tific Reports 14:14809, 2024
14. Wu Y. et. al., Combining Physics-based and Data-driven Models: Advancing the Fron-tiers of Research with Scientific Machine Learning, Expert Systems With Applications 255 124678, 2024
15. Sanderse, Benjamin, Panos Stinis, Romit Maulik, and Shady E. Ahmed. Scientific machine learning for closure models in multiscale problems: a review. arXiv preprint arXiv:2403.02913, 2024
16. Lee J, et. al., Editorial: Integrating Machine Learning With Physics Based Modeling of Physiological Systems. Front. Physiol. 16:1562750, 2025
17. Quinn J.C. and Abarbanel H., Data assimilation using a GPU accelerated path integral Monte Carlo approach, vol. 230, issue 22, pp. 8168-8178, Journal of Computational Physics, 2011
18. Abarbanel, Henry DI, The statistical physics of data assimilation and machine learning. Cambridge University Press, 2022
19. Cheng Y. and Reich S., A McKean optimal transportation perspective on Feynman-Kac formulae with application to data assimilation, Frontiers in Applied Dynamical Systems, http://arxiv.org/abs/1311.6300, 2015
20. Huillet, Thierry. "On Wright–Fisher diffusion and its relatives." Journal of Statistical Mechanics: Theory and Experiment 2007.11 P11006-P11006, 2007
21. Giardina, C. and Redig, F.,. Duality for Markov processes: a Lie algebraic approach. Springer Nature, 2026
22. Etheridge, Alison. An introduction to superprocesses. Vol. 20. Providence, RI: American Mathematical Society, 2000.
23. Chaleyat-Maurel, M., & Genon-Catalot, V. Filtering the Wright-Fisher diffusion. ESAIM: Probability and Statistics, 13, 197-217, 2009
24. Papaspiliopoulos, O., Ruggiero, M., & Spano, D. (2014). Filtering hidden Markov measures. arXiv preprint arXiv:1411.4944,2014.
25. Hoffman, Matthew D., et al. "Stochastic variational inference." *Journal of machine learning research*, 2013.
26. Margossian, Charles C., and David M. Blei. "Amortized variational inference: When and why?." *arXiv preprint arXiv:2307.11018*, 2023.
27. Li, Xuechen, et al. "Scalable gradients for stochastic differential equations." *International conference on artificial intelligence and statistics*. PMLR, 2020.