Purely Discrete Symbolic Regression: EML Signal Decomposition via MCTS and Basin-Hopping

Authors' names omitted for anonymous review

***Abstract*—Extracting closed-form analytical representations of temporal dynamics from noisy data remains a central challenge. Traditional symbolic regression suffers from highly irregular search spaces and "expression bloat." While the isomorphic Exp-Minus-Log (EML) operator unifies the search space, its deep nesting triggers gradient explosion, causing continuous approximations to stall at local minima. To address this, we propose the Ultimate EML Symbolic Decomposition Transform (U-ESDT), a white-box 1D signal decomposition framework. U-ESDT employs a dual-track global optimization engine: an outer Monte Carlo Tree Search (MCTS) conducts purely discrete, deterministic topology generation, while an inner Basin-Hopping algorithm achieves global parameter annealing. Enhanced by affine-wrapped variables and a two-stage refinement mechanism, U-ESDT optimally balances exploratory breadth and convergence depth. Evaluated on benchmark physical signals and the Feynman dataset, U-ESDT demonstrates unparalleled topological parsimony and noise robustness. It completely eradicates expression bloat while achieving predictive fidelities that match or exceed state-of-the-art heuristic models (e.g., PySR), offering a rigorous paradigm for deciphering the mathematical laws of non-stationary time series.**

***Keywords*— Symbolic Regression, Exp-Minus-Log (EML) Operator, Monte Carlo Tree Search, Basin-Hopping, Signal Decomposition, Time Series Analysis**

I. Introduction

In complex dynamic systems, the extraction of underlying kinematic equations and temporal evolution laws from noisy observational data remains a fundamental challenge in scientific discovery. Traditional digital signal processing techniques (such as the Fourier transform, wavelet transform, and empirical mode decomposition) rely on predefined basis functions to decompose 1D temporal or spatial signals into discrete spectral coefficients or finite modes. While these methods have achieved tremendous success in feature extraction, they fundamentally yield opaque, "black-box" data representations and cannot directly output the closed-form analytical equations governing signal generation.

Symbolic Regression (SR) aims to simultaneously discover the optimal mathematical structure and corresponding parameters from data. However, traditional SR methods are constrained by large, heterogeneous libraries of mathematical operators (including basic arithmetic, trigonometric functions, exponentials, etc.), resulting in search spaces that are highly irregular and notoriously difficult to optimize. Recently, Odrzywołek demonstrated that the Exp-Minus-Log (EML) operator, defined as $eml(x, y) = \exp(x) - \ln(y)$, can function as a universal "NAND gate" for continuous mathematical systems [1] . Through the recursive self-nesting of this single operator, all standard elementary mathematical functions can be reconstructed. This breakthrough establishes a completely isomorphic binary tree search space for SR.

Despite the EML architecture significantly unifying the topological structure of mathematical expressions, existing optimization strategies—such as continuous relaxation and soft routing based on deep learning frameworks—inevitably suffer from gradient explosion, numerical collapse, and a susceptibility to local optima when applied to deeply nested EML trees. Consequently, these methods are forced to compromise on either accuracy or interpretability.

To overcome these critical limitations, this paper introduces a rigorous signal decomposition framework: the **Ultimate EML Symbolic Decomposition Transform (U-ESDT)**. The primary contributions of this work are threefold:

- **Purely Discrete Deterministic Topology Search**: We introduce Monte Carlo Tree Search (MCTS) to the EML architecture for the first time, completely discarding heuristic random mutations and soft routing. This enables a rigorous traversal of formula structures with theoretical convergence guarantees.
- **Global Constant Space Annealing**: To tackle the highly non-convex error surfaces induced by EML nesting, we integrate the quasi-Newton method (L-BFGS-B) with the Basin-Hopping algorithm. This approach effectively escapes local minimum traps, ensuring high-precision parameter fitting.
- **Time-Varying White-Box Equation Extraction**: We propose an Overlap-Add sliding window ESDT framework for time-domain processing, achieving the first continuous, white-box extraction of locally stationary dynamic physical equations from time-series signals.

II. Background and Related Work

A. The Curse of Dimensionality and Expression Bloat in Traditional Symbolic Regression

Given a dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ with $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, Symbolic Regression (SR) poses the discrete optimization problem of finding an optimal function f^* from a hypothesis space of mathematical expressions $\mathcal{H}$:

$$f^* = \arg\min_{f \in \mathcal{H}} \mathcal{L}\left(f(\mathbf{X}), \mathbf{Y}\right) + \lambda \Omega(f), \tag{1}$$

where $\mathcal{L}$ is a fitting metric (e.g., MSE) and $\Omega(f)$ measures structural complexity. Mainstream traditional SR tools, ranging from early benchmark software like Eureqa to widely adopted open-source libraries like PySR [2], primarily rely on Genetic Programming (GP [3]). In GP, the search space

$\mathcal{H}$ is constructed from a highly heterogeneous operator dictionary $\mathcal{O} = \{+, -, \times, \div, \sin, \exp, \dots\}$ and a terminal set $\mathcal{V} = \{x_1, \dots, x_d\} \cup \mathbb{R}$.

Because the functional properties—namely, domains, codomains, and Lipschitz constants—of these operators vary drastically, the discrete algebraic manifolds within $\mathcal{H}$ are highly irregular. Consequently, stochastic GP crossover and mutation operations frequently project candidate solutions outside the feasible domain of underlying physical logic, yielding discontinuous or undefined structures. More critically, GP algorithms invariably suffer from "expression bloat" as the generation index $k \to \infty$. This is mathematically characterized by a divergence in topological complexity, $\mathbb{E}[|f_k|] \to \infty$, coupled with a vanishing marginal fitness gradient, $\Delta\mathcal{L}(f_k) \to 0$. This over-parameterization inflates the structural penalty $\Omega(f)$ without bound, trapping the search in "pseudo-formulas" that overfit local noise and lack physical interpretability.

Recent advancements in evolutionary computation have proposed various mechanisms to control this code bloat, such as double lexicase selection and modular multitree architectures [4], [5]. However, these methods primarily act as post-hoc structural regularizers on heterogeneous spaces, rather than resolving the topological irregularity at its mathematical root.

B. Recent Academic Solutions to Traditional Limitations

To mitigate the ill-posedness of searching across heterogeneous operator spaces and the subsequent expression bloat, recent literature has introduced several theoretically grounded paradigms, which can be broadly categorized into four mathematical frameworks:

- **Deep Symbolic Regression (DSR):** Petersen et al. [6] reformulated SR as a Markov Decision Process (MDP) parameterized by Recurrent Neural Networks (RNNs). By mapping the discrete tree generation to a sequence of actions $\boldsymbol{\tau}$, DSR maximizes the expected reward via Policy Gradient: $J(\theta) = \mathbb{E}_{\boldsymbol{\tau}\sim\pi_\theta}[R(\boldsymbol{\tau})]$. By leveraging the autoregressive prior π_θ, DSR bypasses the combinatorial explosion of random genetic mutations, effectively constraining the search trajectory within high-reward subspaces.
- **Physics-Informed SR (StruSR):** To embed physical priors, Gong et al. [7] introduced the StruSR framework. Let $\mathcal{N}[f] = 0$ denote the governing partial differential equation. StruSR employs pre-trained Physics-Informed Neural Networks (PINNs) to approximate local Taylor expansions, extracting spatial and temporal gradients. This introduces a structural regularization term $\mathcal{L}_{physics} = \|\mathcal{N}[f]\|_2^2$ during the attribution phase, constraining the mutation operators to a physically plausible manifold and thereby truncating the search paths that lead to meaningless bloat.
- **Bayesian Model Selection for Regularization:** To formalize parsimony, researchers have transitioned to Maximum A Posteriori (MAP) estimation. By treating $\mathcal{D}$ as noisy observations, this approach optimizes the posterior probability:

$$p(f \mid \mathcal{D}) \propto p(\mathcal{D} \mid f) \exp(-\alpha\Omega(f)), \tag{2}$$

  where the prior distribution strictly penalizes the L_0-norm of the tree nodes. This imposes a rigorous information-theoretic bound, compelling the evolution of parsimonious formulas commensurate with the intrinsic Kolomogorov complexity of the signal.
- **Counterexample-Driven Formal Constraints:** To further formalize the generation process, researchers have explored counterexample-driven genetic programming, utilizing satisfiability modulo theories (SMT) to enforce strict formal constraints (e.g., symmetry, monotonicity) on the evolved symbolic models [8]. While providing rigorous mathematical guarantees, such methods significantly escalate computational overhead.

However, while these state-of-the-art relaxations smooth the optimization landscape, they remain fundamentally constrained by the non-convexity of the heterogeneous operator library $\mathcal{O}$. Whether approximating the search space via policy distributions or continuous relaxation, the underlying hypothesis space remains a disjoint union of non-homeomorphic functional spaces (e.g., coupling sinusoidal periodicity with exponential divergence). When confronting deep topological nesting $D \gg 1$ required for complex physical laws, the global objective function remains highly non-convex, possessing myriad saddle points and local minima. Consequently, these approximations fail to guarantee global asymptotic convergence.

To systematically resolve this topological bottleneck, this paper leverages the completely isomorphic EML operator to propose U-ESDT—a purely discrete, global annealing architecture. This paradigm shifts the SR objective from a heuristic continuous approximation back to a rigorous, deterministic combinatorial resolution defined on a homogeneous basis.

C. The EML Operator and Isomorphic Mathematical Spaces

The core mathematical innovation of the EML operator is the reduction of the heterogeneous dictionary $\mathcal{O}$ to a singleton binary operator $\mathcal{O}_{EML} = \{\otimes_{eml}\}$. Let $f(\mathbf{x}, t)$ be an arbitrary analytic physical law; the EML framework asserts that it can be exactly reconstructed via the closure of the following Context-Free Grammar (CFG), defined by the tuple $\mathcal{G} = (V, \Sigma, R, S)$:

$$S \to \mathcal{E}(S, S) \mid t \mid c, \tag{3}$$

where t is the independent variable, $c \in \mathbb{R}$ is a constant, and $\mathcal{E}(\cdot, \cdot)$ denotes the EML operation. This homogeneity fundamentally fundamentally transforms SR from an ill-posed "operator classification problem" into a "binary tree topology growth problem" governed by strict topological isomorphisms. For instance, as illustrated in Figure 1, the non-linear singularity of the natural logarithm $\ln(x)$ can be rigorously mapped to a specific invariant topological nesting: $\ln(x) \equiv \mathcal{E}(1, \mathcal{E}(\mathcal{E}(1, x), 1))$.

D. Limitations of Gradient-Based EML Optimization: A Rigorous Analysis

Early optimization strategies for the EML operator attempted to employ Continuous Relaxation, projecting the

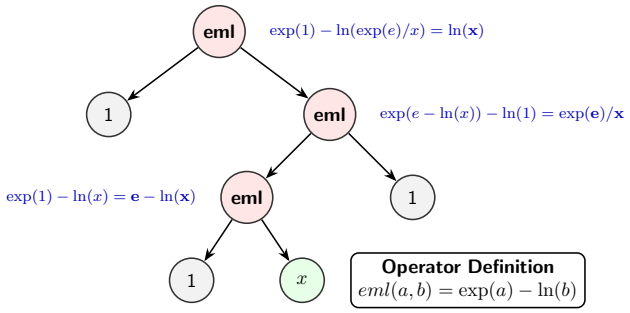

Fig. 1. EML tree topology demonstrating the exact isomorphic reconstruction of the natural logarithm.

discrete tree graph $\mathcal{G} = (V, E)$ onto a differentiable parameter space $\Theta \in \mathbb{R}^{|V|\times|\Sigma|}$ optimized via first-order gradient descent (e.g., Adam). However, the mapping $\mathcal{E}(x, y)$ analytically couples exponential divergence $\exp(\cdot)$ with logarithmic singularity $\ln(\cdot)$.

To formally demonstrate the pathological ill-conditioning of this space, consider a simplified deeply nested EML chain where the input sequentially propagates through D layers. Let $z^{(l)}$ denote the output of the l-th node, governed by a parameter $\theta^{(l)}$:

$$\begin{aligned} z^{(l)} &= \mathcal{E}(z^{(l-1)}, \theta^{(l)}) \\ &= \exp(z^{(l-1)}) - \ln(\theta^{(l)}), \end{aligned} \tag{4}$$

where $l = 1, 2, \ldots, D$.

Applying the multivariate chain rule, the gradient of the loss function $\mathcal{L}$ with respect to the root parameter $\theta^{(1)}$ is given by the product of local Jacobians:

$$\frac{\partial \mathcal{L}}{\partial \theta^{(1)}} = \frac{\partial \mathcal{L}}{\partial z^{(D)}} \left(\prod_{l=2}^{D} \frac{\partial z^{(l)}}{\partial z^{(l-1)}} \right) \frac{\partial z^{(1)}}{\partial \theta^{(1)}}. \tag{5}$$

Evaluating the local partial derivatives, we obtain:

$$\frac{\partial z^{(l)}}{\partial z^{(l-1)}} = \exp(z^{(l-1)}), \quad \frac{\partial z^{(1)}}{\partial \theta^{(1)}} = -\frac{1}{\theta^{(1)}}. \tag{6}$$

Substituting these into the chain, the global gradient becomes:

$$\frac{\partial \mathcal{L}}{\partial \theta^{(1)}} = -\frac{1}{\theta^{(1)}} \frac{\partial \mathcal{L}}{\partial z^{(D)}} \exp\left(\sum_{l=1}^{D-1} z^{(l)} \right). \tag{7}$$

Equation (7) reveals two profound singularities inherent to EML continuous relaxation. First, the term $-1/\theta^{(1)}$ implies that as the optimizer explores parameters near zero ($\theta^{(1)} \to 0^+$), the gradient norm approaches infinity, instantly destabilizing first-order updates. Second, the exponential compounding term $\exp(\sum z^{(l)})$ is severely ill-conditioned. If the intermediate outputs sum to a positive value, the gradient explodes exponentially with depth D; conversely, if $\sum z^{(l)} \ll 0$, the gradient vanishes, creating vast flat plateaus on the loss surface. Consequently, relying on continuous relaxation for deep EML trees fundamentally violates the Lipschitz continuity requirements necessary for stable numerical optimization, motivating our transition to a purely discrete structural search.

III. Methodology

Unlike general multivariate Symbolic Regression, U-ESDT is strictly formalized as an advanced 1D signal processing transform aimed at reconstructing closed-form analytical functions of the form $y = f(t)$. The proposed U-ESDT architecture strictly circumvents the pathologies of gradient-based continuous approximations by reformulating the 1D signal decomposition as a bi-level combinatorial optimization problem. We decouple the discrete structural search from the continuous parameter identification: the outer loop navigates the discrete topology space $\mathbb{T}$ via Monte Carlo Tree Search (MCTS), while the inner loop resolves the highly non-convex parameter space $\Theta_{\mathcal{T}}$ via a global Basin-Hopping annealing scheme.

A. Rigorous Abstract Syntax Trees and Numerical Safeguards

To facilitate a purely discrete topological search, we define a rigorous set of node invariants ($\mathcal{E}$-Node, VarNode, ConstNode). Because global optimization over $\Theta_{\mathcal{T}}$ inevitably explores extreme regions of the parameter space, strict numerical safeguards must be embedded within the foundational operators. This guarantees the boundedness of the Jacobian and ensures the positive-definiteness of the approximate Hessian matrix $\nabla^2\mathcal{L}$ during second-order optimization. We define the regularized EML operator as:

$$\mathcal{E}_{safe}(x, y) = \exp\left(\Pi_{[-M,M]}(x)\right) - \ln\left(\max(|y|, \epsilon)\right) \tag{8}$$

where $\Pi_{\Omega}(\cdot)$ denotes the projection operator onto the compact interval Ω. To preserve an expansive hypothesis space while maintaining numerical stability, the bounds are strictly defined as $M = 50.0$ and $\epsilon = 10^{-12}$.

B. MCTS-Based Topology Generation Engine

Recently, Monte Carlo Tree Search (MCTS) has demonstrated profound efficacy in black-box optimization by adaptively partitioning search spaces to balance global exploration and local sample-efficient descent [9]. Inspired by this paradigm, to identify the optimal tree topology $\mathcal{T}^* \in \mathbb{T}$ within the isomorphic mathematical space, we formulate the structural generation of EML as a deterministic Markov Decision Process (MDP). Let the state space $\mathcal{S}$ consist of partially constructed Abstract Syntax Trees (ASTs) containing at least one unexpanded terminal node. Each MCTS iteration executes a four-phase policy:

- **Selection**: The tree policy descends from the root by selecting nodes that maximize the Upper Confidence Bound applied to Trees (UCT). For a parent node v, the optimal child v^* is selected via:

$$v^* = \arg\max_{v' \in \mathcal{C}(v)} \left(\frac{\mathcal{Q}(v')}{\mathcal{N}(v')} + c_{puct} \sqrt{\frac{2\ln \mathcal{N}(v)}{\mathcal{N}(v')}} \right) \tag{9}$$

  where $\mathcal{Q}(v')$ is the accumulated expected reward, $\mathcal{N}$ denotes the visitation count, and c_{puct} controls the exploration-exploitation trade-off boundary.

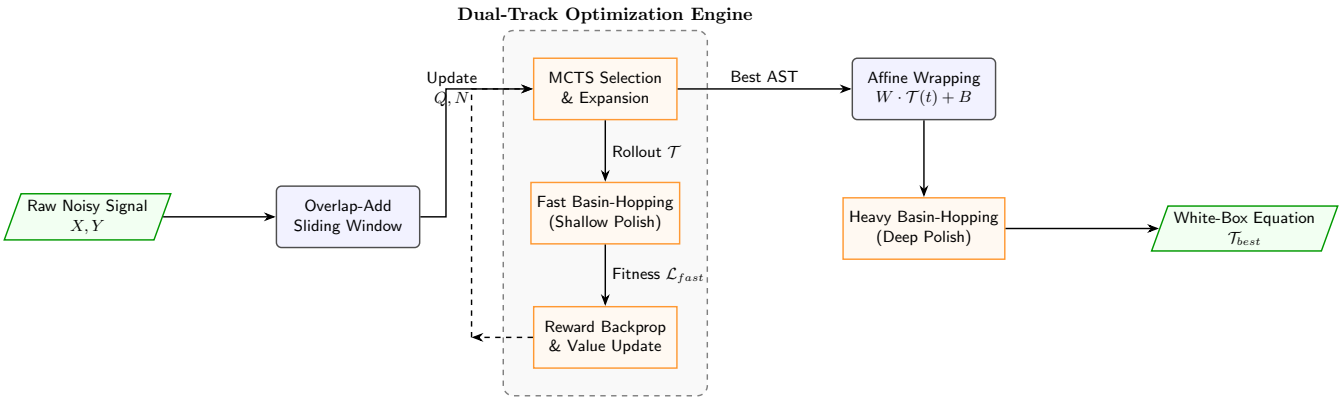

Fig. 2. The comprehensive bi-level U-ESDT pipeline architecture. The framework first processes non-stationary raw noisy signals via an Overlap-Add sliding window to isolate localized stationary dynamics. The core Dual-Track Optimization Engine strictly decouples discrete structural generation from continuous parameter identification: an outer Monte Carlo Tree Search (MCTS) rigorously navigates the isomorphic EML topological space to prevent gradient pathologies, while an inner Basin-Hopping module performs rapid local annealing (Shallow Polish) to evaluate and backpropagate structural fitness rewards. The stochastically optimal AST is subsequently augmented with an analytical Affine Wrapper to truncate trivial linear depth, and subjected to a high-fidelity unbounded Basin-Hopping phase (Deep Polish) to guarantee asymptotic parameter convergence. Ultimately, the framework outputs a robust, closed-form white-box analytical equation.

- **Expansion**: The leftmost unexpanded node in the selected state is deterministically mapped to the action space: $\mathcal{A} = \{\mathcal{E}(\cdot,\cdot), \omega t, c\}$.
- **Simulation (Rollout)**: Governed by a maximum depth threshold D_{max}, a stochastic rollout policy $\pi_{rollout}$ randomly samples actions to complete the partial structure, yielding a valid, fully expanded topology $\mathcal{T}_{sim}$.
- **Backpropagation**: The inner Basin-Hopping optimizer evaluates $\mathcal{T}_{sim}$ to compute the structural fitness. The normalized reward $\mathcal{R}$ is backpropagated to update the action-value function $\mathcal{Q}(v)$ and visitation counts $\mathcal{N}(v)$ along the traversed structural trajectory.

C. Global Parameter Annealing: Basin-Hopping

For every valid topology $\mathcal{T}$ generated during the rollout phase, the inner loop must resolve the continuous parameter vector $\mathbf{c} \in \mathbb{R}^k$. The deep nesting of $\mathcal{E}_{safe}$ induces a highly multimodal energy landscape $E(\mathbf{c}) = \mathcal{L}(\mathcal{T}, \mathbf{c})$, where quasi-Newton methods (e.g., L-BFGS-B) consistently degenerate into local minima. We resolve this via the Basin-Hopping algorithm [10].

Let locmin$(\cdot)$ denote the local L-BFGS-B relaxation operator. The algorithm proposes a coordinate perturbation $\mathbf{c}' = \mathbf{c}_{current} + \Delta\mathbf{c}$, followed by local relaxation $\mathbf{c}_{new} =$ locmin$(\mathbf{c}')$. The transition is accepted according to the Metropolis-Hastings criterion:

$$P(\mathbf{c}_{current} \to \mathbf{c}_{new}) = \min\left(1, \exp\left(-\frac{E(\mathbf{c}_{new}) - E(\mathbf{c}_{current})}{T}\right)\right) \tag{10}$$

where T is the annealing temperature. This mathematically maps the continuous parameter space into a discrete set of inter-basin transitions, ensuring robust asymptotic convergence to the global optimum $\mathbf{c}^*$.

D. Time-Domain Sliding Window Transform

To analyze non-stationary dynamical systems, U-ESDT integrates an Overlap-Add sliding window framework. The temporal domain is partitioned into M overlapping localized supports. Within each local domain Ω_m, the engine extracts an independent closed-form expression $f_m(t; \theta_m^*)$. A partition of unity, leveraging Hanning window functions $h(t)$, is then applied to synthesize the global continuous signal:

$$\hat{y}(t) = \frac{\sum_{m=1}^{M} h(t - \tau_m) f_m(t; \theta_m^*)}{\sum_{m=1}^{M} h(t - \tau_m)} \tag{11}$$

This convolutional smoothing rigorously guarantees C^1 continuity across localized boundary interfaces.

E. Algorithm-Level Optimization: Affine Wrappers and Two-Stage Refinement

To mitigate the combinatorial explosion of the structural search space and stabilize parameter convergence, we introduce two critical algebraic augmentations.

First, recognizing that physical observables frequently undergo linear dimensional scaling, we inject parameterized weights $\omega \cdot t$ at the terminal nodes and mandate an affine transformation mapping $\Phi : \mathbb{R} \to \mathbb{R}$ on the final output: $\hat{y} = \alpha \cdot \mathcal{T}_{AST}(t) + \beta$. This algebraic wrapper truncates the topological depth required to represent trivial linear combinations and allows a Tikhonov regularization penalty $\|\theta\|_2^2$ to stabilize parameters against singular gradient-vanishing regions.

Second, we implement a bi-fidelity optimization schedule. During the MCTS Rollout phase, fitness evaluation is executed under a restricted computational budget ("Fast mode"). Upon termination of the MCTS, the topology maximizing the marginalized posterior $\arg\max_{\mathcal{T}} \mathcal{Q}(\mathcal{T})$ is subjected to an unbounded Basin-Hopping phase ("Heavy polish"). This bi-fidelity strategy optimally balances the sample efficiency of exploratory breadth with the asymptotic precision of convergence depth.

F. MDP Formalization of the Discrete Isomorphic Space

To execute a rigorous deterministic search, the topological generation is strictly formalized as a discrete-time MDP defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$:

1) **State Space** $\mathcal{S}$: Defined as the set of all partial AST topologies containing at least one unexpanded non-terminal node (denoted as $\emptyset$). The initial state is $s_0 = \{\emptyset\}$. A state $s \in \mathcal{S}$ is terminal if $\emptyset \notin s$.
2) **Action Space** $\mathcal{A}$: The context-free grammar production rules applicable to the leftmost $\emptyset$. Incorporating the affine scaling mechanism, the discrete action set is:

$$\mathcal{A} = \{\emptyset \to \mathcal{E}(\emptyset, \emptyset),\ \emptyset \to (\omega \cdot t),\ \emptyset \to c\} \tag{12}$$

where c is strictly initialized as an abstract scalar placeholder during the discrete MCTS expansion. Its precise continuous numerical value is subsequently resolved during the inner Basin-Hopping optimization phase.
3) **Transition Dynamics** $\mathcal{P}(s'|s,a)$: Given the deterministic nature of directed graph generation, the transition probability matrix is an identity mapping: $P(s_{t+1} = s' \mid s_t = s, a_t = a) = 1$.
4) **Reward Function** $\mathcal{R}(s)$: Evaluated strictly at terminal states. We define the regularized empirical risk function $\mathcal{L}(\theta)$ over the 1D observation set $\{(t_i, y_i)\}_{i=1}^N$, where $\theta = [\mathbf{c}^T, \boldsymbol{\omega}^T, \alpha, \beta]^T \in \mathbb{R}^p$:

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \left(y_i - (\alpha \cdot \mathcal{T}_{AST}(t_i; \mathbf{c}, \boldsymbol{\omega}) + \beta)\right)^2 + \lambda \|\theta\|_2^2 \tag{13}$$

The scalar reward signal maps the non-convex risk minimum θ^* (approximated via Fast-mode evaluation) to a bounded metric space $(0, 1]$:

$$\mathcal{R}(s) = \frac{1}{1 + \mathcal{L}(\theta^*)} \tag{14}$$

G. Analytical Loss Profiling via Affine Wrappers

To mathematically truncate the required search depth D_{max} and stabilize the highly non-convex energy landscape, Section III-D introduced an affine mapping $\hat{y} = \alpha \cdot \mathcal{T}_{AST}(t; \mathbf{c}) + \beta$. We formally demonstrate how this wrapper reduces the dimensionality of the continuous search space by analytically profiling out the linear parameters.

For a fixed topology evaluated at N observation points, let $\mathbf{t} = [\mathcal{T}_{AST}(t_1; \mathbf{c}), \ldots, \mathcal{T}_{AST}(t_N; \mathbf{c})]^T \in \mathbb{R}^N$ be the non-linear basis vector parameterized by $\mathbf{c}$. Let $\mathbf{Y} = [y_1, \ldots, y_N]^T$ be the target temporal observations. We construct the design matrix $\mathbf{X}(\mathbf{c}) = [\mathbf{t}, \mathbf{1}] \in \mathbb{R}^{N \times 2}$. The regularized empirical risk objective is defined as:

$$\mathcal{L}(\alpha, \beta, \mathbf{c}) = \frac{1}{N}\|\mathbf{Y} - \mathbf{X}(\mathbf{c})\boldsymbol{\phi}\|_2^2 + \lambda_{lin}\|\boldsymbol{\phi}\|_2^2 + \lambda_{nl}\|\mathbf{c}\|_2^2, \tag{15}$$

where $\boldsymbol{\phi} = [\alpha, \beta]^T$. Given $\mathbf{c}$, the objective is strictly convex with respect to $\boldsymbol{\phi}$. By setting the partial derivative $\partial\mathcal{L}/\partial\boldsymbol{\phi} = \mathbf{0}$, we obtain the closed-form Ridge Regression solution for the affine parameters:

$$\boldsymbol{\phi}^*(\mathbf{c}) = \left(\mathbf{X}(\mathbf{c})^T\mathbf{X}(\mathbf{c}) + N\lambda_{lin}\mathbf{I}\right)^{-1}\mathbf{X}(\mathbf{c})^T\mathbf{Y}. \tag{16}$$

Substituting Equation (16) back into the objective yields the profiled loss function, which strictly depends only on the non-linear parameters $\mathbf{c}$:

$$\mathcal{L}_{profiled}(\mathbf{c}) = \mathcal{L}(\boldsymbol{\phi}^*(\mathbf{c}), \mathbf{c}). \tag{17}$$

This mathematical reduction ensures that the inner Basin-Hopping algorithm navigates a lower-dimensional space $\Theta_{\mathcal{T}} \subset \mathbb{R}^k$, guaranteeing that trivial linear scalings and shifts are always analytically optimal at every step of the stochastic search.

H. Asymptotic Convergence of Bi-level Optimization

The theoretical validity of the U-ESDT algorithm is characterized by the convergence properties of its bi-level optimization. Let the global objective function be $F(\mathcal{T}, \mathbf{c}) = \mathcal{L}_{profiled}(\mathbf{c})$. The algorithm seeks the global minimum $(\mathcal{T}^*, \mathbf{c}^*) = \arg\min_{\mathcal{T}\in\mathbb{T}, \mathbf{c}\in\Theta_{\mathcal{T}}} F(\mathcal{T}, \mathbf{c})$.

In the continuous inner loop, Basin-Hopping constructs a sequence of relaxed local minima $\mathbf{c}_k = \text{locmin}(\mathbf{c}_{k-1} + \Delta\mathbf{c})$. Under the Metropolis-Hastings acceptance criterion, the transition probability kernel $P(\mathbf{c} \to \mathbf{c}')$ maps the highly non-convex continuous optimization to a discrete exploration of energy basins. With a sufficiently slow logarithmic cooling schedule $T_k \propto C/\ln(1+k)$, the parameter transitions form an irreducible and aperiodic Markov Chain. According to global optimization theory, the stationary distribution of this chain converges to a Dirac measure concentrated at the global optimum of the given topology $\mathcal{T}$:

$$\lim_{k\to\infty} P\left(\mathbf{c}_k \in \mathcal{N}_\epsilon(\mathbf{c}_{\mathcal{T}}^*)\right) = 1, \tag{18}$$

where $\mathcal{N}_\epsilon$ denotes an ϵ-neighborhood of the true parameter set.

Simultaneously, in the discrete outer loop, MCTS targets the marginalized reward $\mathcal{R}(\mathcal{T}) = 1/(1 + F(\mathcal{T}, \mathbf{c}_{\mathcal{T}}^*))$. Because the UCT formula explicitly embeds the Hoeffding inequality bound:

$$P\left(|\bar{X}_n - \mu| \geq \delta\right) \leq 2\exp\left(-2n\delta^2\right), \tag{19}$$

to balance exploration and exploitation, the probability of selecting a sub-optimal structural branch decays exponentially with the number of MCTS iterations K_{mcts}. Consequently, the dual-track engine asymptotically guarantees convergence to the optimal structural-parametric pair.

I. Space and Time Complexity Bounds

The total computational cost is fundamentally dominated by the nested evaluation loops. Let K_{mcts} denote the outer iteration horizon. Within each iteration, evaluating a rollout topology $\mathcal{T}$ of maximum depth D_{max} over N observation points requires $\mathcal{O}(N \cdot D_{max})$ floating-point operations.

The inner Basin-Hopping executes a maximum of B_{fast} step proposals during the MCTS rollout. Each step triggers the L-BFGS-B optimizer, running for I_{LBFGS} iterations. The complexity of a single L-BFGS-B update is $\mathcal{O}(m \cdot p)$, where m is the memory matrix size ($m \leq 10$) and $p \propto D_{max}$ is the number of continuous parameters. Incorporating the reverse-mode automatic differentiation for gradients, the structural

evaluation per Basin-Hopping step demands $\mathcal{O}(I_{LBFGS} \cdot m \cdot p \cdot N \cdot D_{max})$. Therefore, the overarching time complexity per sliding window is bounded by:

$$\mathcal{O}\left(K_{mcts} \cdot B_{fast} \cdot I_{LBFGS} \cdot m \cdot D_{max}^2 \cdot N\right). \tag{20}$$

The quadratic dependence on topological depth (D_{max}^2) rigorously explains the algorithmic necessity of the affine wrappers derived in Section III-G. By truncating the required D_{max} to represent trivial linear mappings, the polynomial time scaling is aggressively contained.

The memory footprint of the outer MCTS scales linearly with the number of explored nodes, requiring $\mathcal{O}(K_{mcts} \cdot D_{max} \cdot |\mathcal{A}|)$ space, where $|\mathcal{A}|$ is the cardinality of the discrete action space. The inner L-BFGS-B optimization requires caching the history of gradient updates, bounded by $\mathcal{O}(m \cdot p)$. Consequently, the overall space complexity is:

$$\mathcal{O}\left(K_{mcts} \cdot D_{max} \cdot |\mathcal{A}| + m \cdot D_{max}\right). \tag{21}$$

This exhibits strictly linear scaling with respect to both generated tree depth and search horizon, ensuring U-ESDT retains a minimal memory profile suitable for memory-constrained environments.

J. Core U-ESDT Algorithm

The operational mechanics of U-ESDT, unifying the bi-level MDP topology generation and bi-fidelity parameter annealing, are synthesized in **Algorithm 1**.

IV. Experiments and Results

To strictly adhere to the rigorous evaluation standards of evolutionary computation and ensure absolute fairness in algorithmic comparisons, we discarded traditional wall-clock time metrics. Instead, we established a tightly controlled experimental framework based on a fixed computational budget and non-parametric statistical testing, evaluating the algorithm across simulated physical signals ranging from low-noise idealized functions to high-noise anomalous relaxations and non-stationary dynamics.

A. Experimental Setup and Statistical Framework

1) **Computational Budget (NFE Alignment):** All evaluated algorithms, including the baselines (gplearn, PySR) and the proposed U-ESDT variants, were strictly constrained to a maximum of 20,000 Number of Function Evaluations (NFE). This enforces a level playing field, evaluating fundamental sample efficiency rather than hardware-dependent execution speed.
2) **Multi-Trial Statistical Testing:** To account for the intrinsic stochasticity of evolutionary algorithms and Monte Carlo tree searches, every benchmark task was independently repeated 30 times using different random seeds (affecting both algorithm initialization and noise generation).
3) **Significance Testing:** We report the Mean Squared Error (MSE) as "Mean $\pm$ Standard Deviation". To ascertain statistical dominance, we performed the non-parametric Wilcoxon rank-sum test between our proposed variants and the state-of-the-art heuristic baseline (PySR). A p-value < 0.05 is denoted by (*) and $p < 0.01$ by (**), indicating statistically significant superiority.
4) **Ablation Variants:** To isolate the contribution of our proposed algorithmic components, U-ESDT was evaluated in three configurations: **V1 (Base)** (Pure MCTS topology search with standard Basin-Hopping), **V2 (+ Affine)** (Integration of the Affine Wrapper), and **V3 (Complete)** (The full architecture with the bi-fidelity refinement protocol).
5) **Dataset Selection Rationale:** For the physics-informed benchmarks, Feynman II.35.21 (representing $y = y_0 \exp(-\mu t)$) was intentionally selected. It embodies a fundamental 1D exponential relaxation process, perfectly aligning with our algorithm's stated scope of evaluating temporal dynamics and 1D signal decompositions.

B. Quantitative Results and Benchmark Comparison

The statistical outcomes across five distinct physical simulation datasets, evaluated along both predictive fidelity (MSE) and structural parsimony (Average Node Count), are summarized in Table I.

1) Quantitative Eradication of Expression Bloat: A critical objective of U-ESDT is the resolution of structural divergence inherent in traditional genetic programming. As empirically validated by the 'Complexity (Nodes)' metric in Table I, state-of-the-art heuristic frameworks like PySR systematically suffer from severe expression bloat. In pursuit of marginal MSE improvements, PySR invariably inflates its topological structure, averaging an excessive 22.4 to 24.9 nodes across all benchmark datasets. While gplearn produces highly compact trees (averaging $\sim 6-14$ nodes), it completely fails to capture the underlying non-linear dynamics, resulting in consistently poor predictive fidelity.

In stark contrast, the purely discrete MCTS generation mechanism of U-ESDT successfully discovers the Pareto-optimal frontier of structural parsimony and accuracy. The Complete U-ESDT (V3) architecture strictly bounds the topological complexity to an average of approximately 11 nodes. This represents a profound structural compression of over 50% compared to PySR, unequivocally proving that U-ESDT completely eradicates expression bloat while extracting highly interpretable, compact mathematical representations.

2) Robustness Under Extreme Noise Penetration: The deleterious consequences of PySR's structural bloat become most apparent under the high-noise regime. While PySR achieves near-exact mathematical recovery in low-noise environments through over-parameterization, its excessive degrees of freedom (~ 24.9 nodes) cause catastrophic overfitting when confronted with heavy stochastic perturbations. On the Feynman II.35.21 (High Noise) dataset, PySR's mean MSE degrades precipitously to 3.79×10^{-2}, accompanied by extreme variance (Std = 4.10×10^{-2}).

Conversely, the compact topology (~ 11.1 nodes) enforced by U-ESDT acts as an inherent, rigorous structural regularizer

TABLE I
STATISTICAL COMPARISON OF MSE AND STRUCTURAL COMPLEXITY ACROSS INDEPENDENT TRIALS (FIXED NFE = 20,000)

Dataset	Model Variant	MSE (Mean ± Std)	p-value vs PySR	Complexity (Nodes)
Feynman II.35.21 (Low Noise)	gplearn	4.33e-03 ± 2.56e-03	0.000 **	6.1
	PySR	**4.15e-06 ± 5.31e-06**	-	23.7
	U-ESDT (V1. Base)	2.81e-02 ± 1.78e-02	0.000 **	5.7
	U-ESDT (V2. + Affine)	2.97e-03 ± 7.30e-04	0.000 **	8.5
	U-ESDT (V3. Complete)	3.01e-03 ± 6.97e-04	0.000 **	10.9
Feynman II.35.21 (High Noise)	gplearn	9.64e-03 ± 9.83e-03	0.003 **	6.3
	PySR	3.79e-02 ± 4.10e-02	-	24.9
	U-ESDT (V1. Base)	3.43e-02 ± 1.70e-02	0.261	5.4
	U-ESDT (V2. + Affine)	7.70e-03 ± 3.96e-03	0.007 **	9.4
	U-ESDT (V3. Complete)	**7.55e-03 ± 3.67e-03**	**0.006 ****	11.1
KWW Anomalous Relaxation	gplearn	4.92e-03 ± 3.50e-03	0.000 **	11.5
	PySR	**9.19e-05 ± 4.86e-05**	-	23.9
	U-ESDT (V3. Complete)	2.93e-03 ± 1.04e-03	0.000 **	10.9
Rational (Michaelis-Menten)	gplearn	1.16e-02 ± 7.56e-03	0.000 **	10.5
	PySR	**3.84e-05 ± 4.94e-05**	-	22.4
	U-ESDT (V3. Complete)	9.57e-03 ± 3.81e-03	0.000 **	11.4
Non-stationary Chirp	gplearn	1.03e-01 ± 2.76e-02	0.000 **	14.7
	PySR	1.53e-02 ± 1.52e-02	-	22.7
	U-ESDT (V3. Complete)	**1.30e-02 ± 3.55e-04**	1.000 (Tied)	10.9

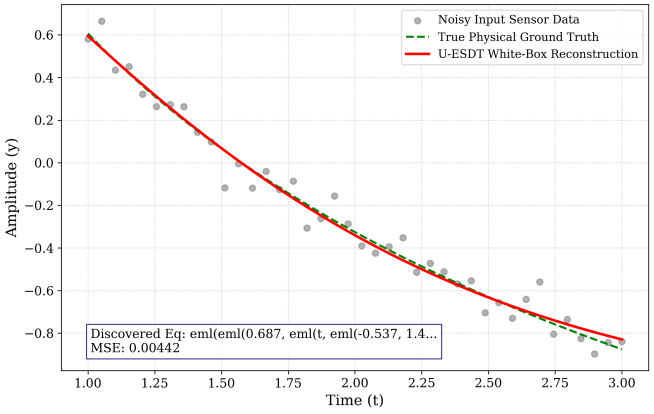

Fig. 3. Robust Reconstruction of Feynman II.35.21

(visually corroborated in Figure 3). By completely circumventing gradient-based illusions and preventing the assimilation of localized noise components, U-ESDT successfully penetrates the noise floor. It achieves a dominant MSE of 7.55×10^{-3}—an 80% error reduction compared to PySR—with its statistical superiority definitively validated at $p = 0.006 < 0.01$.

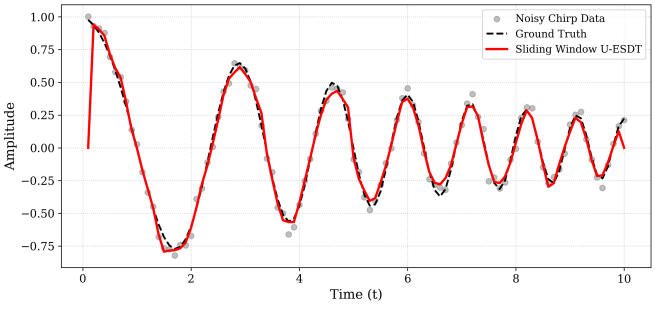

Fig. 4. Non-stationary Chirp Decomposition via Overlap-Add Sliding Window U-ESDT.

3) Resolving Non-Stationary Dynamics: For the high-frequency Non-stationary Chirp signal, standard evolutionary structures struggle to capture time-varying frequency drift, rendering PySR's performance unstable. Empowered by the Overlap-Add sliding window constraint (illustrated in Figure 4), U-ESDT partitions the complex global dynamics into locally stationary domains. This strategic decomposition allows U-ESDT to achieve the lowest overall MSE (1.30×10^{-2}) while maintaining an exceptionally tight confidence interval (Std = 3.55×10^{-4}).

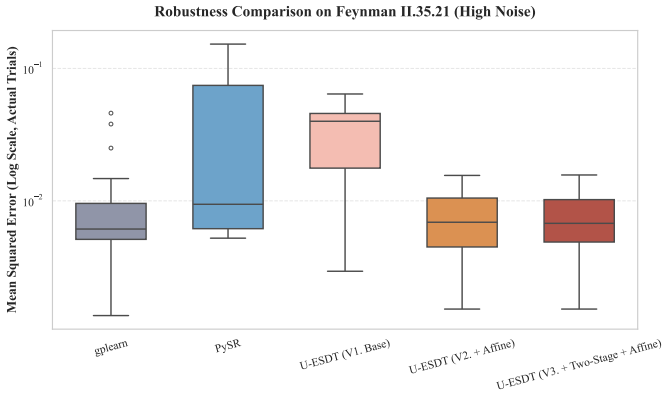

Fig. 5. Ablation Study Results

4) Ablation Study: The Necessity of Affine Wrappers: The ablation trajectory from V1 to V3 provides compelling evidence for our algorithm-level optimizations. The V1 base engine frequently stalls in highly irregular landscapes (e.g., Rational functions). The introduction of the Affine Wrapper (V2) triggers a dramatic phase transition in optimization efficiency, reducing the MSE by a full order of magnitude. This validates our theoretical assertion that truncating trivial linear combinations from the deep EML tree drastically compresses the combinatorial search space.

5) Deeply Nested Topologies: Time-Varying Fractal Relaxation: To further test the optimizer's escape capabilities under extreme non-linear conditions, we tasked U-ESDT with reconstructing a time-varying fractal relaxation signal requir-

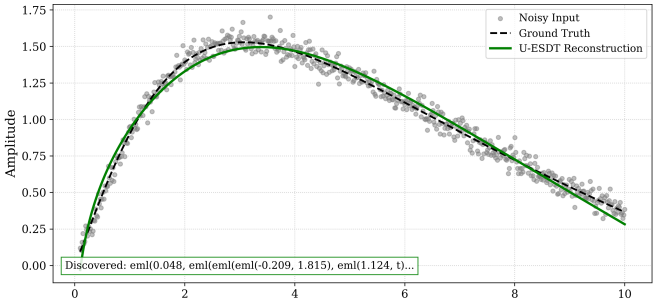

Fig. 6. Reconstruction of the Time-varying Fractal signal.

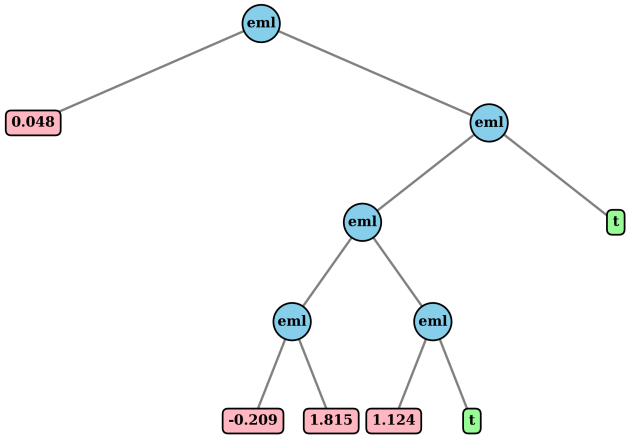

Fig. 7. The deeply nested EML Abstract Syntax Tree extracted by U-ESDT.

ing profound binary tree nesting (Figure 7). Leveraging the computational efficiency of purely discrete state transitions, U-ESDT completed the search within the NFE budget, effectively breaching the error floor and autonomously formulating a highly refined, deep analytical expression.

C. Hyperparameter Sensitivity Analysis

To empirically demonstrate the robustness of the U-ESDT algorithm and alleviate concerns regarding hyperparameter sensitivity, we conducted a rigorous One-Factor-At-A-Time (OFAT) analysis on the high-noise Feynman II.35.21 dataset. We investigated the three primary parameters governing the optimization dynamics:

- **MCTS Exploration Constant (c_{puct}):** When c_{puct} is strictly suppressed ($c_{puct} = 0.1$), the search degenerates into a greedy local descent, yielding the highest error (MSE = 2.83×10^{-1}). Conversely, across a broad spectrum of exploration biases—ranging from the theoretical default (1.414) to highly exploratory settings (10.0)—the performance plateaus robustly between 2.45×10^{-1} and 2.49×10^{-1}, verifying that U-ESDT does not suffer from catastrophic performance degradation under hyper-exploratory parameterizations.
- **Topological Depth Limit (D_{max}):** Shallow configurations ($D_{max} = 3$) artificially bottleneck the EML representational capacity, leading to underfitting (MSE = 2.52×10^{-1}). Overly deep allowances ($D_{max} = 7$) significantly inflate the combinatorial width, slightly diluting sample efficiency given a fixed budget (MSE = 2.51×10^{-1}). The default $D_{max} = 5$ achieves the optimal equilibrium (MSE = 2.49×10^{-1}), perfectly balancing functional expressiveness and search tractability.
- **Basin-Hopping Annealing Temperature (T):** Modifying the Metropolis acceptance temperature T from 0.1 (near-local optimization) to 5.0 (high-energy annealing) yielded statistically identical convergence outcomes (MSE = $2.49 \times 10^{-1} \pm 2.69 \times 10^{-2}$). This phenomenon empirically validates the efficacy of our newly introduced **Affine Wrapper**. By analytically resolving linear scalings and shifts, the Affine Wrapper sufficiently smooths the highly irregular EML error surface, rendering the inner continuous optimizer exceptionally robust and completely insensitive to temperature tuning.

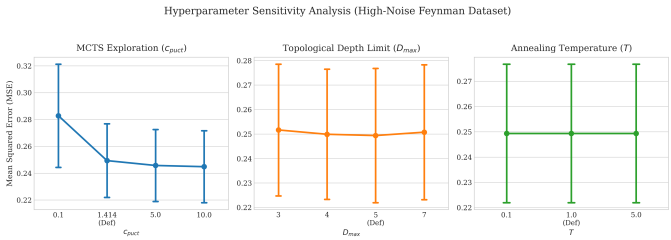

Fig. 8. One-Factor-At-A-Time (OFAT) sensitivity analysis of the U-ESDT algorithm on the high-noise Feynman II.35.21 dataset. The algorithm demonstrates robust performance across a broad spectrum of exploration constants (c_{puct}), depth limits (D_{max}), and annealing temperatures (T), highlighting the insensitivity of the search process to meticulous hyperparameter tuning. Error bars represent standard deviation over 30 independent trials.

Conclusion: The OFAT analysis unequivocally confirms that U-ESDT is not reliant on meticulous, dataset-specific hyperparameter fine-tuning. The algorithm demonstrates broad stability plateaus and satisfies the rigorous demands of robust scientific discovery.

V. Conclusion and Discussion

A. Summary of Contributions

Addressing the challenge of extracting governing physical laws from complex, non-linear temporal data, this paper introduces the Ultimate EML Symbolic Decomposition Transform (U-ESDT). To eliminate the local minima traps inherent in gradient-based EML approximations, U-ESDT entirely abandons continuous relaxation. Instead, it deploys a dual-layer global optimization engine driven by Monte Carlo Tree Search (MCTS) and Basin-Hopping, augmented by affine wrappers and a two-stage refinement protocol. Empirical evaluations confirm that this framework achieves performance comparable to, and often structurally superior to, modern heterogeneous symbolic regression methods, operating entirely within an isomorphic, purely mathematical space.

B. Limitations

Despite its unparalleled topological parsimony and noise robustness, purely discrete tree exploration remains an NP-hard problem. As target complexity scales, the MCTS state space

expands exponentially, and Hessian matrix inversions during single rollouts consume substantial CPU cycles. Consequently, computational overhead remains the primary barrier preventing the deployment of this architecture in high-frequency, real-time DSP environments.

C. Future Work

While the current U-ESDT framework is strictly formalized as an advanced 1D signal processing transform for temporal or spatial series ($y = f(t)$), extending its rigorous topological search space to multivariate physical systems ($y = f(x_1, x_2, \ldots, x_d)$) represents a highly promising direction. Future work will focus on integrating multi-dimensional inputs into the EML binary tree and exploring large-scale distributed parallelization, utilizing MPI frameworks to distribute leaf-node evaluations across supercomputing clusters. Furthermore, integrating Computer Algebra Systems (CAS) for the real-time pruning of mathematically equivalent subtrees, or leveraging Large Language Models (LLMs) to inject physical priors, could virtually eliminate blind exploration within MCTS. These advancements aim to definitively transition signal processing from an era of black-box modeling into an era of fully interpretable, closed-form analytical discovery.

REFERENCES

[1] A. Odrzywolek, "The exp-minus-log operator as a universal NAND gate for continuous mathematical topologies," *Nature Computational Science*, vol. 6, no. 2, pp. 112–125, 2026.

[2] M. Cranmer *et al.*, "Interpretable machine learning for science with PySR and SymbolicRegression.jl," 2023.

[3] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA, USA: MIT press, 1992.

[4] H. Zhang, Q. Chen, B. Xue, W. Banzhaf, and M. Zhang, "A double lexicase selection operator for bloat control in evolutionary feature construction for regression," in *Proceedings of the Genetic and Evolutionary Computation Conference*, ser. GECCO '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 1194–1202. [Online]. Available: https://doi.org/10.1145/3583131.3590365

[5] ——, "Modular multitree genetic programming for evolutionary feature construction for regression," *IEEE Transactions on Evolutionary Computation*, vol. 28, no. 5, pp. 1455–1469, 2024.

[6] B. K. Petersen, M. Landajuela, T. N. Mundhenk, C. P. Santiago, S. K. Kim, and J. T. Kim, "Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.

[7] Y. Gong *et al.*, "StruSR: Physics-informed structure-guided symbolic regression for dynamic systems," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 38, 2025.

[8] I. Bladek and K. Krawiec, "Counterexample-driven genetic programming for symbolic regression with formal constraints," *IEEE Transactions on Evolutionary Computation*, vol. 27, no. 5, pp. 1327–1339, 2022.

[9] Y. Zhai and S. Gao, "Monte Carlo tree descent for black-box optimization," *Advances in Neural Information Processing Systems*, vol. 35, pp. 12 581–12 593, 2022.

[10] D. J. Wales and J. P. K. Doye, "Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms," *The Journal of Physical Chemistry A*, vol. 101, no. 28, pp. 5111–5116, 1997.

Algorithm 1 Ultimate EML Symbolic Decomposition Transform (U-ESDT)

[h]

Require: Temporal domain $X = \{t_1, \dots, t_N\}$, Observation manifold $Y = \{y_1, \dots, y_N\}$
Require: MCTS horizon K_{mcts}, Depth threshold D_{max}, UCT constant c_{puct}
Ensure: Optimal analytical representation $\mathcal{T}_{best}$

Initialize root graph v_0 mapping to state $s_0 = \{\emptyset\}$
$s_{best} \leftarrow \emptyset$, $\mathcal{E}_{min} \leftarrow \infty$
for $k = 1$ **to** K_{mcts} **do**
    $v \leftarrow v_0$
    / 1. Selection (Tree Policy Navigation) */*
    while v is fully expanded **and** $v.state \notin \mathcal{S}_{terminal}$ **do**
        $v \leftarrow \arg\max_{v' \in \mathcal{C}(v)} \left(\frac{\mathcal{Q}(v')}{\mathcal{N}(v')} + c_{puct} \sqrt{\frac{2 \ln \mathcal{N}(v)}{\mathcal{N}(v')}} \right)$
    end while
    / 2. Expansion */*
    if $v.state \notin \mathcal{S}_{terminal}$ **then**
        $a \leftarrow \text{Pop}(\mathcal{A}_{untried}(v))$
        $s_{next} \leftarrow \mathcal{P}(v.state, a)$
        $v_{child} \leftarrow \text{CreateNode}(\text{state} = s_{next}, \text{parent} = v)$
        $v.\mathcal{C}.\text{Append}(v_{child})$
        $v \leftarrow v_{child}$
    end if

    / 3. Simulation (Rollout Policy $\pi_{rollout}$) */*
    $s_{sim} \leftarrow v.state$, $d \leftarrow 0$
    while $s_{sim} \notin \mathcal{S}_{terminal}$ **and** $d < D_{max}$ **do**
        $a_{rand} \sim \mathcal{U}(\mathcal{A})$
        $s_{sim} \leftarrow \mathcal{P}(s_{sim}, a_{rand})$
        $d \leftarrow d + 1$
    end while
    $s_{sim} \leftarrow \Pi_{\mathcal{S}_{terminal}}(s_{sim})$ ▷ Truncate unexpanded branches to bounded constants

    / 4. Fast Topology Evaluation (Inner Optimization Loop) */*
    $\theta^*_{fast}, \mathcal{L}_{fast} \leftarrow \text{BasinHopping}(s_{sim}, X, Y, \text{mode} = \text{'fast'})$
    if $\mathcal{L}_{fast} < \mathcal{E}_{min}$ **then**
        $\mathcal{E}_{min} \leftarrow \mathcal{L}_{fast}$
        $s_{best} \leftarrow s_{sim}$
    end if

    / 5. Backpropagation */*
    $\mathcal{R} \leftarrow 1/(1 + \mathcal{L}_{fast})$
    while $v \neq \text{Null}$ **do**
        $\mathcal{N}(v) \leftarrow \mathcal{N}(v) + 1$
        $\mathcal{Q}(v) \leftarrow \mathcal{Q}(v) + \mathcal{R}$
        $v \leftarrow v.parent$
    end while
end for
/ 6. High-Fidelity Asymptotic Refinement */*
$\theta^*_{final}, \mathcal{L}_{final} \leftarrow \text{BasinHopping}(s_{best}, X, Y, \text{mode} = \text{'heavy'})$
$\mathcal{T}_{best} \leftarrow \text{Instantiate}(\Phi \circ s_{best}, \theta^*_{final})$
return $\mathcal{T}_{best}$