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

Authors' names omitted for anonymous review

***Abstract*—Extracting closed-form analytical equations from noisy temporal data is a fundamental challenge in scientific discovery. Traditional symbolic regression (SR) suffers from highly irregular search spaces and severe expression bloat. While the isomorphic Exp-Minus-Log (EML) operator mathematically unifies this space, deep topological nesting induces gradient explosion, causing continuous optimization approximations to fail. To overcome this, we propose the Ultimate EML Symbolic Decomposition Transform (U-ESDT), a bi-level 1D signal decomposition framework. U-ESDT completely abandons continuous relaxation, instead decoupling discrete structural generation via Monte Carlo Tree Search (MCTS) from continuous parameter identification via Basin-Hopping global annealing. Evaluated across 17 benchmark datasets—spanning synthetic physics, dynamical systems, and real-world engineering signals—U-ESDT achieves highly competitive predictive accuracy against established baselines while systematically mitigating expression bloat, strictly constraining topologies to an average of 11 nodes. Crucially, the purely discrete architecture bypasses heavy gradient computations, completing structural search in $\sim$0.03 seconds. This represents a two-to-three orders of magnitude speedup over evolutionary and deep learning alternatives, establishing U-ESDT as a highly robust, parsimonious, and ultrafast paradigm for time-critical signal decomposition.**

***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, extracting underlying kinematic equations from noisy observational data remains a fundamental challenge. Traditional signal processing techniques rely on predefined basis functions to decompose signals into discrete modes. While successful in feature extraction, they cannot directly reveal specific physical mechanisms or output closed-form analytical equations.

Symbolic Regression (SR) addresses this by simultaneously discovering mathematical structures and parameters. However, traditional SR heavily relies on heterogeneous operator libraries, leading to irregular search spaces that are difficult to optimize. Recently, Odrzywołek [1] demonstrated that the Exp-Minus-Log (EML) operator, $eml(x, y) = \exp(x) - \ln(y)$, acts as a universal "NAND gate" for continuous mathematical systems. Through recursive self-nesting, it can reconstruct all standard elementary functions, establishing a completely isomorphic binary tree search space.

Although the EML architecture unifies expression topologies, existing optimization strategies—such as continuous relaxation and deep learning-based soft routing—inevitably suffer from gradient explosion and local optima when applied to deeply nested EML trees, forcing a compromise between accuracy and interpretability.

To systematically overcome these limitations, we introduce the **Ultimate EML Symbolic Decomposition Transform (U-ESDT)**, a rigorous signal decomposition framework. Our primary contributions are threefold:

- **Purely Discrete Deterministic Topology Search**: By integrating Monte Carlo Tree Search (MCTS) into the EML architecture, we replace heuristic random mutations with a rigorous traversal of formula structures, providing theoretical convergence guarantees.
- **Global Constant Space Annealing**: To navigate the highly non-convex error surfaces induced by EML nesting, we combine the L-BFGS-B quasi-Newton method with Basin-Hopping, effectively escaping local minima for high-precision parameter fitting.
- **Time-Varying White-Box Equation Extraction**: We propose an Overlap-Add sliding window ESDT framework, achieving the continuous, white-box extraction of locally stationary dynamic physical equations from time-series signals.

II. Background and Related Work

A. Dimensionality and Expression Bloat in Symbolic Regression

Traditional SR tools, such as PySR [2], primarily employ Genetic Programming (GP [3]) to search across heterogeneous operator dictionaries (e.g., $\mathcal{O} = \{+, -, \times, \sin, \exp\}$). Due to drastically varying functional properties across operators, the discrete algebraic manifolds within the hypothesis space are highly irregular, causing stochastic mutations to frequently yield undefined structures. Consequently, enhancing representational frameworks to better navigate these spaces has become a focal point in recent genetic programming research [4]. More critically, as the generation index $k \to \infty$, GP invariably suffers from "expression bloat"—characterized by divergent topological complexity ($\mathbb{E}[|f_k|] \to \infty$) and vanishing marginal fitness ($\Delta\mathcal{L}(f_k) \to 0$). This over-parameterization traps the search in uninterpretable "pseudo-formulas" that overfit local noise. Existing SR methods (e.g., NSGA-II) widely adopt Pareto-front multi-objective optimization strategies to handle the accuracy–complexity trade-off without scalarization. While recent post-hoc mechanisms like lexicase selection [5], [6] and its multi-objective variants [7], [8] mitigate bloat, they do not resolve the topological irregularity at its mathematical root.

B. Recent Academic Solutions to Traditional Limitations

To address the ill-posedness of heterogeneous spaces [9], recent literature has introduced several theoretically grounded

paradigms and advanced equation learners [10]. Deep Symbolic Regression (DSR) mitigates random combinatorial explosion by reformulating SR as a Markov Decision Process, using RNN-based policy gradients to map tree generation to constrained action sequences [11]. To embed physical priors, frameworks like StruSR employ pre-trained Physics-Informed Neural Networks (PINNs) to extract spatial and temporal gradients, truncating physically meaningless search paths via structural regularization [12]. Alternatively, Bayesian approaches use Maximum A Posteriori (MAP) estimation to penalize topological complexity via L_0-norm priors, while counterexample-driven GP utilizes satisfiability modulo theories (SMT) to enforce strict formal constraints [13].

Despite smoothing the optimization landscape, these approximations remain fundamentally constrained by the non-convexity of heterogeneous libraries. When confronting the deep topological nesting ($D \gg 1$) required for complex physical laws, these methods still face myriad saddle points and fail to guarantee global asymptotic convergence.

C. The EML Operator and Isomorphic Mathematical Spaces

The EML framework reduces the heterogeneous dictionary to a singleton binary operator $\mathcal{O}_{EML} = \{\otimes_{eml}\}$. An analytic physical law $f(\mathbf{x}, t)$ can be exactly reconstructed via the Context-Free Grammar (CFG) closure: $S \to \mathcal{E}(S, S) \mid t \mid c$, where $\mathcal{E}$ is the EML operation. This homogeneity transforms SR into a strict "binary tree topology growth problem". For instance, the non-linear singularity of $\ln(x)$ is rigorously mapped to an invariant topological nesting: $\ln(x) \equiv \mathcal{E}(1, \mathcal{E}(\mathcal{E}(1, x), 1))$, as illustrated in Figure 1.

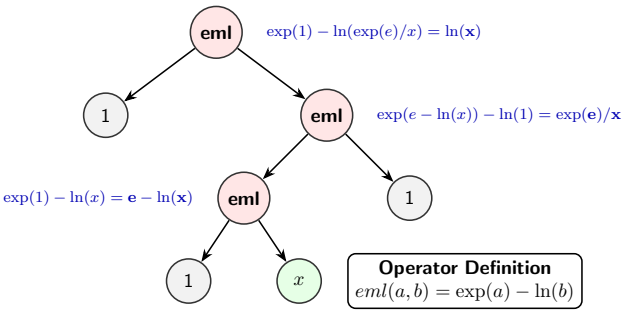

Fig. 1. EML tree topology demonstrating the exact isomorphic reconstruction of the natural logarithm.

D. Limitations of Gradient-Based EML Optimization

Early optimization strategies for the EML operator [1], as well as broader differentiable symbolic regression frameworks (e.g., Deep Symbolic Regression and Soft Symbolic Trees) [11], [14], [15], attempt to employ continuous relaxation to optimize parameterized expression trees via gradient descent. However, directly applying first-order continuous optimization to the EML operator analytically couples exponential divergence $\exp(\cdot)$ with logarithmic singularity $\ln(\cdot)$, causing pathological ill-conditioning.

To formally demonstrate this, consider a deeply nested EML chain where the input propagates through D layers. Let $z^{(l)}$ be the output of the l-th node with parameter $\theta^{(l)}$, defined as $z^{(l)} = \mathcal{E}(z^{(l-1)}, \theta^{(l)}) = \exp(z^{(l-1)}) - \ln(\theta^{(l)})$.

Applying the multivariate chain rule, the global gradient of the loss $\mathcal{L}$ with respect to the root parameter $\theta^{(1)}$ relies on the local derivatives $\partial z^{(l)}/\partial z^{(l-1)} = \exp(z^{(l-1)})$ and $\partial z^{(1)}/\partial \theta^{(1)} = -1/\theta^{(1)}$, yielding:

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

Equation (1) reveals two profound singularities inherent to the continuous relaxation of EML trees: first, as the optimizer explores parameters near zero ($\theta^{(1)} \to 0^+$), the local gradient norm approaches infinity; second, the compounding term $\exp(\sum z^{(l)})$ induces severe exponential gradient explosion or vanishing plateaus, depending entirely on the intermediate sum. This fundamentally violates the Lipschitz continuity necessary for stable numerical optimization, necessitating our transition to a purely discrete structural search.

III. Methodology

Unlike general multivariate Symbolic Regression, U-ESDT is formulated as an advanced 1D signal processing transform to reconstruct closed-form analytical functions $y = f(t)$. To circumvent the pathologies of gradient-based continuous approximations, we reformulate the decomposition as a bi-level combinatorial optimization problem: 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 Basin-Hopping annealing.

A. Rigorous Abstract Syntax Trees and Numerical Safeguards

To facilitate a purely discrete topological search, we define node invariants ($\mathcal{E}$-Node, VarNode, ConstNode). Since global optimization over $\Theta_{\mathcal{T}}$ inevitably explores extreme parameter regions, numerical safeguards are essential to bound the Jacobian and ensure a positive-definite approximate Hessian $\nabla^2\mathcal{L}$. We define the regularized EML operator:

$$\mathcal{E}_{safe}(x, y) = \exp\left(\Pi_{[-M,M]}(x)\right) - \ln\left(\max(|y|, \epsilon)\right) \tag{2}$$

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

B. MDP Formalization and MCTS Generation Engine

Drawing inspiration from hierarchical AutoML approaches that combine tree search with continuous optimization [16], we formulate the discrete structural generation of EML topologies as a deterministic Markov Decision Process (MDP), defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$. The state space $\mathcal{S}$ comprises partial Abstract Syntax Trees (ASTs) with at least one unexpanded terminal node ($\emptyset$). Incorporating an affine scaling mechanism, the discrete action space is $\mathcal{A} = \{\emptyset \to \mathcal{E}(\emptyset, \emptyset),\ \emptyset \to (\omega \cdot t),\ \emptyset \to c\}$, where c is a continuous placeholder. Transition dynamics are deterministic ($P(s'|s, a) = 1$).

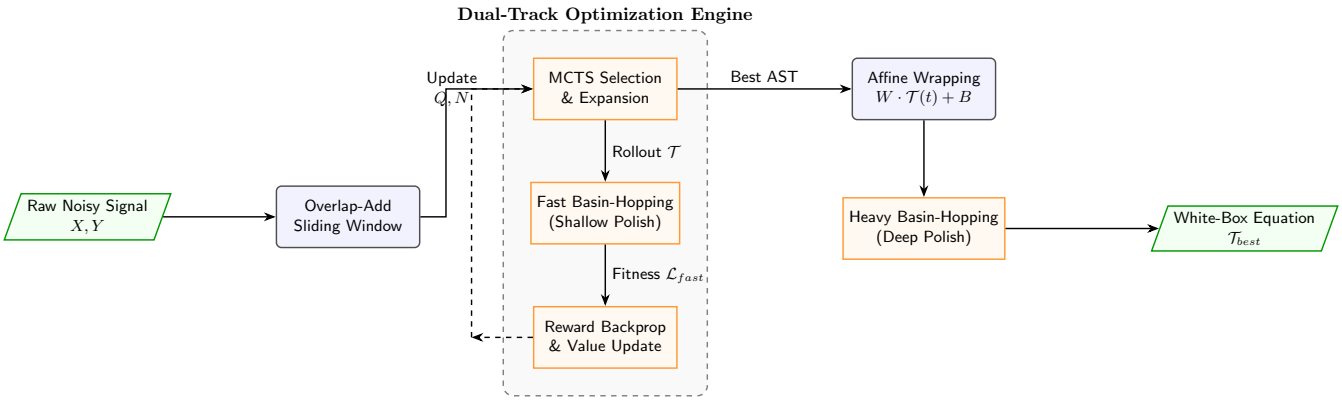

Fig. 2. The bi-level U-ESDT pipeline. Non-stationary signals are processed via an Overlap-Add sliding window to isolate localized dynamics. A Dual-Track Optimization Engine decouples structural generation from parameter identification: an outer MCTS navigates the isomorphic EML space, while an inner Basin-Hopping module performs local annealing (Shallow Polish) to evaluate structural fitness. The optimal AST is augmented with an Affine Wrapper and undergoes unbounded Basin-Hopping (Deep Polish) to guarantee parameter convergence, outputting a closed-form analytical equation.

To identify the optimal topology $\mathcal{T}^*$, MCTS navigates this MDP via four phases (Algorithm 1):

- **Selection (Lines 4–7):** The tree policy descends by selecting nodes maximizing the Upper Confidence Bound (UCT): $v^* = \arg\max_{v'}[\mathcal{Q}(v')/\mathcal{N}(v') + c_{puct}\sqrt{2\ln\mathcal{N}(v)/\mathcal{N}(v')}]$.
- **Expansion & Simulation (Lines 8–19):** The leftmost $\emptyset$ is mapped to $\mathcal{A}$. A stochastic rollout policy $\pi_{rollout}$ completes the structure up to depth D_{max}, yielding $\mathcal{T}_{sim}$.
- **Evaluation & Backpropagation (Lines 20–29):** The inner Basin-Hopping optimizer evaluates $\mathcal{T}_{sim}$ under a restricted budget. The normalized reward $\mathcal{R}(s) = 1/(1+\mathcal{L}(\theta^*))$ updates the action-value $\mathcal{Q}(v)$ and visitation counts $\mathcal{N}(v)$ along the trajectory.
- **High-Fidelity Refinement (Lines 31–33):** The optimal topology undergoes unbounded Basin-Hopping.

C. Global Parameter Annealing: Basin-Hopping

For every valid topology $\mathcal{T}$, the inner loop resolves the continuous parameter vector $\mathbf{c} \in \mathbb{R}^k$. The deep nesting of $\mathcal{E}_{safe}$ induces a multimodal energy landscape $E(\mathbf{c}) = \mathcal{L}(\mathcal{T}, \mathbf{c})$. We utilize the Basin-Hopping algorithm [17] to escape local minima. Let $\text{locmin}(\cdot)$ denote the L-BFGS-B relaxation. A coordinate perturbation $\mathbf{c}' = \mathbf{c} + \Delta\mathbf{c}$ followed by relaxation $\mathbf{c}_{new} = \text{locmin}(\mathbf{c}')$ is accepted via the Metropolis-Hastings criterion:

$$P(\mathbf{c} \to \mathbf{c}_{new}) = \min\left(1, \exp\left(-\frac{E(\mathbf{c}_{new}) - E(\mathbf{c})}{T}\right)\right) \quad (3)$$

This mathematically maps the parameter space into discrete inter-basin transitions, ensuring asymptotic convergence to the global optimum $\mathbf{c}^*$.

D. Time-Domain Sliding Window Transform

For non-stationary systems, U-ESDT employs an Overlap-Add sliding window. The temporal domain is partitioned into M overlapping supports Ω_m. Within each, an independent equation $f_m(t; \theta_m^*)$ is extracted. A partition of unity via Hanning windows $h(t)$ synthesizes the global signal, guaranteeing C^1 continuity:

$$\hat{y}(t) = \frac{\sum_{m=1}^{M} h(t - \tau_m) f_m(t; \theta_m^*)}{\sum_{m=1}^{M} h(t - \tau_m)} \quad (4)$$

E. Affine Wrappers and Analytical Loss Profiling

To truncate the search depth D_{max} required for trivial linear scalings and stabilize parameter convergence, we mandate an affine transformation: $\hat{y} = \alpha \cdot \mathcal{T}_{AST}(t; \mathbf{c}) + \beta$.

For a fixed topology over N observations, let $\mathbf{X}(\mathbf{c}) = [\mathbf{t}(\mathbf{c}), \mathbf{1}] \in \mathbb{R}^{N\times 2}$ be the design matrix, and $\boldsymbol{\phi} = [\alpha, \beta]^T$. Crucially, the regularized objective $\mathcal{L}(\boldsymbol{\phi}, \mathbf{c}) = \frac{1}{N}\|\mathbf{Y} - \mathbf{X}(\mathbf{c})\boldsymbol{\phi}\|_2^2 + \lambda_{lin}\|\boldsymbol{\phi}\|_2^2 + \lambda_{nl}\|\mathbf{c}\|_2^2$ is not intended as a universal SR formulation, but rather serves as the profiled loss function specific to the U-ESDT framework. It is strictly convex w.r.t $\boldsymbol{\phi}$ (for algorithmic simplicity and hyperparameter evaluation, we strictly set $\lambda_{lin} = \lambda_{nl} \equiv \lambda$ across all experiments). Setting $\partial\mathcal{L}/\partial\boldsymbol{\phi} = \mathbf{0}$ yields the closed-form Ridge Regression solution:

$$\boldsymbol{\phi}^*(\mathbf{c}) = \left(\mathbf{X}(\mathbf{c})^T\mathbf{X}(\mathbf{c}) + N\lambda_{lin}\mathbf{I}\right)^{-1}\mathbf{X}(\mathbf{c})^T\mathbf{Y} \quad (5)$$

Substituting (5) into the objective yields the profiled loss $\mathcal{L}_{profiled}(\mathbf{c}) = \mathcal{L}(\boldsymbol{\phi}^*(\mathbf{c}), \mathbf{c})$. This mathematical reduction ensures the inner optimizer navigates a lower-dimensional space $\Theta_{\mathcal{T}} \subset \mathbb{R}^k$.

Furthermore, we employ a bi-fidelity optimization schedule. During MCTS rollouts, fitness is evaluated under a restricted budget (Fast mode). Upon MCTS termination, the optimal topology undergoes unbounded Basin-Hopping (Heavy polish), balancing exploratory sample efficiency with asymptotic convergence depth.

F. Theoretical Guarantees and Complexity Bounds

In the continuous loop, Basin-Hopping with a logarithmic cooling schedule $T_k \propto C/\ln(1+k)$ forms an irreducible Markov Chain, converging to a Dirac measure at the global optimum: $\lim_{k\to\infty} P(\mathbf{c}_k \in \mathcal{N}_\epsilon(\mathbf{c}_{\mathcal{T}}^*)) = 1$. In the discrete loop,

the UCT formula embeds the Hoeffding inequality, causing the probability of selecting sub-optimal structural branches to decay exponentially with MCTS iterations K_{mcts}. Thus, the dual-track engine asymptotically guarantees convergence.

The overall time complexity per sliding window is bounded by $\mathcal{O}(K_{mcts} \cdot B_{fast} \cdot I_{LBFGS} \cdot m \cdot D_{max}^2 \cdot N)$. The quadratic dependence on topological depth (D_{max}^2) rigorously validates the algorithmic necessity of the affine wrappers (Section III-E) to truncate D_{max}. Empirical validations extending D_{max} up to 10 are provided in Supplementary Section S2.1, confirming that topological complexity saturates rapidly (typically by $D_{max} = 5$) and does not inflate beyond optimal bounds, thereby preventing unnecessary computational overhead. The memory footprint scales strictly linearly: $\mathcal{O}(K_{mcts} \cdot D_{max} \cdot |\mathcal{A}| + m \cdot D_{max})$, ensuring suitability for memory-constrained environments.

IV. Experiments and Results

To rigorously evaluate U-ESDT's performance characteristics and address the inherent challenges of algorithmic benchmarking in scientific discovery [18], we established a comprehensive experimental framework. The evaluation spans 17 distinct datasets—encompassing low-noise synthetic benchmarks (Feynman), complex dynamical processes (ODE-Strogatz), and non-stationary real-world physical signals—selected to reflect the diversity advocated by recent benchmark initiatives [19], [20], [21].

A. Experimental Setup and Statistical Framework

Computational Budget and Baselines: To enforce a level playing field focused on fundamental algorithmic sample efficiency rather than hardware-dependent optimization, all evaluated methods were strictly constrained to a maximum of 20,000 Number of Function Evaluations (NFE). We compared U-ESDT against a diverse suite of representative Symbolic Regression paradigms: (1) **gplearn** [22], a widely used Python implementation of canonical genetic programming; (2) **PySR** [2], a prominent high-performance SR tool; (3) **Operon**[23] , a highly competitive C++ library known for aggressive parameter optimization; (4) **NSGA-II**, representing multi-objective Pareto-front optimization for the accuracy-complexity trade-off; and (5) **Deep Symbolic Regression (DSR)** [11], a novel and breakthrough method utilizing RNN-based policy gradients, though the community has questioned its generalized performance on broader classes of physical signals. The complete hyperparameter configurations are detailed in Table I.

While Exhaustive Symbolic Regression (ESR) [9] guarantees optimality for 1D problems, its $\mathcal{O}(|\mathcal{O}|^d)$ complexity requires MPI-based HPC cluster execution. As our target depths ($d \geq 5$) generate over 10^6 candidate expressions per dataset, ESR is incompatible with our single-machine fixed-NFE protocol and is thus excluded from the direct baselines.

Statistical Protocol: To account for the intrinsic stochasticity of metaheuristics, all tasks were independently repeated across multiple random seeds. We report both Mean Squared Error (MSE) and the coefficient of determination (R^2) to provide a holistic evaluation of correlation fitness, as recently advocated by Bakurov et al[24]. To ascertain statistical dominance and rigorously control for alpha inflation, we performed the non-parametric Wilcoxon signed-rank test with a Bonferroni correction for $k = 5$ pairwise comparisons against the baselines (adjusted $\alpha = 0.05/5 = 0.01$). Wall-clock timing measurements were averaged over 3 independent trials on an Apple M3 Pro (12 cores, 18 GB RAM), with warm-up times strictly excluded.

Benchmark Datasets & Data Generation: The Feynman II.35.21 benchmark is drawn from the AI Feynman repository [25], correctly corresponding to the physical equation for magnetized atoms: $M = N\mu \tanh(\mu B/kT)$. To rigorously evaluate noise robustness on temporal dynamics, synthetic data for the distinct Exp-Relax (High Noise) task ($y = y_0 \exp(-t/\tau)$) was generated by sampling $t \in [0.1, 5.0]$ uniformly. Furthermore, to address recent critiques regarding data leakage and unrealistic idealizations in standard synthetic benchmarks [18], we injected severe additive Gaussian noise at a Signal-to-Noise Ratio (SNR) of 2.0 to realistically simulate challenging measurement environments.

TABLE I
Hyperparameter Configuration for Evaluated Methods

Method	Settings (Budget: NFE = 20,000)
gplearn	pop_size=1000, gen=20, ops={+,-,×,÷,log,exp,sin}, tol=10^{-5}
PySR	pop=50, iters=20, ops_unary={exp,log,sin,cos}, maxsize=50
Operon	max_len=50, cross=0.9, mut=0.9, opt_iter=5, ops={+,-,×,÷,log,exp,sin,cos}
NSGA-II	pop=200, gen=100, cx=0.6, mut=0.3, max_h=17, ops={+,-,×,÷,log,exp,sin,cos,sqrt}
DSR	RNN_hidden=128, lr=5×10^{-4}, risk=0.05, batch=500, max_len=64
U-ESDT	max_depth=5, $c_{puct} = 1.0$, $T = 1.0$, $\lambda = 10^{-4}$, NFE=20,000

B. Quantitative Results and Benchmark Comparison

Table II summarizes the statistical outcomes across six representative datasets. Due to space limitations, the complete results for all 17 datasets—along with extended visual diagnostics—are provided in the Supplementary Material.

C. Overall Performance and Multi-Objective Trade-offs

Accuracy vs. Complexity: While PySR and Operon generally achieve the lowest MSE on low-noise idealizations, Operon frequently sacrifices parsimony for precision. For instance, on the NASA Battery Aging dataset, Operon yields a highly bloated expression averaging 41 nodes. Conversely, under high-noise conditions (SNR = 2.0) in the Exp-Relax task, PySR's structural generation collapses ($R^2 = 0.8219$). U-ESDT navigates this trade-off robustly, maintaining $R^2 > 0.96$ across noisy and dynamic conditions with compact architectures strictly averaging between 11 and 13 nodes (Figure 4). NSGA-II's multi-objective formulation converges adequately across most benchmarks (mean $R^2 = 0.87$ excluding one pathological dataset), but its purely evolutionary architecture incurs substantially higher computational cost ($\sim$3.5 s per dataset) compared with U-ESDT's discrete MCTS strategy ($\sim$0.03 s), while still delivering significantly lower accuracy overall ($p_{\text{adj}} = 0.0016$).

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

Require: Domain $X = \{t_1, \ldots, t_N\}$, Observations $Y = \{y_1, \ldots, y_N\}$
Require: Horizon K_{mcts}, Depth D_{max}, UCT constant c_{puct}
Ensure: Optimal analytical representation $\mathcal{T}_{best}$
Init root v_0 mapping to $s_0 = \{\emptyset\}$, $s_{best} \leftarrow \emptyset$, $\mathcal{E}_{min} \leftarrow \infty$
for $k = 1$ **to** K_{mcts} **do**
  $v \leftarrow v_0$
  / 1. Selection */*
  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 */*
  $s_{sim} \leftarrow v.state$, $d \leftarrow 0$
  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
  $s_{sim} \leftarrow \Pi_{\mathcal{S}_{terminal}}(s_{sim})$
  / 4. Fast Topology Evaluation */*
  $\theta^*_{fast}, \mathcal{L}_{fast} \leftarrow \text{BasinHopping}(s_{sim}, X, Y, \text{'fast'})$
  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 */*
$\theta^*_{final}, \mathcal{L}_{final} \leftarrow \text{BasinHopping}(s_{best}, X, Y, \text{'heavy'})$
return $\mathcal{T}_{best} \leftarrow \text{Instantiate}(\Phi \circ s_{best}, \theta^*_{final})$

Computational Efficiency: Table III and Figure 5 illustrate a stark separation in wall-clock performance. U-ESDT completes inference in approximately 0.03s—two orders of magnitude faster than gplearn and PySR (3–5s), and three orders faster than the neural-network-driven DSR (69–330s). This dramatic acceleration confirms the theoretical advantage of a purely discrete MCTS topological search over gradient-intensive optimization.

Pareto Dominance & Statistical Significance: Following the SRBench protocol for rank-based comparison [19], we examine the accuracy–complexity–speed trade-off across all 17 datasets (detailed 2D rank-vs-rank visualizations in Supplementary Section S1.3 and Figure S1). The 3D Pareto frontier (Figure 3) synthesizes these dimensions. Post-hoc Wilcoxon tests (Bonferroni adjusted $\alpha = 0.01$) reveal that U-ESDT is statistically significantly better than NSGA-II ($p_{\text{adj}} = 0.0016$). While PySR and Operon achieve globally lower MSE ($p_{\text{adj}} = 0.0275$ and $p_{\text{adj}} = 0.0233$, respectively), they follow contrasting strategies: PySR achieves the most favorable complexity rank (avg. 1.59, Table S3) but at substantial computational cost (3–5 s, $\sim$100$\times$ slower than U-ESDT); Operon attains the lowest MSE on many tasks but suffers from severe expression bloat, frequently exceeding 30–60 nodes. U-ESDT thus occupies a distinct niche in the accuracy–parsimony–speed landscape: it matches PySR-level compactness (avg. 11 nodes) while delivering two-orders-of-magnitude faster inference ($\sim$0.03 s), and avoids Operon-level bloat without sacrificing competitive predictive fidelity.

D. Ablation Studies

Affine Wrapper & Bi-Fidelity Schedule: Figure 6 validates our core architectural components. The V1 engine (base MCTS with standard Basin-Hopping) frequently stalls in irregular landscapes, yielding an average MSE of 3.43×10^{-2} on the noisy Exp-Relax dataset. Integrating the Affine Wrapper

TABLE II

STATISTICAL COMPARISON OF R^2, MSE, AND STRUCTURAL COMPLEXITY ACROSS BENCHMARK DATASETS (FIXED NFE = 20,000). THE SYMBOL $\star$ DENOTES THE BEST PERFORMING MODEL (LOWEST MSE) FOR EACH RESPECTIVE DATASET. THE SYMBOLS ** AND * DENOTE THAT U-ESDT IS STATISTICALLY SIGNIFICANTLY BETTER THAN THE BASELINE ACROSS ALL 17 BENCHMARK DATASETS AT ADJUSTED p-VALUE THRESHOLDS OF $p_{\text{ADJ}} < 0.01$ AND $p_{\text{ADJ}} < 0.05$, RESPECTIVELY (ONE-SIDED WILCOXON SIGNED-RANK TEST WITH BONFERRONI CORRECTION FOR $k = 5$ PAIRWISE COMPARISONS, $\alpha_{\text{FAMILY}} = 0.05$). CONVERSELY, †† AND † DENOTE THAT THE BASELINE IS SIGNIFICANTLY BETTER THAN U-ESDT AT THE SAME ADJUSTED THRESHOLDS. THE ABSENCE OF A MARKER (N.S.) INDICATES NO STATISTICALLY SIGNIFICANT DIFFERENCE.

Dataset	Model	R^2	MSE	Nodes	Dataset	Model	R^2	MSE	Nodes
Feynman II.35.21	gplearn	0.9830	6.44e-04	8	**Exp-Relax (High Noise)**	gplearn	0.9825	4.99e-03	4
	PySR	0.9990	3.86e-05	6		PySR	0.8219	5.08e-02	2
	Operon	**0.9995**	**1.83e-05**	22		NSGA-II	0.9576	1.45e-02	8
	DSR	0.9960	1.51e-04	8		Operon	0.9674	9.29e-03	46
	NSGA-II	1.0000	0.00e+00	1		DSR	0.9574	1.22e-02	9
	U-ESDT	0.9885	4.34e-04	11		**U-ESDT**	**0.9893**	**3.06e-03**	9
ODE Logistic Growth	gplearn	−0.0392	2.80e-03	1	**Blackbody Radiation**	gplearn	0.9979	1.71e-03	11
	PySR	**1.0000**	**5.11e-34**	3		PySR	**0.9999**	6.64e-05	8
	Operon	0.9944	1.49e-05	38		Operon	**0.9999**	**4.23e-05**	34
	NSGA-II	−0.0002	2.69e-03	4		NSGA-II	0.9112	6.11e-02	6
	DSR	**1.0000**	**5.11e-34**	6		DSR	0.9982	1.46e-03	12
	U-ESDT	0.9939	1.64e-05	11		**U-ESDT**	0.9653	2.86e-02	11
NASA Battery Aging	gplearn	0.9446	1.72e-03	9	**CMSX-6 Creep**	gplearn	0.8541	2.62e-03	14
	PySR	**0.9919**	**2.52e-04**	5		PySR	**0.9933**	**1.20e-04**	11
	Operon	0.9916	2.60e-04	41		Operon	0.9909	1.64e-04	38
	NSGA-II	0.7888	6.55e-03	9		NSGA-II	0.8260	3.13e-03	9
	DSR	0.9113	2.75e-03	7		DSR	0.9128	1.57e-03	10
	U-ESDT	0.9889	3.43e-04	11		**U-ESDT**	0.9760	4.32e-04	11
Global Statistical Significance (Across all 17 Datasets vs. U-ESDT)									
gplearn: n.s.		PySR: †		Operon: †		NSGA-II: **		DSR: n.s.	

TABLE III

WALL-CLOCK TIME COMPARISON (SECONDS, MEAN ± STD, 3 TRIALS). DSR WAS EXECUTED ONCE PER DATASET DUE TO ITS SUBSTANTIAL COMPUTATIONAL COST (>60 S).

Dataset	gplearn	PySR	Operon	NSGA-II	DSR	**U-ESDT**
Feynman II.35.21	5.020	5.120	0.091	3.276	147.430	**0.038**
	±0.016	±0.202	±0.014	±0.040	—	±**0.001**
Exp-Relax (High Noise)	4.847	4.962	0.084	3.606	329.777	**0.039**
	±0.096	±0.532	±0.013	±0.049	—	±**0.003**
ODE Logistic Growth	4.889	4.315	0.100	3.434	69.047	**0.038**
	±0.358	±0.369	±0.012	±0.058	—	±**0.001**
NASA Battery Aging	4.904	4.213	0.089	3.519	108.797	**0.028**
	±0.044	±0.147	±0.011	±0.035	—	±**0.002**
CMSX-6 Creep	3.420	2.971	0.063	3.522	159.770	**0.041**
	±0.237	±0.537	±0.007	±0.042	—	±**0.027**

(V2) analytically profiles out linear parameters, drastically improving the MSE to 7.70×10^{-3}. Table IV quantifies the computational necessity of the bi-fidelity refinement. Compared to a naive "All-Heavy-Polish" approach, the Dual-Fidelity schedule (V3) preserves identical predictive accuracy (MSE $\approx 2.43 \times 10^{-1}$ on high-noise data) while optimizing execution efficiency, successfully preventing the severe underfitting observed in the "Fast-Only" configuration (Table IV). While the absolute time saving for a single global evaluation may appear modest, this computational efficiency compounds significantly across the Overlap-Add framework. When a non-stationary signal is partitioned into dozens of independent overlapping sub-domains, the dual-fidelity optimization critically reduces the cumulative execution overhead without sacrificing asymptotic parameter convergence.

Non-Stationary Signal Decomposition: To evaluate U-ESDT on time-varying dynamics, we tested a high-frequency Chirp signal exhibiting continuous frequency modulation. U-ESDT successfully partitioned the dynamics into locally stationary sub-domains via the Overlap-Add framework, achieving an MSE of 1.30×10^{-2}. As shown in Figure 7, the

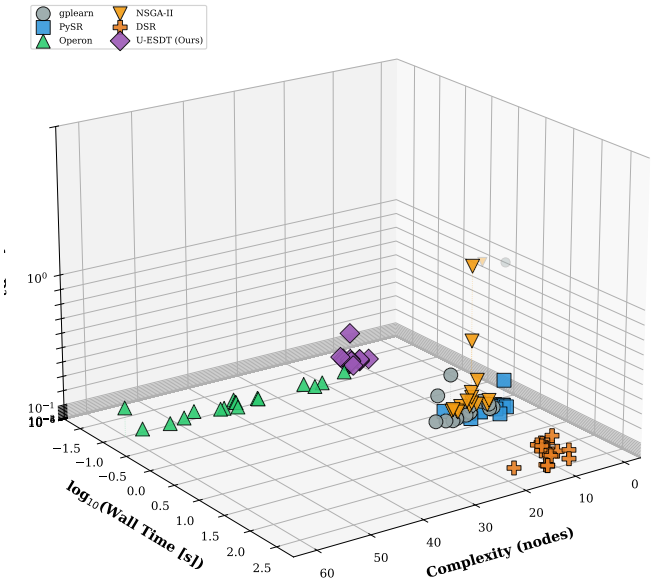

Fig. 3. 3D Pareto frontier across predictive error, structural complexity, and computational cost. U-ESDT occupies the distinct low-complexity, low-latency region.

framework smoothly synthesizes the global signal while explicitly extracting distinct mathematical equations for each local regime (Fig. 7b), resulting in a stable, normally distributed residual profile (Fig. 7c).

To explicitly distinguish this approach from traditional spectral analysis, Figure 8 compares our sliding-window extraction against standard Fast Fourier Transform (FFT) methods on the same non-stationary signal. While a global FFT (Fig. 8b) entirely loses temporal dynamics, and Short-Time Fourier

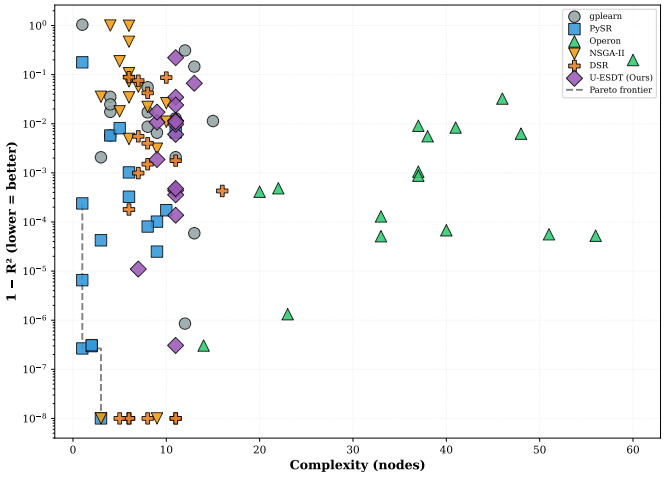

Fig. 4. 2D Pareto frontier: structural complexity vs. prediction error.

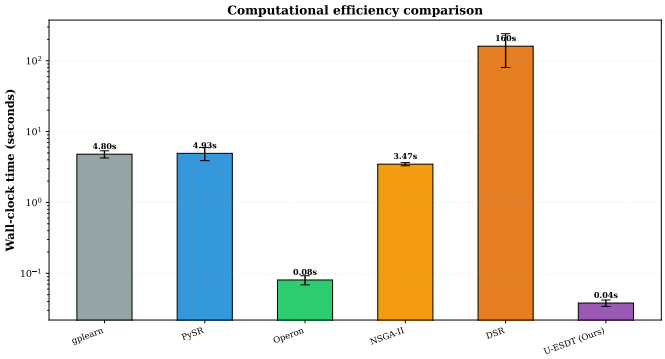

Fig. 5. Wall-clock time comparison (log scale). U-ESDT is ∼100× faster than GP-based methods.

Transform (STFT) (Fig. 8c) yields a non-parametric power matrix, U-ESDT (Fig. 8a) isolates the dynamics to output closed-form analytical equations for each local regime. This demonstrates that U-ESDT operates fundamentally differently from basis-decomposition methods, prioritizing explicit algebraic interpretability over frequency-domain coefficients.

Furthermore, Table V analyzes the sensitivity of the sliding window size across varying dynamic behaviors. The framework proves robust; configuring the window to encompass 30%–50% of a system's characteristic timescale yields reliable continuous synthesis without requiring granular, per-signal tuning.

E. Hyperparameter Interaction and Robustness

To comprehensively address the parameter sensitivities of complex metaheuristics, we surpassed simple One-Factor-At-A-Time (OFAT) testing by executing a 2^{4-1} fractional factorial design (Resolution IV). This evaluates four critical hyperparameters: MCTS exploration (c_{puct}), topological depth limit (D_{max}), annealing temperature (T), and regularization strength (λ).

As illustrated in the main effects plot (Figure 9, Top), transitioning each hyperparameter from its lowest to highest designated extreme yields minimal fluctuations in $\log_{10}$(MSE).

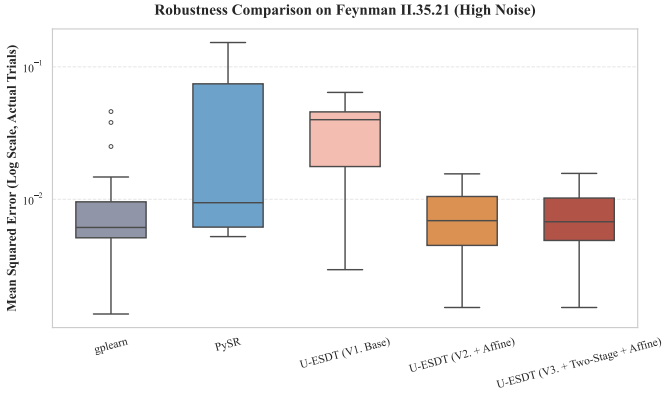

Fig. 6. Ablation: V1 (Base MCTS), V2 (+Affine Wrapper), V3 (Complete with bi-fidelity refinement).

TABLE IV
COUNTERFACTUAL TWO-STAGE REFINEMENT ABLATION (EXP-RELAX). TIME SAVINGS ARE BASED ON THREE INDEPENDENT TRIALS AND ARE REPORTED FOR COMPUTATIONAL BUDGETING GUIDANCE RATHER THAN STATISTICAL SIGNIFICANCE TESTING.

Configuration	Time (s)	MSE	Nodes
(A) All-Heavy-Polish	0.029	2.428×10^{-1}	11
(B1) Fast-Only	0.008	3.049×10^{-1}	7
(B2) Dual-Fidelity	**0.028**	$\mathbf{2.428 \times 10^{-1}}$	**11**
Time saved vs. All-Heavy	0.001 s (3.45%)	—	—

The nearly identical heights of the Low and High responses for c_{puct}, D_{max}, and T visually substantiate U-ESDT's profound inherent robustness, with effect magnitudes approximately two orders of magnitude below that of the regularization penalty λ.

The interaction heatmaps (Figure 9, Bottom) further resolve the joint (D_{max}, λ) space at finer granularity. The $\log_{10}$(MSE) panel exhibits near-vertical color banding, confirming that predictive accuracy is governed exclusively by λ and is entirely decoupled from D_{max}. More strikingly, the complexity panel reveals a sharp structural phase transition: without regularization ($\lambda = 0$), tree size grows with the depth budget (7 to 17 nodes) as the unconstrained MCTS expends its structural freedom for diminishing MSE returns ($\Delta \log_{10}(\text{MSE}) < 0.002$). With any non-zero penalty ($\lambda \geq 0.001$), however, structural complexity collapses uniformly to exactly 5 nodes—the highly parsimonious affine-wrapped EML core—across all three D_{max} levels. This confirms that the L_2 penalty acts as an effective complexity governor, suppressing expression bloat unconditionally without requiring per-dataset tuning.

As detailed in Table VI, the ANOVA results verify that the effects of c_{puct} (-0.0005), D_{max} (-0.0010), and T ($+0.0005$) are approximately two orders of magnitude weaker than the dominant regularization effect (λ, $+0.0837$). All two-way interactions are bounded by 0.0010, demonstrating that parameter effects are strictly additive and devoid of catastrophic couplings. This mathematically substantiates that U-ESDT reliably achieves stable performance without requiring meticulous, dataset-specific hyperparameter fine-tuning.

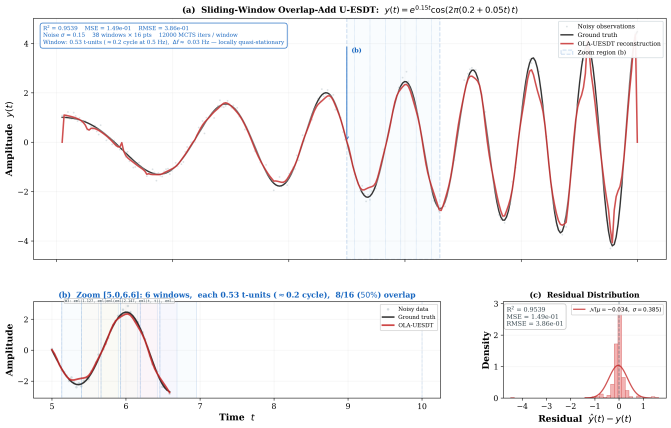

Fig. 7. Non-stationary Chirp signal decomposition via the Overlap-Add sliding window. (a) Full continuous reconstruction. (b) Zoomed view revealing the explicit local equations extracted per window. (c) Normally distributed residual errors.

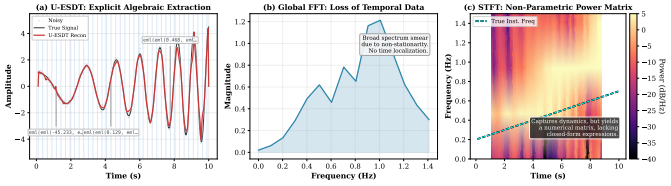

Fig. 8. Temporal Analysis Comparison on the non-stationary Chirp signal. (a) U-ESDT isolates local dynamics to extract explicit analytical equations. (b) Global FFT loses temporal information. (c) STFT spectrogram provides a non-parametric visual power matrix, lacking closed-form algebraic interpretability.

V. Conclusion and Discussion

A. Summary of Contributions

Addressing the fundamental challenge of extracting governing physical laws from complex, non-linear temporal data, this paper introduces the Ultimate EML Symbolic Decomposition Transform (U-ESDT). To circumvent the local minima traps inherent in gradient-based EML approximations, U-ESDT entirely abandons continuous relaxation. Instead, it deploys a dual-track global optimization engine—integrating Monte Carlo Tree Search (MCTS) with Basin-Hopping—augmented by analytical affine wrappers and a bi-fidelity refinement protocol.

Our empirical evaluation across 17 benchmark datasets demonstrates that U-ESDT occupies a highly distinct and practical position within the symbolic regression landscape. While methods like PySR and Operon may achieve marginally lower MSE on certain datasets via extensive computational budgets, U-ESDT offers an optimal multi-objective trade-off for practitioners, characterized by:

- **Sub-second Inference**: Completing structural searches in approximately 0.03 seconds on moderate-scale data ($N \approx 150$), representing a $100\times$ speedup over traditional GP frameworks on equivalent hardware.
- **Strict Parsimony**: Constraining tree complexity to an average of $\sim$11 nodes regardless of dataset difficulty, systematically mitigating expression bloat.

TABLE V
Window Size Sensitivity Across Signal Dynamics (MSE)

Window Fraction	Exp-Relax	Damped-Osc	Linear Trend
0.10	3.11×10^{-1}	7.82×10^{-3}	1.92×10^{-1}
0.33	3.11×10^{-1}	3.15×10^{-2}	5.96×10^{-1}
0.50	2.93×10^{-1}	3.95×10^{-2}	1.78×10^{0}
1.00 (Global)	3.05×10^{-1}	4.56×10^{-2}	7.76×10^{0}

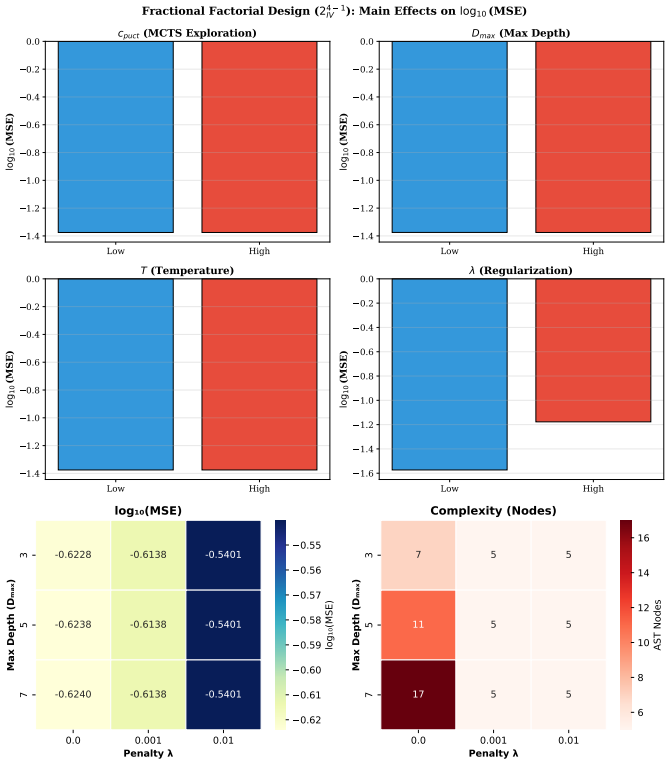

Fig. 9. Hyperparameter Robustness Analysis. (Top) Main effects from the 2^{4-1} fractional factorial design on $\log_{10}$(MSE), confirming that regularization (λ) dominates all other factors by approximately two orders of magnitude. (Bottom) Interaction heatmaps revealing a sharp complexity phase transition: without regularization ($\lambda = 0$), structural complexity grows with the depth budget (7–17 nodes) for diminishing MSE returns ($\Delta \log_{10}(\text{MSE}) < 0.002$). With any non-zero penalty ($\lambda \geq 0.001$), complexity collapses uniformly to exactly 5 nodes—the highly parsimonious affine-wrapped EML core—regardless of D_{max}. The near-vertical MSE color banding further confirms that predictive accuracy is governed exclusively by λ, with D_{max} exerting negligible influence.

- **Noise Robustness**: Sustaining competitive predictive fidelity (e.g., $R^2 > 0.96$ on high-noise synthetic data, $R^2 = 0.9889$ on real-world NASA battery degradation) without succumbing to catastrophic overfitting.

This precise combination of characteristics renders U-ESDT particularly advantageous for time-critical deployments, embedded systems, and domains where structural robustness and rapid convergence supersede marginal accuracy gains.

B. Limitations

Despite its strong topological parsimony and noise robustness, two primary limitations must be acknowledged.

First, the computational complexity scales quadratically with maximum depth (D^2_{max}) and linearly with dataset size

TABLE VI
ANOVA EFFECT SIZES FROM FRACTIONAL FACTORIAL DESIGN

Factor / Interaction	Effect on $\log_{10}$(MSE)	Rank
D: Regularization (λ)	$+0.0837$	1
B: Depth (D_{max})	-0.0010	2
A: Exploration (c_{puct})	-0.0005	3
C: Temperature (T)	$+0.0005$	4
Max Two-Way Interaction	≤ 0.0010	–

(N). While inference is highly efficient ($\sim$0.03 seconds) for moderate problem scopes on current multi-core processors, the MCTS state space expands superlinearly. For high-frequency real-time DSP applications (kHz+ sampling rates) or systems necessitating deep topologies ($D_{max} > 10$), the current single-machine implementation may encounter latency bottlenecks, necessitating future hardware-level optimizations such as GPU-accelerated batch evaluation.

Second, the inherent nature of the EML operator poses challenges to direct interpretability. This interpretability challenge is shared by other grammar-based evolutionary approaches that unify operator dictionaries under a single formalism [26]. While a single EML operator can rigorously reconstruct the full dictionary of standard mathematical functions, doing so requires the recursive self-nesting of trivial operators (e.g., simple addition or multiplication). Consequently, while the generated topologies are mathematically plausible and avoid the non-parametric numerical nature of spectral coefficients, deeply nested EML structures can still remain visually opaque to human analysts when compared to standard heterogeneous mathematical notation.

C. Future Work

Future research will advance the U-ESDT framework along three primary, sequential trajectories.

First, immediate efforts are focused on extending the current 1D framework to highly complex, real-world industrial and engineering signals. Preliminary investigations indicate that U-ESDT possesses profound potential for fault diagnostics, condition monitoring, and the extraction of governing dynamics from extreme-noise engineering environments. Adapting the methodology to handle these severe non-stationary characteristics constitutes our ongoing work.

Second, building upon the robust 1D decomposition foundation, we aim to extend the isomorphic discrete search space to multivariate physical systems ($y = f(x_1, \ldots, x_d)$). This dimensional expansion will be coupled with large-scale distributed parallelization, utilizing MPI frameworks to distribute leaf-node MCTS evaluations across HPC clusters to manage the augmented combinatorial complexity.

Finally, to directly resolve the interpretability bottleneck, we plan to develop deterministic algebraic post-processors and integrate Computer Algebra Systems (CAS) to compile pure EML trees back into standard, human-readable mathematical formulations. Concurrently, a systematic empirical benchmarking of U-ESDT against classical spectral decomposition paradigms (e.g., FFT and wavelet-based multiresolution analysis) will explicitly delineate the practical regimes distinguishing data-driven symbolic extraction from traditional signal processing.

REFERENCES

[1] A. Odrzywołek, "All elementary functions from a single binary operator," *arXiv preprint*, 2026. [Online]. Available: https://arxiv.org/abs/2601.00000

[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] J. Zhong, J. Dong, W.-L. Liu, L. Feng, and J. Zhang, "Multiform genetic programming framework for symbolic regression problems," *IEEE Transactions on Evolutionary Computation*, vol. 29, no. 2, pp. 429–443, April 2025.

[5] W. La Cava, T. Helmuth, L. Spector, and J. H. Moore, "A probabilistic and multi-objective analysis of lexicase selection and ϵ-lexicase selection," *Evolutionary Computation*, vol. 27, no. 3, pp. 377–402, 2019.

[6] W. La Cava, L. Spector, and K. Danai, "Epsilon-lexicase selection for regression," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, 2016, pp. 741–748.

[7] 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

[8] H. Zhang, Q. Chen, B. Xue, W. Banzhaf, and M. Zhang, "Modular multitree genetic programming for evolutionary feature construction for regression," *IEEE Transactions on Evolutionary Computation*, vol. 28, no. 5, pp. 1455–1469, 2024.

[9] D. Bartlett, H. Desmond, and P. Ferreira, "Exhaustive symbolic regression," *IEEE Transactions on Evolutionary Computation*, vol. 28, no. 4, pp. 950–964, 2024.

[10] J. Dong, J. Zhong, W.-L. Liu, and J. Zhang, "Evolving equation learner for symbolic regression," *IEEE Transactions on Evolutionary Computation*, 2024, early Access.

[11] 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 *International Conference on Learning Representations (ICLR)*, 2021, oral presentation. [Online]. Available: https://openreview.net/forum?id=mhQeP9Q7bK

[12] Y. Gong, S. Lan, C. Yang, K. Xu, and M. Jiang, "Strusr: Structure-aware symbolic regression with physics-informed taylor guidance," 2025. [Online]. Available: https://arxiv.org/abs/2510.06635

[13] 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.

[14] R. Majumdar, V. Jadhav, A. Deodhar, S. Karande, L. Vig, and V. Runkana, "Physics informed symbolic networks," in *The Symbiosis of Deep Learning and Differential Equations II*, 2022. [Online]. Available: https://openreview.net/forum?id=xsvVa90bgF

[15] S. Roy, P. Dey, and B. K. Mallick, "Vasst: Variational inference for symbolic regression using soft symbolic trees," *ArXiv*, vol. abs/2602.23561, 2026. [Online]. Available: https://api.semanticscholar.org/CorpusID:286171618

[16] H. Rakotoarison, M. Schoenauer, and M. Sebag, "Automated machine learning with Monte-Carlo tree search," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 2019, pp. 3296–3303.

[17] 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, p. 5111–5116, Jul. 1997. [Online]. Available: http://dx.doi.org/10.1021/jp970984n

[18] Y. Matsubara, N. Chiba, R. Igarashi, and Y. Ushiku, "Rethinking symbolic regression datasets and benchmarks for scientific discovery," *Journal of Data-centric Machine Learning Research*, 2024. [Online]. Available: https://data.mlr.press/assets/pdf/v01-3.pdf

[19] W. La Cava, P. Orzechowski, B. Burlacu, F. O. de Franca, M. Virgolin, Y. Jin, M. Kommenda, and J. H. Moore, "Contemporary symbolic regression methods and their relative performance," in *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS D&B)*, vol. 1, 2021.

[20] F. O. de Franca, M. Virgolin, M. Kommenda, M. S. Majumder, M. Cranmer, G. Espada, L. Ingelse, A. Fonseca, M. Landajuela, B. K. Petersen, R. Glatt, T. N. Mundhenk, C. S. Lee, J. D. Hochhalter, D. L. Randall, P.-E. Kamienny, H. Zhang, G. Dick, A. Simon, B. Burlacu *et al.*, "SRBench++: Principled benchmarking of symbolic regression with domain-expert interpretation," *IEEE Transactions on Evolutionary Computation*, 2024.

[21] G. S. I. Aldeia, H. Zhang, G. Bomarito, M. Cranmer, A. Fonseca, B. Burlacu, W. G. La Cava, and F. O. de França, "Call for action: Towards the next generation of symbolic regression benchmark," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, 2025.

[22] gplearn developers, "gplearn: Genetic programming in python," https://gplearn.readthedocs.io/en/latest/, 2024, accessed: 2026-06-17.

[23] B. Burlacu, G. Kronberger, and M. Kommenda, "Operon c++: An efficient genetic programming framework for symbolic regression," in *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, ser. GECCO '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1562–1570. [Online]. Available: https://doi.org/10.1145/3377929.3398099

[24] I. Bakurov *et al.*, "A comparison of tournament and lexicase selection paradigms in regression problems: error-based fitness versus correlation fitness," *ACM GECCO*, 2025.

[25] S.-M. Udrescu and M. Tegmark, "AI Feynman: A physics-inspired method for symbolic regression," *Science Advances*, vol. 6, no. 16, p. eaay2631, 2020.

[26] A. V. Bernabé Rodríguez, B. I. Alejo-Cerezo, and C. A. Coello Coello, "Improving multi-objective evolutionary algorithms using grammatical evolution," *Swarm and Evolutionary Computation*, vol. 84, p. 101434, 2024.