Concept Paper

A Physics-Guided and Self-Adaptive Multi-Agent Framework for Jet Anomaly Detection

Asifullah Khan * , Hamna Asif , Aleesha Zainab

Posted Date: 27 March 2026

doi: 10.20944/preprints202603.0958.v2

Keywords: jet anomaly detection; high-energy physics; graph neural networks; multi-agent systems; meta-learning; reinforcement learning; physics-guided learning; pile-up mitigation; relational reasoning; attention mechanisms

Concept Paper

A Physics-Guided and Self-Adaptive Multi-Agent Framework for Jet Anomaly Detection

Asifullah Khan [1,2,3,*], **Hamna Asif** [2] **and Aleesha Zainab** [1,2]

[1] Pattern Recognition Lab, DCIS, PIEAS, Nilore, Islamabad 45650, Pakistan
[2] PIEAS Artificial Intelligence Center (PAIC), PIEAS, Nilore, Islamabad 45650, Pakistan
[3] Deep Learning Lab, Center for Mathematical Sciences, PIEAS, Nilore, Islamabad 45650, Pakistan
* Correspondence: asif@pieas.edu.pk

Abstract

Jet anomaly detection in a high-energy physics is a non-stationary task that is fuelled by shifts in the domain due to pile-up and predominantly by background noise, and dynamically changing relationships between jet constituents in such a scenario, where a conventional graph neural network architecture is frequently inadequate in terms of robustness and interpretability. Physics-Self-Adaptive Multi-Agent System (PhySA-MAS) is a physics-directed, self-adaptive multi-agent architecture that proposes jet analysis as a decentralized and dynamically reconfigurable reasoning scheme. It does not use one monolithic model but instead integrates specialist agents dealing with meta-learning, relational reasoning, communication and topology control which can vary their interactions depending on event-level physics. The energy conservation constraints are embedded within graph message passing to ensure physical consistency, while a reinforcement-driven topology controller dynamically rewires inter-agent communication in response to anomalous patterns. An additional communication strategy, anchor-peer communication, ensures the further stabilization of learning through the reduction of gradient conflict and the amplification of the signals related to anomalies, which, in combination, offers a powerful and structurally understandable alternative to fixed deep learning models.

Keywords: jet anomaly detection; high-energy physics; graph neural networks; multi-agent systems; meta-learning; reinforcement learning; physics-guided learning; pile-up mitigation; relational reasoning; attention mechanisms

1. Motivation

Jet anomaly detection in HEP is a difficult learning problem which exposes structural limitations of standard deep learning models. Pile-up effects and Luminosity-dependent domain shifts add non-stationarity to the problem, in violation of the assumptions of fixed distribution on which most models are trained [1,2]. At the same time, rare anomalous signatures are often being diluted by dominant Quantum Chromodynamics (QCD) background radiation which reduces the discriminative power of globally pooled representations [3,4]. Heterogeneous feature relevance across jet constituents can also lead to gradient interference in uniform aggregation [5], whereas both intrinsic event dependency of relational structure of jets and the fixed graph connectivity definition are no longer sufficient in describing dynamic hierarchical interactions [6].

Graph neural networks (GNNs) have demonstrated great potential in the modeling of relationships in addition to jet tagging and particle-based representations [6,7]. Similarly, weakly-supervised, or unsupervised, methods for anomaly detection have shown potential in HEP [3,8]. However, these approaches are still essentially monolithic in nature: message passing is static, communication ways are static, and architectural assumptions do not change from one event to another [6,7]. Even physics-informed approaches to learning usually use domain constraints only at the level of loss instead of built into the relational propagation [9].

These observations indicate that robustness and interpretability issues do not only arise due to parameter learning, but can also be caused by architectural rigidity. We therefore propose to reframe the problem of jet anomaly detection as a distributed reasoning process rather than a single model inference process. This is the motivation behind PhySA-MAS, we re-conceptualize the anomaly detection to a coordinated behaviour of specialized agents that adapt and reorganize based on the physics of events in a dynamic way embedding structural adaptiveness and physics-aware computation right in the architecture.

The proposed framework introduces a novel architectural perspective by reframing jet anomaly detection as a distributed reasoning process rather than a conventional single-model inference task. Rather than relying on a monolithic deep learning architecture, the approach decomposes the detection pipeline into a coordinated set of specialized agents, each responsible for distinct yet interrelated functions, including meta-learning based adaptation, relational reasoning over particle interactions, structured information fusion, and dynamic topology control. This multi-agent formulation enables a shift from static computational pipelines to a collaborative and adaptive learning paradigm, in which agents communicate and reorganize their interactions in response to event-level physics conditions. By embedding physical constraints directly within relational message propagation and incorporating reinforcement-driven topology adaptation, the framework facilitates reasoning-oriented learning that captures complex particle dependencies beyond purely statistical patterns. Consequently, the model achieves improved robustness under non-stationary conditions such as pile-up and detector noise.

2. Overview of PhySA-MAS Architecture

In HEP, jet anomaly detection is not easy because of high-dimensional particle representations, non-stationarity caused by pile-up, dominant QCD background noise and event-dependent relational structure [1,2]. Rare anomalous signatures are usually lost in a sea of standard-model processes with overwhelming number, which translates into poor discriminative separability [3]. Conventional monolithic models such as jet images of CNN or static graph neural networks (GNNs) and ensemble methods often have problems to adapt with the dynamic luminance conditions and shifting detector response [4,5]. PhySa-MAS solves this limitation by reformulating the problem of anomaly detection as a distributed, physics-guided, prowling process in which several specialized agents work together and adapt based on event-level physics instead of using fixed inference process pipeline.

2.1. Core Concept

The basic concept behind PhySA-MAS is to incorporate the physics-awareness and structural adaptability directly into the architecture. Instead of using a big one-shot model, the framework orchestrates the actions of self-adaptive agents that adjust parameters, reason on the relationship of particles, fuse multiple outputs from different agents, and dynamically change forms of communication. By incorporating energy conservation principles and physically meaningful particle interactions directly into the relational propagation mechanism rather than enforcing physics only at the loss level, the system maintains physical consistency while selectively enhancing anomaly-relevant features [6]. This design further alleviates gradient conflicts and representation dilution commonly observed in monolithic deep networks [7].

2.2. Architecture Overview of the Proposed PhySA-MAS

PhySa-MAs comprises of four tightly coupled modules as shown in Figure 1:

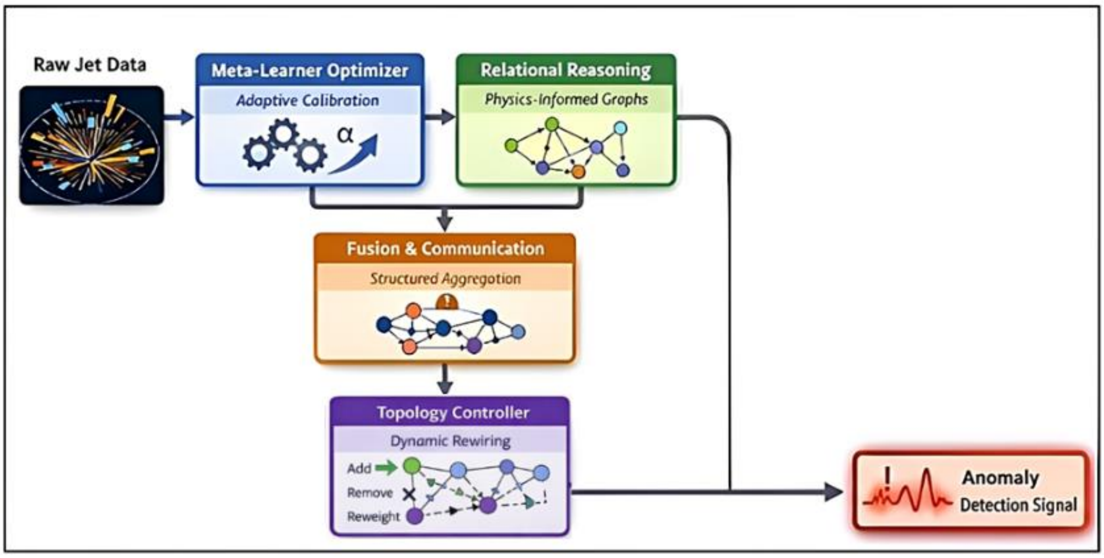


Figure 1. Conceptual Flow Diagram of PhySA-MAS.

1. Meta-Learner Optimizer Agent: To overcome the effects of domain shift and pile-up variability by fast-weight adaptation based on Model-Agnostic Meta-Learning (MAML) [8]. Instead of static parameters, the agent updates model weights with inner and outer loops, depending on the current luminosity and noise level. It takes raw jet features and event context as input and produces **calibrated parameters**, enabling robustness to pile-up and domain shifts as shown in Figure 2.

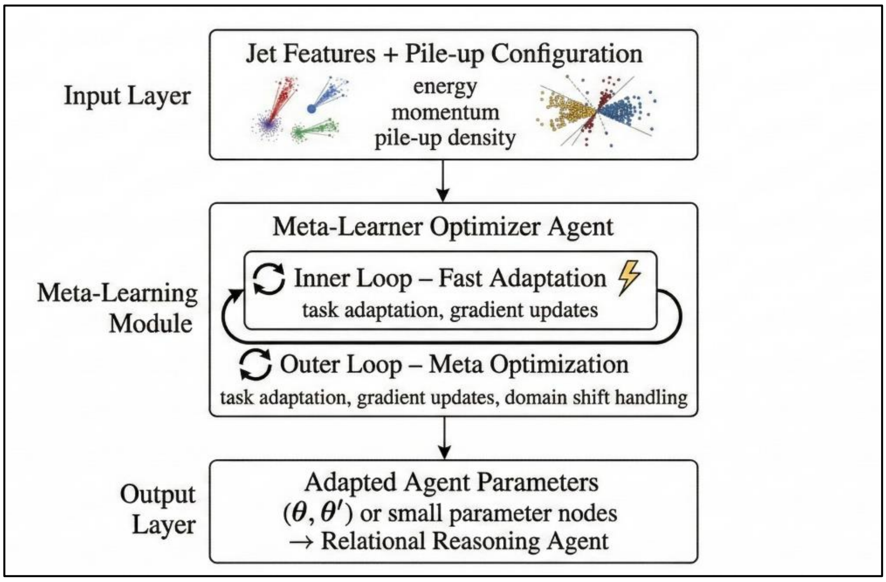


Figure 2. Meta-Learner Optimizer Agent performing inner-loop fast adaptation and outer-loop meta-optimization over jet features and pile-up configurations to produce calibrated parameters for downstream relational reasoning.

2. Relational Reasoning Agent: Overcomes limitations of Static GNN [5] by establishing a directed, asymmetric graph reasoning. It employs physics-aware confidence gating that prune or reweights the edges based on the kinematic constraints so that message passing only happens between physically correlated clusters of particles as shown in Figure 3.

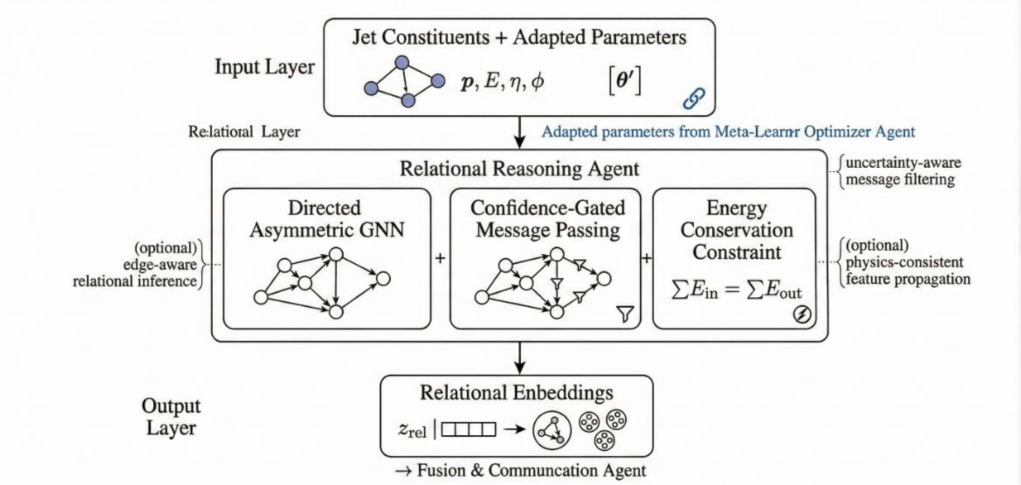


Figure 3. Relational Reasoning Agent combining directed asymmetric GNN, confidence-gated message passing, and energy conservation constraints to generate physics-consistent relational embeddings from jet constituent inputs.

3. Fusion & Communication Agent: This agent aggregates outputs from multiple agents using an anchor–peer broadcast mechanism with hyper-attention and manages how information is shared across the multi-agent system without getting "diluted". Mitigates gradient interference and signal dilution [9]. It produces a fused anomaly representation, along with auxiliary signals such as agent states, confidence scores, and uncertainty estimates, which serve as inputs to the topology controller as illustrated in Figure 4.

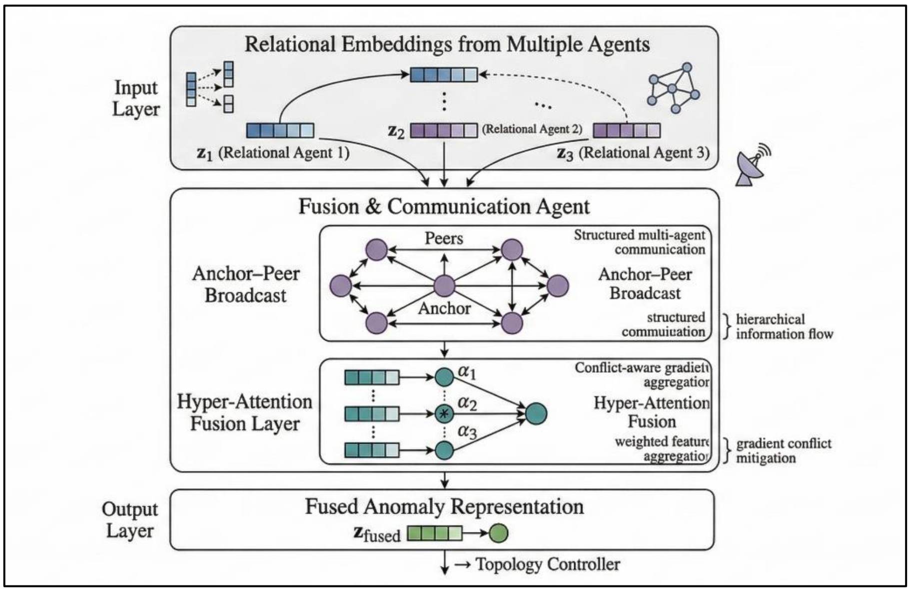


Figure 4. Fusion & Communication Agent aggregating relational embeddings from multiple reasoning agents via anchor-peer broadcast and hyper-attention fusion to produce a unified anomaly representation for the Topology Controller.

4. Topology Controller Agent: This agent is the "architect" that determines the structure of the team itself. The Topology Controller receives the fused anomaly representation and agent-level features as its state. It employs a **reinforcement learning policy $\pi(a \mid s)$** to dynamically modify the communication structure between agents as shown in Figure 5.

Through a topology adaptation mechanism, it performs operations such as adding, removing, or reweighting edges, resulting in an updated agent connectivity graph. This enables event-dependent communication and provides structural interpretability.

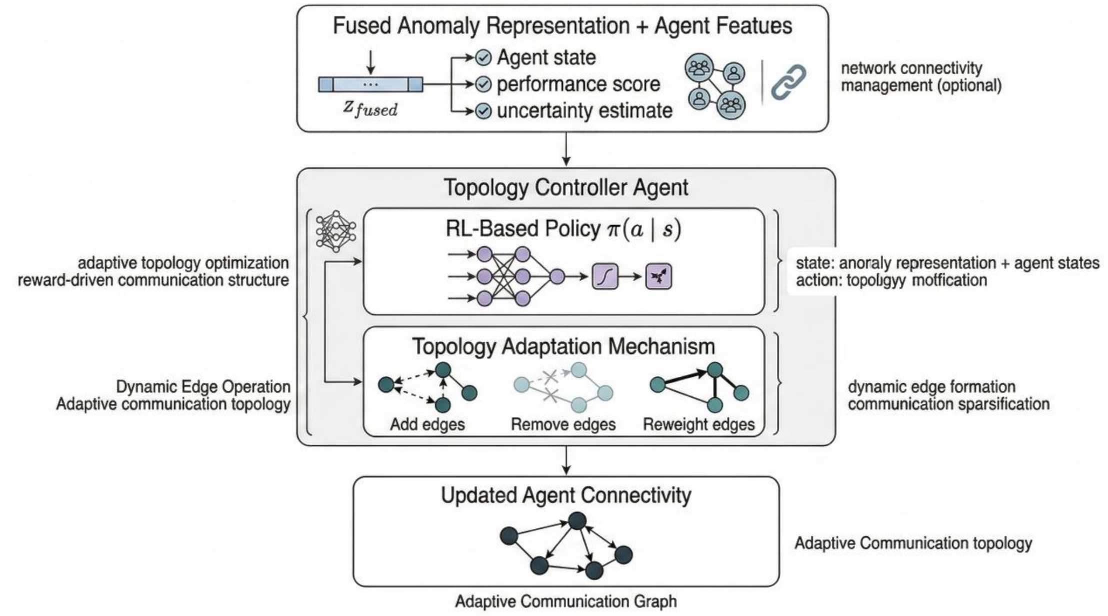


Figure 5. Topology Controller Agent dynamically rewiring inter-agent communication using an RL-based policy, performing event-conditioned edge addition, removal, and reweighting to yield an adaptive communication graph.

The Meta-Learner treats each jet as a task instance and employs rapid adaptation to improve robustness and maintain physical consistency [8]. The Relational Reasoning Agent builds directed constituent graphs for the hierarchical jet formation and controls the propagation using the confidence scores and constraints sensitive to conservation [5]. The Fusion & Communication Agent aggregates representations using structured attention [9], while the Topology Controller optimizes connectivity policies through reinforcement learning (RL) [10].

All four agents are a part of a closed feedback loop with calibrated parameters providing input for relational reasoning; relational outputs being fused together into a common representation of the anomaly; the fused signal providing input for topology adaptation and updated connectivity providing input for further reasoning and fusion. This constant adaptation helps PhySa-MAS to be robust to noise, domain shifts, and complex event interactions.

Table 1. Agent-Level Functional Design and Physics Constraints in the Proposed PhySA-MAS.

Agent	Function	Input	Output	Physics Integration	Adaptivity
Meta-Learner Optimizer	Event-level adaptation	Raw jet features, event context	Calibrated agent parameters	Maintains energy/momentum consistency	Inner/outer-loop meta-learning
Relational Reasoning	Physics-informed asymmetric GNN	Node embeddings, calibrated parameters	Node messages, confidence scores	Energy conservation, directional edges	Event-aware message weighting
Fusion & Communication	Structured multi-agent	Node messages,	Unified anomaly signal	Preserves local feature consistency	Anchor–peer hyper-attention

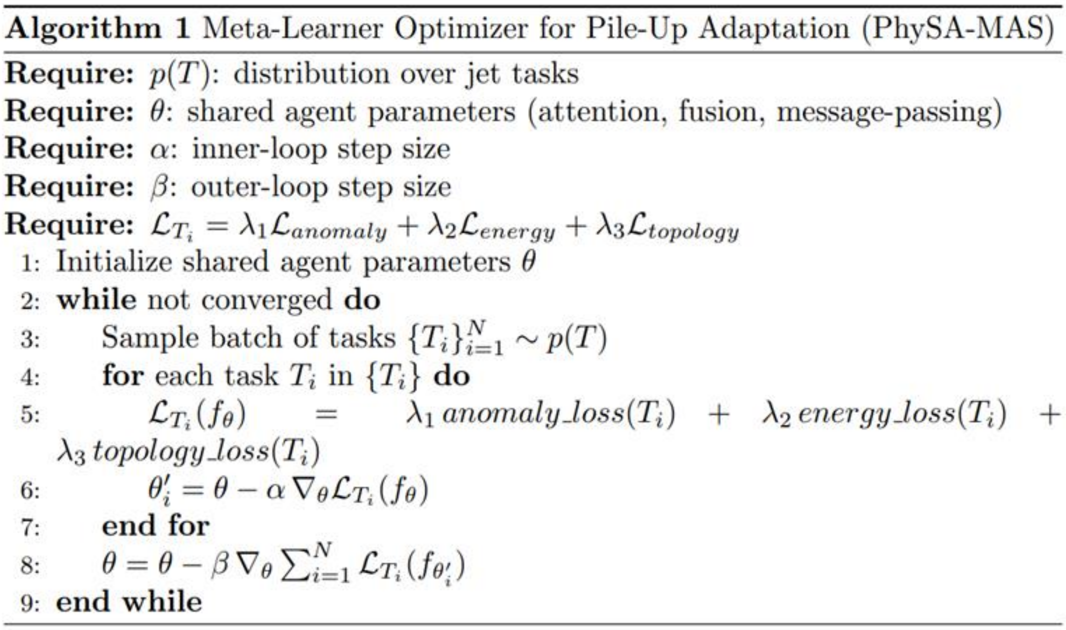
	aggregation	confidence scores			
Topology Controller	Dynamic inter-agent graph rewiring	Fused signals, intermediate outputs	Updated agent connectivity	Physically plausible edges	RL-based edge optimization

3. Meta-Learner Optimizer: Fast Adaptation to Pile-Up

In high-energy particle physics, pileup occurs when multiple proton-proton collisions happen simultaneously during a single bunch crossing, creating a complex background of overlapping signals [1,2]. This background is dynamic, with its density scaling with the instantaneous luminosity of the beams. For example, earlier experiments experienced an average pileup multiplicity $\langle N_{pu} \rangle \approx 5$ [2], Run 3 routinely reaches $\langle N_{pu} \rangle \approx 60$ interactions per crossing, and the upcoming High-Luminosity LHC is expected to operate with up to ~200 [4].

In jet-based anomaly detection, different pileup conditions pose essential challenges [1,2]. In poor luminosity regimes, the detector signals are relatively clean and have minimal, overlapping collisions. On the other hand, the high luminosity regimes introduce severe stochastic fluctuations in the energy density so that they can obfuscate real anomalous signatures [4]. Standard machine learning models trained according to a set luminosity assumption tend to become invalid under such distribution shifts, resulting in false positives in high noise or a failure to detect signals in cleaner regimes [5].

To tackle this variance, various meta-learning techniques known as model-agnostic meta-learning (MAML) inspired fast-weight adaptation [11], can be used to effectively implement a luminosity-aware tuner. Each of the jet events or mini-batches is treated as a task under a given pileup configuration, and some parameters of the agent, such as attention scaling factors and fusion coefficients are made to be conditional on local pileup properties. The meta-learner is therefore able to learn initialization parameters that can quickly adapt to new pileup regimes through only a small number of gradient steps [11], which guarantees the robustness and pileup-aware nature of anomaly detection.



```
Algorithm 1 Meta-Learner Optimizer for Pile-Up Adaptation (PhySA-MAS)
Require: $p(T)$: distribution over jet tasks
Require: $\theta$: shared agent parameters (attention, fusion, message-passing)
Require: $\alpha$: inner-loop step size
Require: $\beta$: outer-loop step size
Require: $\mathcal{L}_{T_i} = \lambda_1 \mathcal{L}_{anomaly} + \lambda_2 \mathcal{L}_{energy} + \lambda_3 \mathcal{L}_{topology}$
Initialize shared agent parameters $\theta$
while not converged do
    Sample batch of tasks $\{T_i\}_{i=1}^{N} \sim p(T)$
    for each task $T_i$ in $\{T_i\}$ do
        $\mathcal{L}_{T_i}(f_\theta) = \lambda_1\, anomaly\_loss(T_i) + \lambda_2\, energy\_loss(T_i) + \lambda_3\, topology\_loss(T_i)$
        $\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{T_i}(f_\theta)$
    end for
    $\theta = \theta - \beta \nabla_\theta \sum_{i=1}^{N} \mathcal{L}_{T_i}(f_{\theta_i'})$
end while
```

Figure 2. Physics-Informed Meta-Learning for Adaptive Jet Anomaly Detection under Pile-Up Variations.

Formally, let each task correspond to a jet event or mini-batch with a specific pile up density, detector noise profile and energy contamination pattern. The model is comprised of the coordinated agents (or a subset of the agents that share common parameters) while optimization is done with respect to a physics-aware loss function with relational consistency and conservation constraints [9]. This formulation allows for the tasks to be adapted specially without losing the structural coherence in particle-based representations [6,7,16].

The MAML-based Meta-Learner agent learns quantities such that a single step of gradient descent under a new pile-up condition takes the system towards a physically consistent representation, therefore reducing false anomaly signals, resulting from detector artifacts [11].

Table 2. Inner–Outer Loop Meta-Learning Mechanism and Its Physics-Constrained Integration within the Proposed PhySA-MAS Framework.

Aspect	Description	Inputs	Outputs	Physics Integration	Adaptivity
Inner-Loop Updates	Task-specific parameter updates to reduce pile-up and detector noise	Jet features, event context	Adapted agent parameters	Maintains energy/momentum consistency	Event-specific, fast updates
Outer-Loop Optimization	Generalization across tasks to maintain robust performance	Distribution of tasks	Updated meta-parameters	Ensures consistent physical reasoning	Global, across-luminosity
Integration with Relational Agent	Calibrated parameters improve downstream reasoning	Calibrated inner-loop outputs	Refined relational features	Preserves physically meaningful relationships	Adaptive to changing pile-up
Real-Time Application	Enables high-throughput anomaly detection	Live jet data	Rapidly adapted embeddings	Physically consistent features	Low-latency, scalable

4. Asymmetric GNNs for Relational Reasoning

Relational modeling of jet constituents in HEP benefits from the explicit inclusion of physical hierarchies and energy conservation in graph structures [1,2,6,7]. The Relational Reasoning Agent in PhySA-MAS makes use of asymmetric graph building, confident message passing, and energy conservation regularizer to generate physics consistent and anomaly sensitive representations.

Node features for each constituent include standard kinematic variables (p_T, η, φ, E) relative to the jet axis, following established practice in constituent-level jet representations. Edge features encode pairwise geometric and kinematic quantities including ΔR_{ij} and momentum fraction z_{ij}. Jet constituents are assumed to be particle-flow candidates pre-clustered via the anti-k_T algorithm.

4.1. Asymmetric Graph Construction

Jet constituents are modeled as nodes in a **directed graph**, where edges capture asymmetric relationships such as leading-to-soft radiation influence, reflecting the hierarchical structure of jet formation [3,4]. Formally, each jet is represented as $G = (V, E)$, where (V) contains (N) constituents. Directed edges $(v_j, v_i) \in E$ are defined so that information flows from higher p_T leading nodes to lower p_T daughter nodes within a local radius ($\Delta R_{ij} < R_{max}$). This ensures that the core jet representation remains robust to noise from the soft sector, while preserving the causal structure of parton fragmentation.

4.2. Confidence-Gated Message Passing

Each node has a learnable confidence, which represents the degree to which it is relevant to potentially anomalous substructures [6,7]. During message passing, this score gates outgoing signals so that high-confidence nodes propagate stronger messages, while nodes dominated by pile-up, soft radiation, or detector noise are suppressed. In training, nodes that consistently improve anomaly detection gradually acquire higher confidence scores, enabling selective relational reasoning and reducing interference from irrelevant features.

4.3. Energy Conservation Regularizer (Physics Constraint)

An energy conservation regularizer is incorporated in every message-passing layer in order to implement physical consistency. At the node level, the accumulated outgoing message energies may not be more than the input energy of any node, so there is no amplification of unphysical influence. At the graph level, the overall energy sent through is made soft to be equal to the reassembled jet energy. It is a soft constraint that penalizes energy deviation measured by the energy, which directs the network to concentrate on significant energy flows, and stabilizing training [9]. The inclusion of these physics' principles in the purely relationship propagation also helps the agent to generate more reliable anomaly detection without the need of post-hoc corrections.

These elements can be combined together to allow the Relational Reasoning Agent to reason about physics, direction-aware, and anomaly-oriented graph reasoning and to have strong embeddings in downstream multi-agent fusion and topology modules.

5. Fusion & Communication: Anchor-Peer Hyper Attention

In PhySA-MAS, Fusion & Communication Agent combines the reports of several other dedicated agents into one anomaly signal [6,7]. In contrast to more traditional models which use simple concatenation or pooling, it uses a designed Anchor-peer-Broadcast scheme to selectively boost the relevant signal and minimise conflicts on the feature heterogeneity [12,13].

5.1. Anchor-Peer Framework

The anchor-peer design differentiates between the anchor agents, the ones that give high-confidence anomaly signals and the peer agents, the ones that give contextual information. These anchors transmit their outputs to peers, which use a hyper-attention mechanism that is conditioned on event-level features, such as particle kinematics and a graph topology [12]. This is a selective aggregation, which avoids soaking up anomaly signals that are rare and thus learning becomes concentrated on the most useful interactions.

5.2. Reducing Gradient Conflict

Broadcasting is structured, which removes gradient conflict [13]. Gradient flow is guided by anchor, while peers influence each other depending on their relevance, which limits the discontinuous interference of opposing signals. This avoids overfitting to noise by stabilizing training and converting into the jet pattern anomalies. Furthermore, it is fast.

5.3. Selective Amplification of Anomaly Signals

The fusion agent selectively enhances patterns that are relevant to anomaly by conditioning peer attention to features of events. Outliers in jet substructure or anomalous energy patterns receive enhanced importance in a fused signal, and this increases the sensitivity of detection, when the background processes are dominant (e.g., in comparison to the background signal) [12,14].

5.4. Integration Within PhySA-MAS

The combined output is sent to the Topology Controller, which is the completion of the multi-agent loop. This allows structural adjustment downstream and interpretability, showing what anchors and peer interactions were the most important to anomaly detection. The mechanism offers an integrated and physics-conscious mechanism, which increases the strength and sensitivity of jet abnormality recognition.

6. Self-Organizing Topology via Reinforcement Learning

6.1. Motivation

Generally, the communication graphs of interest are fixed, and this is not realistic in highly fluctuating jet events [7,10]. PhySA-MAS addresses this shortcoming by permitting the adaptation of the agent topology dynamically according to each event and enhancing its robustness, sensitivity and the exploitation of anomalies.

6.2. Topology Controller Based on RL

Topology Controller Agent tries to control the inter-agent communication with the help of RL [10,15]. It unlearns a policy and adds, removes or updates weights between agents using jet-level observables and intermediate representations. Reward functionality provides a balance between anomaly recognition capabilities, communication performance, and physical realism to provide topology adjustment to be effective and comprehensible.

6.3. Agent Re-Wiring of Anomalous Jets

There is also topological interpretation that is offered by topology adaptation. When anomalies occur, the pattern of communication is not the same as that of the background jets, and some of the agents, like relational or fusion anchors, become central to the whole process. The particular topological patterns tend to be associated with the particular classes of anomalies [10,15]. The patterns of rewiring are valuable to visualize the reasoning of the system, and then it becomes clear why certain events are identified as anomalous.

7. Training Strategy

PhySA-MAS uses a training strategy based on the curriculum to have stable learning in its multi-agent architecture [6,10]. In the first technique, the relational and fusion agents are trained with a fixed communication topology, and the network is able to learn strong representations without the additional complexity of adaptive connections. This is followed by the introduction of meta-learning adaptation, operating under controlled variations of the pile-up, favoring the adaptation of agent parameters to each event, remaining physically consistent [6,8]. Lastly, the topology learning component gets triggered, in which the Topology Controller tries adding, removing, and reweighting edges using constrained action space in order to prevent unphysical or unstable structures [10,15].

This re-arranged curriculum alleviates the instability that is typical when meta-learning is used together with RL. PhySa-MAS does so by first learning complexity before adapting, and lastly by maximizing the structure. Consequently, PhySA-MAS is able to learn to have robust, physics-aware anomaly detection with interpretable agent interactions and stable training dynamics.

8. Predicted Benefits of Jet Anomaly Detection

PhySA-MAS offers a number of unique benefits to the jet anomaly detection:

- **Stability to pile-up and domain change:** The Meta-Learner Optimizer enables event-specific learning, ensuring stability across a range of luminosity and collision conditions [1,2,6,15].

- **Less signal dilution:** Relational reasoning based on asymmetric and confidence-gated features ensures that anomaly-relevant patterns are amplified, while background signals and pile-up contributions are suppressed [4,7,14].
- **Quicker convergence:** The anchor-peer hyper-attention mechanism of conflict-stable gradient flow synchronizes training and speeds it up by eliminating interference due to dissimilar feature importance [5].
- **Event-level adaptivity and interpretability:** The self-organizing agent topology, controlled by RL, reveals important inter-agent interactions and topological structures that emerge during anomalous events. These structures provide structural insights into the reasoning process of the system [10].

Collectively, these characteristics can be used to detect anomalies with physics sensitivity, adaptability, and interpretability, as well as mitigating the weaknesses of entirely stationary, monolithic designs and providing well-behaved detection in the presence of complicated, non-stationary processes.

9. Conclusion

The idea of MAS, its adaptability and emergent intelligence reasoning capabilities can be applied to diverse problems of HEP. PhySA-MAS is a move in the direction of physics-native agentic learning, in which inductive biases, physical constraints and structural adaptivity are built into the architecture as opposed to being imposed on it externally. The framework shows how multi-agent systems can evolve in non-stationary, high-dimensional, and noisy environments as it is interpretable and physically consistent by implementing meta-learning, relational reasoning, structured fusion, and self-organizing topology.

Even though proposed to monitor jet anomalies in HEP, the concepts supporting PhySA-MAS can be generally extended to other non-stationary physics systems, such as astrophysics, computer-based plasma simulation and experimental particle detector measurements. Subsequent research looks towards the formalisation of theoretical guarantees of topology adaptation, the framework to impose more conservation laws and symmetries [9], the scaling of the system to run on real time LHC data streams. The idea is that agentic, physics-aware learning will become a workable and comprehensible method of inference in complex science with these developments.

References

1. M. Cacciari, G. P. Salam, and G. Soyez, “Pileup subtraction using jet areas,” *Phys. Lett. B*, vol. 659, no. 1–2, pp. 119–126, 2008.
2. G. Soyez et al., “Pileup mitigation at the Large Hadron Collider,” *Phys. Rev. Lett.*, vol. 110, no. 16, 162001, 2013.
3. L. M. Dery et al., “Weakly supervised anomaly detection in particle physics,” *JHEP*, vol. 2021, no. 5, 2021.
4. A. J. Larkoski, I. Moult, and B. Nachman, “Jet substructure at the Large Hadron Collider: A review,” *Phys. Rept.*, vol. 841, pp. 1–63, 2020.
5. T. Chen et al., “Understanding deep learning requires rethinking generalization,” *ICLR*, 2017.
6. P. T. Komiske, E. M. Metodiev, and J. Thaler, “ParticleNet: Jet tagging via particle clouds,” *Phys. Rev. D*, vol. 101, 056019, 2020.
7. J. Shlomi, P. Battaglia, and J.-R. Vlimant, “Graph neural networks in particle physics,” *Mach. Learn. Sci. Technol.*, vol. 2, 021001, 2020.
8. B. Nachman and D. Shih, “Anomaly detection with density estimation,” *Phys. Rev. D*, vol. 101, 075042, 2020.
9. B. Cranmer et al., “Lagrangian neural networks,” *ICLR Workshop*, 2020.
10. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed., MIT Press, 2018.
11. C. Finn, P. Abbeel, and S. Levine, “Model-agnostic meta-learning for fast adaptation of deep networks,” in *Proc. ICML*, 2017.

12. A. Vaswani et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
13. H. Kim et al., "Hypergraph attention networks," *ICLR*, 2020.
14. J. Lee et al., "Graph attention networks," *ICLR*, 2019.
15. M. Gasse et al., "Exact combinatorial optimization with graph convolutional networks,"
16. M. Usman et al., "Particle Multi-Axis Transformer for Jet Tagging," arXiv:2406.06638, 2024.