

Article

STEGMN: Spatiotemporal Equivariant Graph Mechanics Networks for Molecular Trajectory Prediction

Yangyang Miao ¹  and Quan Qian ^{1,2,3,*} ¹ School of Computer Engineering & Science, Shanghai University, Shanghai 200444, China; miaoyy@shu.edu.cn² Center of Materials Informatics and Data Science, Materials Genome Institute, Shanghai University, Shanghai 200444, China³ Key Laboratory of Silicate Cultural Relics Conservation, Shanghai University, Ministry of Education, Shanghai 200444, China

* Correspondence: qqian@shu.edu.cn

Abstract

Molecular trajectory prediction is fundamental to computational chemistry, drug discovery, and materials simulation, enabling insights into dynamics, reaction pathways, and conformational stability. Its natural alignment with graph-structured spatiotemporal data has made it a key frontier in GNN research. However, current mainstream spatiotemporal GNNs, while enforcing E(3)-equivariance, treat atoms as unconstrained point masses and lack explicit rigid geometric constraints, often yielding unphysical deformations that compromise predictive interpretability. To address this challenge, we propose STEGMN—the first spatiotemporal graph architecture for molecular trajectory prediction that explicitly encodes rigid constraints. Inspired by Graph Mechanics Networks, we design a constraint-preserving equivariant spatiotemporal attention mechanism that captures temporal dependencies while rigorously maintaining both E(3)-equivariance and rigid-body constraints. Additionally, we introduce a constraint-preserving equivariant pooling module that generates future states by performing a learnable weighted aggregation of historical angular velocities, followed by forward kinematics mapping. This ensures that all outputs simultaneously satisfy E(3)-equivariance and strict bond-length conservation. Evaluated on real-world molecular dynamics datasets, STEGMN consistently outperforms strong baselines. On the rMD17 benchmark, it achieves an average ~40% reduction in prediction MSE relative to representative spatiotemporal graph models (ST-GNN, ST-GCN, and ST-EGNN) across eight small-molecule systems, highlighting the critical value of explicit constraint modeling for physically stable trajectory prediction.

Keywords: equivariant graph neural networks; spatiotemporal graph neural networks; graph mechanics networks; molecular trajectory prediction



Academic Editors: Teng Zhou, Zhizhe Lin and Chunjie Cao

Received: 12 April 2026

Revised: 30 April 2026

Accepted: 4 May 2026

Published: 1 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Accurately predicting the 3D motion trajectories of atoms in a molecular system over time—commonly referred to as molecular trajectory prediction—is a central challenge at the intersection of machine learning and the physical sciences. This capability has far-reaching applications: in drug discovery, it enables the simulation of protein–ligand binding pathways and facilitates the screening of high-affinity compounds [1,2]; in materials science, it underpins the analysis of lattice dynamics, phase transitions, and defect evolution [3]; and in reaction modeling, it allows for the tracking of bond-breaking and

bond-forming events, as well as the prediction of energy barriers and intermediate configurations [4]. Classical Molecular Dynamics (CMD) generates atomic trajectories efficiently by numerically integrating Newton's equations of motion with empirical force fields, yet its accuracy is inherently limited by the transferability of the force field parameters; in contrast, Ab Initio Molecular Dynamics (AIMD) computes atomic forces by solving the electronic structure in real time, offering superior accuracy for describing chemical reactions but remaining prohibitively expensive for large-scale or long-timescale simulations. With the increasing integration of machine learning into scientific computing, this task is gradually shifting from computationally expensive first-principles simulations toward an efficient, data-driven modeling paradigm. This paradigm shift is made feasible by the intrinsic graph structure of molecular systems: atoms naturally serve as graph nodes, while chemical bonds or spatial proximity define the edges. The dynamic evolution of these systems is then captured as the spatiotemporal propagation of node states. Message Passing Neural Networks (MPNNs), due to their excellent scalability and structural adaptability, have naturally become the preferred architecture for modeling this task [5–7]. With the increasing demand for modeling dynamic systems, researchers have extended Graph Neural Networks (GNNs) to incorporate temporal dynamics, leading to the development of Spatiotemporal Graph Neural Networks (STGNNs) [8,9]. These models are designed to jointly capture the spatial dependencies encoded in the graph topology and the temporal evolution patterns of the system. However, when applied to physical systems with inherent geometric structure, conventional spatiotemporal graph models encounter significant limitations. Such systems exhibit not only complex spatiotemporal dynamics but also obey strict physical symmetries—such as invariance or equivariance under rotations, translations, and reflections. Equivariant Graph Neural Networks (GNNs) address this by explicitly embedding symmetry constraints into their architecture, ensuring that model predictions transform covariantly with the input under the action of relevant symmetry groups [10–12]. This inductive bias substantially improves both generalization and physical plausibility, particularly in tasks like molecular dynamics simulation and atomic force prediction [5,13]. Despite these advances, existing equivariant spatiotemporal GNNs still face key challenges when modeling dynamic systems subject to rigid geometric constraints:

- Current spatiotemporal GNNs mainly rely on soft constraint mechanisms rather than architectural hard constraints for geometric preservation. In the absence of explicit geometric constraint modeling, conventional networks may occasionally yield trajectories exhibiting structural distortions or deviations from physical laws, which can compromise prediction reliability. Most existing message-passing frameworks primarily depend on soft constraint strategies, such as regularized loss terms or implicit coordinate updates, which provide training flexibility but lack inherent architectural mechanisms for strict geometric preservation [14–16]. While Graph Mechanics Networks (GMN) successfully enforce hard constraints through generalized coordinates and forward kinematics within an E(3)-equivariant framework, their formulation remains inherently static. This omission of explicit temporal dynamics limits their ability to capture time-evolving representations, thereby constraining predictive accuracy for dynamically evolving molecular systems [17].
- Explicit multi-frame spatiotemporal fusion mechanisms that rigorously preserve both geometric constraints and E(3)-equivariance remain largely underexplored. Although several equivariant spatiotemporal GNNs incorporate temporal pooling mechanisms—such as LoCS's roto-translated recurrent aggregation [18], Eq-Motion's [19] trajectory compression followed by equivariant refinement and ES-TAG's [20] equivariant linear pooling over relative coordinates—these approaches perform temporal fusion directly in Cartesian coordinate space. Without explicit

enforcement of rigid geometric constraints via generalized coordinates or forward kinematics, their pooling operations may produce outputs that, while E(3)-equivariant, violate bond-length conservation or induce unphysical structural deformations. This limitation motivates our design of a constraint-preserving equivariant pooling mechanism that operates on angular velocities within a generalized-coordinate framework.

In response to the aforementioned challenges, we propose a Spatiotemporal Equivariant Graph Mechanics Network (STEGMN). The principal contributions of this work are as follows:

- To prevent violations of geometric constraints during spatiotemporal message passing, we introduce the constraint-preserving equivariant Temporal Mechanics Attention Layer (TMAL). TMAL tightly integrates generalized coordinates with temporal dynamics via forward kinematics, enabling the model to capture temporal dependencies while rigorously preserving E(3)-equivariance and rigid-body constraints. Specifically, it employs angular acceleration—a quantity inherently constrained by the system’s kinematic structure—as the attention value for message aggregation. This design intrinsically respects rigid connectivity, effectively mitigating bond-length distortions induced by coordinate drift and significantly enhancing geometric stability in trajectory prediction.
- To enable effective multi-frame spatiotemporal fusion, we propose the Temporal Mechanics Pooling Layer (TMPL). Built upon generalized coordinates, this layer explicitly maintains the system’s symmetries and geometric constraints during temporal aggregation. It performs learnable weighted pooling over historical angular velocities and integrates forward kinematics to map pooled representations to future states. This design achieves temporal compression while rigorously preserving E(3)-equivariance and the topological integrity of the molecular structure throughout the spatiotemporal hierarchy.

The paper is organized as follows: Section 2 reviews prior work and its limitations in preserving physical symmetries and rigid constraints. Section 3 formalizes the trajectory prediction task with spatiotemporal graphs, rigid constraints, and E(3)-equivariance. Section 4 introduces STEGMN—our equivariant architecture with spatial and temporal mechanics modules—and provides equivariance proofs. Section 5 shows STEGMN outperforms baselines on both small (rMD17) and large (AdK) molecular systems. Section 7 concludes by emphasizing STEGMN’s ability to unify physical constraints with spatiotemporal dynamics for modeling complex biomolecules.

2. Related Work

Graph Neural Networks (GNNs) have been widely adopted in molecular dynamics simulation tasks due to their powerful capability in modeling non-Euclidean structured data [21]. Molecules naturally lend themselves to graph representations, where atoms serve as nodes and chemical bonds or interactions act as edges [22–24]. The Message Passing Neural Networks (MPNNs) framework captures complex atomic interactions by iteratively passing and aggregating local information among nodes, enabling accurate predictions of energy, forces, and conformational evolution [25–27]. SchNet [3] employs a distance-based continuous-filter convolution mechanism and achieves high-accuracy predictions of molecular energies and atomic forces on benchmark datasets such as QM9 [28] and MD17 [29]. Subsequently, PhysNet [30] further improves performance in modeling molecular dynamics trajectories by introducing deeper residual architectures and richer feature interactions built upon SchNet. However, these models are not equivariant: their message-passing architectures rely exclusively on rotationally invariant scalar features—typically derived

from interatomic distances—and therefore lack the representational capacity to directly predict vector- or tensor-valued physical quantities (e.g., atomic forces or dipole moments) in an equivariant manner [17]. While physically consistent forces can be recovered a posteriori via the negative gradient of the predicted energy, this equivariance emerges from the mathematical structure of differentiation rather than from the network’s internal equivariant design.

With the growing recognition of equivariance as a fundamental physical prior, equivariant Graph Neural Networks (Equivariant GNNs) have gradually become the dominant paradigm in molecular modeling. Tensor Field Networks [31] were the first to systematically integrate spherical harmonics and Clebsch–Gordan coupling, enabling strictly SE(3)-equivariant message passing. Models based on the Transformer architecture [32–34], exemplified by the SE(3)-Transformer [35], introduce a family of equivariant self-attention mechanisms that combine invariant attention weights with equivariant feature updates, effectively balancing long-range dependency modeling and symmetry preservation, thereby significantly enhancing expressiveness and modeling efficiency. The E(n) Equivariant Graph Neural Network (EGNN) [12] further simplified equivariant design by explicitly constructing E(n)-equivariant coordinate update terms (e.g., learnable scalings based on relative positions) within the message function, achieving strong equivariance while substantially reducing computational complexity—making it more suitable for large-scale molecular dynamics simulations. GMN [17] employs generalized coordinates to impose hard constraints on the model, thereby inherently ensuring rigid-body motion invariance and physical feasibility of molecular conformations. However, most of these equivariant models focus primarily on spatial domain modeling, that is, leveraging equivariance constraints to accurately capture geometric symmetries and local interactions of molecular structures in 3D space, while largely neglecting explicit modeling of temporal dynamics. Since molecular motion is inherently a spatiotemporal coupled process, relying solely on static or instantaneous spatial snapshots is insufficient to recover true dynamical laws.

Consequently, researchers have begun integrating equivariance with temporal modeling to jointly capture both spatial symmetries and temporal evolution of molecular systems. LoCS [18] employs a GRU to maintain memory of past frames and further incorporates rotational invariance to improve generalization. In contrast to LoCS’s recurrent update mechanism, EqMotion [19] condenses each node’s historical trajectory into a multi-dimensional vector, thereby compressing the spatiotemporal graph into a spatial graph, and then designs an equivariant module together with an interaction reasoning module to predict future frames. ESTAG [20] extends equivariance to spatiotemporal GNNs by constructing separate equivariant spatial and temporal modules—each equipped with forward-looking attention and equivariant pooling—to handle spatial and temporal message passing, respectively. Despite their success in modeling symmetries, most of these approaches still treat molecules as systems of freely moving particles, neglecting the rigid geometric constraints imposed by bond lengths. This often leads to unphysical structural deformations in dynamic predictions.

Our method not only inherits equivariant properties but further integrates generalized coordinates and a forward kinematics mechanism to explicitly embed rigid constraints into the spatiotemporal modeling process. By combining historical dynamic information with constraint-preserving equivariant pooling, our model effectively avoids structural distortions commonly caused in conventional approaches by coordinate drift or insufficient implicit constraint learning, thereby enhancing the physical consistency of molecular trajectory predictions while maintaining high prediction accuracy.

3. Problem Description

This paper employs the concept of spatiotemporal graphs to formally represent the dynamic processes of physical systems such as molecules. A spatiotemporal graph with N nodes, E edges and temporal duration T is denoted as $\{G_t = (N_t, E_t)\}_{t=0}^{T-1}$. Throughout this work, the node set N_t and the edge set E_t is time-invariant, with $N_t = N$ and $E_t = E$ for all t ; however, each node P_i is associated with time-dependent scalar features $h_i(t)$ and spatial position vectors $x_i(t)$ at each time frame t . The edge set E corresponds to a fixed adjacency tensor $A \in \mathbb{R}^{N \times N \times m}$, where N denotes the number of nodes and m the number of edge feature dimensions; each element $A_{ij} \in \mathbb{R}^m$ encodes the edge features between node P_i and node P_j . The collective node features and positions at time t are compactly represented by matrices $H(t) \in \mathbb{R}^{N \times c}$ and $X(t) \in \mathbb{R}^{N \times 3}$, where c denotes the dimensionality of the per-node feature vectors. For notational simplicity, we assume a unit time step $\Delta t = 1$. Conventionally, the physical time corresponding to frame t is given by $T - t\Delta t$. The model's objective is, given the system states from time $t = 0$ to $t = T - 1$, to predict the 3D positions of all nodes at time T , represented as $X(t) \in \mathbb{R}^{N \times 3}$, thereby forecasting the future spatial configuration of the system.

In physical systems, a rigid connection refers to a bonding or structural constraint that maintains a fixed geometric relationship between certain atoms or nodes through chemical bonds or other structural constraints. Therefore, our model should satisfy such constraints. Figure 1 illustrates a rigid connection structure O_0 at time t . Here, $x_1(t)$ and $x_2(t)$ denote the Cartesian coordinates of two atoms P_i and P_j in 3D space, respectively. The variable $q_0(t)$ represents the center-of-mass coordinate of O_0 , while $\ddot{\theta}_{01}(t)$ and $\ddot{\theta}_{02}(t)$ denote the generalized angular accelerations of the respective atoms relative to the center of mass, which describe local rotational motion about it. The vectors $x_{01}(t)$ and $x_{02}(t)$ represent the relative position vectors from the center of mass to each atom.

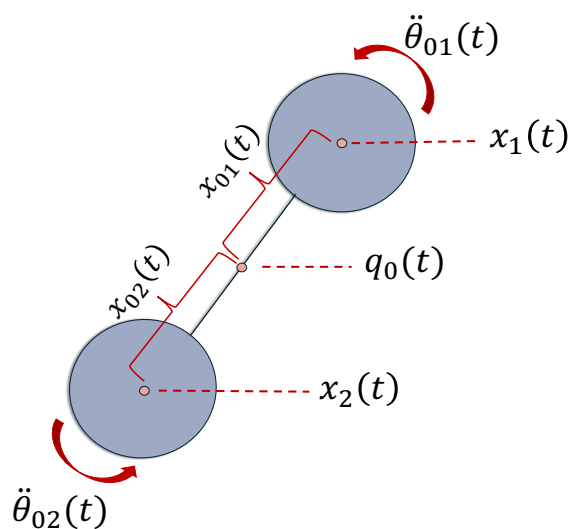


Figure 1. Rigid connection O_0 at time t .

If a rigid connection exists between nodes P_1 and P_2 , their relative distance must remain constant throughout the evolution process, i.e., the following constraint must be satisfied:

$$\|x_1(t) - x_2(t)\| = d_{12}, \quad \forall t \in [0, T - 1], \quad (1)$$

where d_{12} is a constant distance (e.g., covalent bond length) that does not vary with time. As in Graphical Mechanics Networks (GMN), we define a structural object $O_k = \{P_i\}_{i=1}^{n_k}$ as a group of n_k particles connected via rigid bonds (as illustrated in Figure 1). To preserve

the interatomic distances, the dynamics of all particles within O_k should be updated jointly, using angular velocities rather than directly manipulating Cartesian coordinates, ensuring that the rigid constraints are strictly enforced during the simulation.

The model must also satisfy E(3) equivariance: for any Euclidean transformation $g \in E(3)$ (i.e., any combination of rotation, reflection, and translation), we have

$$\mathcal{M}(g \cdot X(t), H(t)) = g \cdot \mathcal{M}(X(t), H(t)), \quad (2)$$

where the action of $g = (R, b)$ on the position matrix $X(t) \in \mathbb{R}^{N \times 3}$ is defined node-wise as $(g \cdot X(t))_i = R x_i(t) + b$ for all $i = 1, \dots, N$, with $R \in O(3)$ and $t \in \mathbb{R}^3$. The scalar node features $H(t)$ are invariant under g (i.e., unchanged by Euclidean transformations).

4. STEGMN

In this section, we present a detailed description of the implementation of STEGMN. The overall architecture of the model is illustrated in Figure 2. We first adopt the original Graph Mechanics Network (GMN) [17] as the spatial modeling module to explicitly encode rigid geometric constraints while ensuring E(3)-equivariance. Subsequently, we introduce a constraint-preserving equivariant spatiotemporal attention layer to capture dynamic dependencies across time steps, enabling effective modeling of temporal evolution. The spatial and temporal modules are executed alternately, progressively fusing spatiotemporal features layer by layer to refine high-level representations while preserving physical consistency. Finally, a constraint-preserving equivariant pooling layer aggregates structural information to generate the system configuration at the next time step.

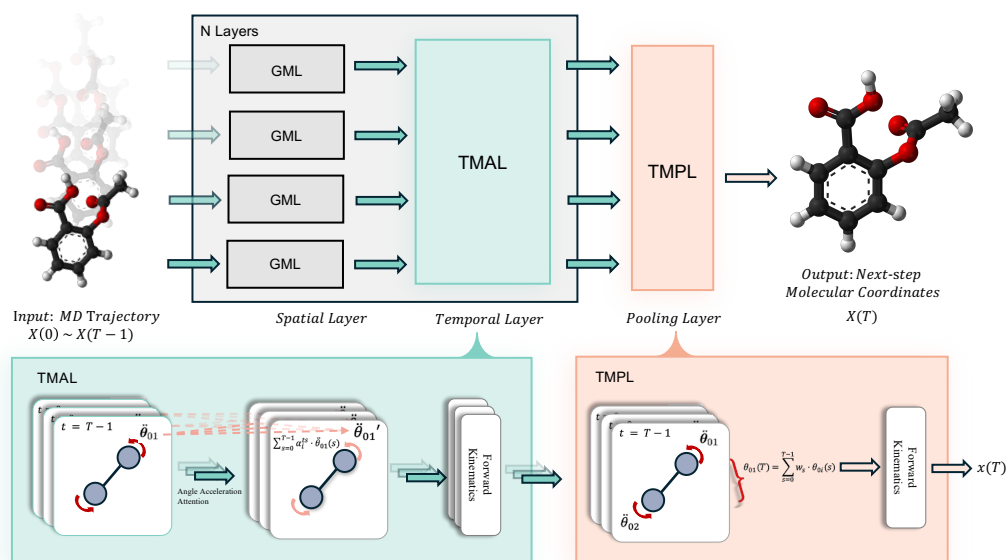


Figure 2. Overall architecture of STEGMN. Spheres represent atomic nodes and dashed lines indicate interatomic interactions or temporal dependencies. The model processes a historical molecular trajectory $\{G_t\}_{t=0}^{T-1}$ with time-invariant topology, where each frame comprises node features $H(t) \in \mathbb{R}^{N \times c}$, 3D Cartesian coordinates $X(t) \in \mathbb{R}^{N \times 3}$, and rigid connectivity priors. Structured as spatiotemporal tensors, these inputs flow through an alternating pipeline of GMN-based spatial modules and Temporal Mechanics Attention Layers (TMALs) to jointly encode local geometry and dynamic dependencies. Finally, the Temporal Mechanics Pooling Layer (TMPL) aggregates historical angular velocities via forward kinematics to predict the molecular configuration at time step T .

4.1. Spatial Modeling: Graph Mechanics Layer Based on GMN

We first adopt the original Graph Mechanics Network (GMN) as the foundational module for spatial modeling. Following the GMN, we define the center of mass of a rigid body O_k as its generalized Cartesian coordinate q_k , and the rotation angles θ_{k1}, θ_{k2} (with $\theta_{k1} = \theta_{k2}$) at both ends of the rigid connection as generalized angular coordinates. As with EGNN, we then compute the interaction force $f_i^l(t)$ acting on each particle i at layer l and time t , relative to all other particles:

$$f_i^l(t), h_i^l(t) = \sum_{j \in N(i)} \varphi_1(x_{ji}^{l-1}(t), h_j^{l-1}(t), h_i^{l-1}(t), e_{ji}(t)) \quad (3)$$

t denotes the current time step; $h_i^{l-1}(t)$ and $h_j^{l-1}(t)$ are the node features of node i and its neighbor j , respectively; $x_{ji}^{l-1}(t)$ is the relative position between nodes i and j at time t , defined as $x_{ji}^{l-1}(t) = x_j^{l-1} - x_i^{l-1}$; $e_{ji}(t)$ represents the edge feature; and φ_1 is a learnable equivariant function. To aggregate information from all particles within each structural object O_k , we compute the generalized Cartesian acceleration $\ddot{q}_k^l(t)$ via another learnable equivariant function φ_2 :

$$\ddot{q}_k^l(t) = \sum_{i \in O_k} \varphi_2(f_i^l(t), x_{ki}^{l-1}(t), v_{ki}^{l-1}(t)) \quad (4)$$

Through the above steps, we obtain the generalized Cartesian acceleration—i.e., the Cartesian acceleration of the center of mass of the rigid structure. To enable forward kinematics, we further compute the angular acceleration of each endpoint relative to the center of mass. Based on the interaction forces and Cartesian coordinates acting on each endpoint, we can derive the generalized angular accelerations $\ddot{\theta}_{01}$ and $\ddot{\theta}_{02}$ for the two particles:

$$\ddot{\theta}_{ki}(t) = \frac{\sum_{i \in O_k} x_{ki}(t) \times f_i(t)}{\sum_{i \in O_k} \|x_{ki}(t)\|^2}, \quad \forall i \in O_k \quad (5)$$

Given the generalized Cartesian acceleration and the angular accelerations of the two particles, we can iteratively compute the generalized Cartesian velocity $\dot{q}_k^l(t)$ and coordinate $q_k^l(t)$ via forward kinematics:

$$\dot{q}_k^l(t) = \psi_1 \left(\sum_{i \in O_k} h_i^{l-1}(t) \right) \dot{q}_k^{l-1}(t) + \ddot{q}_k^l(t) \quad (6)$$

$$q_k^l(t) = q_k^{l-1}(t) + \dot{q}_k^l(t) \quad (7)$$

Here, $\dot{q}_k^{l-1}(t)$ denotes the current velocity of the center of mass of the rigid structure, which can be computed from existing conditions. The function ψ_1 acts as an adaptive weighting mechanism, enabling self-tuned control over velocity updates. Subsequently, we update the angular velocities $\dot{\theta}_{ki}^l(t)$ of the endpoints using forward kinematics:

$$\dot{\theta}_{ki}^l(t) = \psi_2 \left(\sum_{i \in O_k} h_i^{l-1}(t) \right) \dot{\theta}_{ki}^{l-1}(t) + \ddot{\theta}_{ki}^l(t), \quad \forall i \in O_k. \quad (8)$$

Similarly, ψ_2 serves as an adaptive weight. Finally, we apply forward kinematics to each particle in O_k to obtain its position and velocity at time t , where rot denotes the rotation matrix. Since updates of atoms within rigid connections are driven by generalized angular velocities, the hard constraint $d_{ij} = \text{const}$ is inherently satisfied.

$$x_i^l(t) = q_k^l(t) + \text{rot}(\dot{\theta}_{ki}^l(t))x_{ki}^{l-1}(t), \quad \forall i \in O_k \quad (9)$$

$$v_i^l(t) = \dot{q}_k^l(t) + \dot{\theta}_{ki}^l(t) \times x_{ki}^l(t), \quad \forall i \in O_k \quad (10)$$

It is worth noting that, in this layer design, we employ GMN solely as an equivariant feature extractor in the spatial domain, without introducing cross-temporal state propagation. Therefore, for each time frame $t \in [0, T - 1]$, the spatial update of the graph structure G_t is performed independently:

$$X^l(t), V^l(t) = \text{GMNLayer}(X^{l-1}(t), V^{l-1}(t), H^{l-1}(t)) \quad (11)$$

4.2. Temporal Modeling: Temporal Mechanics Attention Layer

To capture temporal dynamics while preserving both E(3)-equivariance and rigid geometric constraints, we introduce a constrain-preserving equivariant attention mechanism. Following the standard attention mechanism, each node's temporal state is associated with query and key representations to compute time-aware attention weights. Specifically, for node P_i , we treat its node feature at the current time step t as the query, and its node feature at historical time steps $s \in [0, T - 1]$ as the keys:

$$q_i^t = W_q h_i(t), \quad k_i^s = W_k h_i(s), \quad (12)$$

where W_q and W_k are learnable projection matrices. The attention score is computed as

$$\alpha_i^{ts} = \frac{1}{Z_i} \exp\left(\frac{q_i^t \cdot k_i^s}{\sqrt{d}}\right) \quad (13)$$

$Z_i = \sum_{s=0}^{T-1} \exp(\dots)$ is the normalization factor. The parameter d denotes the feature dimensionality of projected queries and keys ($q_i^t, k_i^s \in \mathbb{R}^d$). Consistent with the scaled dot-product attention in Transformers [36], the scaling factor $1/\sqrt{d}$ normalizes the dot-product variance to unity, preventing softmax saturation and ensuring stable gradient flow for temporal learning.

Unlike conventional equivariant temporal attention mechanisms, we define the value as the angular acceleration of node P_i at time s to ensure that hard constraints are preserved during temporal computation. Since angular acceleration reflects the rate of rotational change around local axes—and in rigid structures, motion is constrained by bond lengths—using it as the attention output naturally regularizes node dynamics, avoiding structural distortions caused by feature drift or overfitting. Specifically, we first compute the angular accelerations for all nodes using Equation (5), then apply the attention mechanism to compute the angular acceleration increment $\Delta\ddot{\theta}_{ki}(t)$ from each historical time step s to the current time t :

$$\Delta\ddot{\theta}_{ki}(t) = \sum_{s=0}^{T-1} \alpha_i^{ts} \cdot \ddot{\theta}_{ki}(s) \quad (14)$$

We then update the angular acceleration:

$$\ddot{\theta}_{ki}(t)_{\text{temporal}} = \ddot{\theta}_{ki}(t) + \Delta\ddot{\theta}_{ki}(t) \quad (15)$$

Up to this point, by using node angular accelerations as the value vectors in the attention mechanism, we successfully integrate temporal dynamic information into the model. As a measure of rotational acceleration, angular acceleration reflects the force and inertial

response of nodes within their local reference frames; its weighted aggregation within the attention mechanism is equivalent to physically informed perception and prediction of historical motion trends.

Therefore, STEGMN not only fuses high-order temporal information but also preserves structural rigidity and E(3)-equivariance throughout evolution. We can stack spatial and temporal layers, iteratively performing GMN-based spatial message passing alongside constrain-preserving equivariant temporal mechanics attention layer, thereby achieving hierarchical fusion of spatiotemporal features. Moreover, since each layer's update operates within the framework of generalized coordinates and forward kinematics, the entire process inherently maintains E(3)-equivariance and rigid constraints, avoiding potential structural distortions or symmetry breaking during deep propagation. (The proof of equivariance is provided in Appendix A.1).

4.3. Temporal Mechanics Pooling Layer

After extracting multi-layer spatiotemporal features, to further aggregate global dynamic information and predict the system configuration at time T , we design a constrain-preserving equivariant pooling layer. This layer introduces a set of node-level learnable weights $\{W_s\}_{s=1}^{T-1}$, directly applied to the angular velocities of node P_i across time steps 0 to $T - 1$, enabling lightweight modeling of overall motion trends. According to rigid-body rotational dynamics, if two endpoints rotate about their midpoint, the angular velocity can be derived from the cross product of relative velocity and relative position:

$$\dot{\theta}_{ki}(s) = \dot{\theta}_{kj}(s) = \frac{(x_i(s) - x_j(s)) \times (v_i(s) - v_j(s))}{\|x_i(s) - x_j(s)\|^2} \quad (16)$$

Here, P_i and P_j denote the two endpoints of a rigid connection, and k represents the rigid structure containing P_i and P_j . After obtaining the angular velocities for each rigid unit over time steps $s = 0$ to $T - 1$, we introduce a set of global learnable temporal weights $\{w_s\}_{s=0}^{T-1}$, satisfying $\sum_{s=0}^{T-1} w_s = 1$, to predict the angular velocity at time T :

$$\dot{\theta}_{ki}(T) = \sum_{s=0}^{T-1} w_s \cdot \dot{\theta}_{ki}(s) \quad (17)$$

Once the angular velocity at time T is computed, we derive the final position and velocity at time T based on the state at time $T - 1$, using forward kinematics:

$$x_i(T) = q_k(T - 1) + \text{rot}(\dot{\theta}_{ki}(T))x_{ki}(T - 1) \quad (18)$$

While maintaining lightweight computation, our method strictly preserves E(3)-equivariance: the weighting scheme is equivariant under rotation, translation and reflection. Combined with the forward kinematic mapping based on generalized coordinates, the predicted configuration naturally satisfies rigid geometric constraints. (The proof of equivariance is provided in Appendix A.2).

5. Results

5.1. Baseline

We compare STEGMN against several baseline models to systematically evaluate the importance of equivariance, temporal modeling, and explicit geometric constraints. The selected baselines include equivariant models without temporal dynamics (EGNN [12]), non-equivariant spatiotemporal models (e.g., ST-GCN [8], ST-GNN [37], AGL-STAN [38]), and equivariant spatiotemporal models that lack explicit rigid constraints (ST-EGNN [12]).

- EGNN [12]: We select EGNN as a baseline due to its status as a representative equivariant graph neural network. By explicitly constructing $E(n)$ -equivariant coordinate updates within its message-passing scheme, EGNN demonstrates strong performance and generalization in molecular energy and force prediction tasks. However, EGNN is primarily designed for static graphs or single-frame conformations and does not explicitly model temporal dynamics. In our experiments, we adopt a single-frame prediction protocol.
- ST-GNN [37]: ST-GNN is a general spatiotemporal GNN framework that jointly models node state sequences by stacking recurrent or convolutional temporal modules atop spatial graph convolutions. It performs message passing directly in Cartesian coordinate space, lacking both $E(3)$ -equivariance and explicit enforcement of rigid geometric constraints.
- ST-GCN [8]: As a classic non-equivariant spatiotemporal graph model, ST-GCN has been widely applied to tasks such as traffic flow forecasting. Its “temporal-spatial-temporal” architecture effectively captures spatiotemporal dependencies. We adapt ST-GCN to the molecular dynamics prediction task to establish a performance upper bound for purely data-driven approaches that lack equivariance constraints.
- AGL-STAN [38]: AGL-STAN is an advanced GNN framework for spatiotemporal sequence prediction. It dynamically infers latent spatial dependencies via an adaptive graph learning mechanism and integrates spatiotemporal attention modules to capture complex temporal patterns and spatial interactions. Originally excelling in traffic forecasting under non-fixed topologies, we transfer AGL-STAN to molecular trajectory prediction by taking multi-frame historical conformations as input and leveraging its adaptive graph structure and attention mechanism to model dynamic atomic interactions and predict coordinate residuals at the next time step.
- ST-EGNN [12]: ST-EGNN is a natural extension of EGNN to spatiotemporal modeling, inheriting EGNN’s strengths in incorporating physical priors. It represents the current paradigm of equivariant spatiotemporal modeling. Nevertheless, ST-EGNN still operates in Cartesian coordinates and does not explicitly enforce rigid geometric constraints, relying instead on implicit learning of bond-length conservation from data.

5.2. Dataset

To evaluate the effectiveness of the proposed method in modeling dynamics of physical systems, we conduct experiments on two representative molecular dynamics datasets: the rMD17 small-molecule dataset [39] and the AdK equilibrium trajectory dataset [29], both containing multiple continuous long trajectories. The original datasets consist of long trajectory sequences. We randomize the starting point and extract $T + 1$ consecutive frames at interval ΔT , using the first T time steps as observed conformations and the final timestamp as the target future position label.

- The revised MD17 dataset (rMD17) [39] is a refined version of the MD17 dataset, comprising long-timescale trajectories of eight small organic molecules (e.g., Aspirin, Benzene, Ethanol) generated via ab initio molecular dynamics simulations. Each trajectory is sampled at a 2 fs interval and spans several hundred picoseconds, capturing rich vibrational and conformational dynamics including bond stretching, angle bending, and torsional degrees of freedom. The dataset provides, at an atomic resolution, the 3D coordinates, atomic forces, and total energy for each frame, making it well-suited for evaluating a model’s ability to capture local vibrations and conformational transitions. In our experiments, we sample the original trajectories using contiguous time windows: a random starting point is selected from the long trajectory, and $T + 1$ consecutive conformations are extracted at a fixed time interval ΔT . The first T frames

serve as the input historical states, and the model predicts the 3D coordinates of all atoms in the $(T + 1)$ -th frame. For each molecule, the trajectory is split into training, validation, and test sets containing 500, 2000, and 2000 samples, respectively (see Table 1).

- The AdK equilibrium dataset [29] originates from nanosecond-scale molecular dynamics simulations of adenylate kinase (AdK) in its equilibrium state. AdK is a canonical allosteric protein whose function relies on large-scale hinge-like motions between two domains, exhibiting multiscale conformational changes. The dataset provides a continuous 1-microsecond trajectory sampled at 1 ps intervals. Due to the high computational cost of modeling thousands of atoms directly at atomic resolution, we apply a coarse-graining procedure: each amino acid residue is treated as a single node, and the 3D coordinates of four backbone atoms (N, C_α , C, O) are extracted as a four-channel geometric feature for that node. This coarse-graining is performed using the MDAnalysis toolkit (v2.9.0), which significantly reduces computational complexity while preserving the essential conformational dynamics.

Table 1. Dataset composition. For the AdK equilibrium dataset, the numbers of atoms and edges in parentheses correspond to the original values before coarse-graining.

Dataset	Total	Training	Validation	Test	Atoms	Edges
rMD17 [39]	4500	500	2000	2000	9–20	8–25
AdK Equilibrium [29]	1068	641	214	213	214 (3341)	852 (3365)

5.3. Small-Molecule Dataset: rMD17

We use atomic numbers as time-invariant node features $h^{(0)}$. Two atoms are considered one-hop neighbors if their distance is below a pre-defined threshold λ . All models are optimized using Adam with an initial learning rate of 5×10^{-5} , coupled with a learning rate scheduler. Each model is trained for 500 epochs under identical hardware and hyperparameter settings, and performance is evaluated on an independent test set using MSE. The selection of edges for rigid connections follows the same criterion as in Graphical Mechanics Networks (GMN). Since including all bonded edges would lead to redundant updates, we adopt a strategy that maximizes the number of selected edges while ensuring they are non-overlapping. All experiments were conducted on a single NVIDIA A100 GPU (80 GB).

Table 2 presents the MSE and MAE of all models across the eight molecules. Notably, the baseline EGNN—relying solely on single-frame input without modeling temporal state sequences—fails to capture the temporal evolution patterns in molecular dynamics, resulting in significantly lower accuracy compared to spatiotemporal models that leverage multi-frame sequences. STEGMN achieves the best performance on six molecules and substantially outperforms the runner-up on several, demonstrating that jointly preserving equivariance and rigid geometric constraints within spatiotemporal modeling effectively enhances trajectory prediction accuracy. On the remaining two molecules (ethanol and naphthalene), STEGMN delivers highly competitive results—slightly higher than the best model but still significantly superior to all other baselines.

Table 2. Molecular trajectory prediction results on the rMD17 dataset (MSE in 10^{-3} , MAE in Å, bold values indicate the best result for each metric.). Molecule abbreviations: Asp = Aspirin, Bzn = Benzene, Eth = Ethanol, Mal = Malonaldehyde, Nap = Naphthalene, Sal = Salicylic acid, Tol = Toluene, Ura = Uracil.

Method	Metric	Asp	Bzn	Eth	Mal	Nap	Sal	Tol	Ura
EGNN [12]	MSE	5.937	1.885	1.531	4.850	3.333	4.826	2.276	4.121
	MAE	0.059	0.033	0.030	0.054	0.045	0.051	0.038	0.044
AGL-STAN [38]	MSE	3.681	1.649	1.389	1.089	0.900	1.098	1.506	0.828
	MAE	0.049	0.031	0.026	0.026	0.023	0.026	0.030	0.022
ST-GNN [37]	MSE	1.014	0.210	0.487	0.664	0.748	0.789	0.713	0.680
	MAE	0.024	0.012	0.018	0.020	0.021	0.020	0.021	0.020
ST-GCN [8]	MSE	0.715	0.106	0.456	0.596	0.737	0.682	0.687	0.673
	MAE	0.021	0.009	0.017	0.019	0.022	0.021	0.021	0.021
ST-EGNN [12]	MSE	0.735	0.163	0.245	0.427	0.359	0.687	0.553	0.445
	MAE	0.021	0.013	0.012	0.016	0.019	0.019	0.016	0.019
STEGMN (Ours)	MSE	0.393	0.084	0.331	0.264	0.398	0.365	0.338	0.251
	MAE	0.014	0.007	0.013	0.011	0.016	0.014	0.014	0.011

5.4. Protein Dataset: AdK Equilibrium Dataset

We use the atomic numbers of the four backbone atoms as time-invariant input node features. Two atoms are connected by an edge if their distance is below a threshold λ . We evaluate our method on the AdK equilibrium molecular dynamics trajectory dataset, employing the MDAnalysis toolkit to coarse-grain the protein structure into a residue-level graph (one C_α node per residue), thereby simplifying its spatial representation. We also adopt a strategy that maximizes the number of selected edges while ensuring they are non-overlapping. We use $T = 10$ past frames with stride $\Delta = 5$ to predict the next frame. We train for 500 epochs using Adam with a learning rate of 5×10^{-5} , weight decay of 10^{-12} , batch size of 4 (limited by GPU memory due to the large protein system size), and report results every five epochs. All experiments were conducted on a single NVIDIA A100 GPU (80 GB).

Table 3 presents the prediction performance of all models on the AdK equilibrium conformation dataset. It is evident that all equivariant models (STEGMN, ST-EGNN) significantly outperform non-equivariant baselines (e.g., AGL-STAN, ST-GCN), underscoring the critical role of E(3)-equivariance in modeling conformational dynamics for biomolecular systems—such as proteins—that exhibit complex spatial symmetries. By enforcing outputs to transform covariantly with input rotations and translations, equivariant models learn more fundamental physical laws, reducing reliance on data augmentation and thereby improving generalization. Among all models, STEGMN achieves the lowest prediction error, surpassing other equivariant counterparts, indicating that explicitly incorporating hard constraints further enhances trajectory prediction accuracy once equivariance is fully satisfied.

Figures 3 and 4 compare the residue-level prediction errors and their distributions between STEGMN and ST-EGNN. The error distribution of STEGMN peaks near 0.5 Å (density ≈ 0.55), indicating that most predictions are highly accurate. In contrast, ST-EGNN exhibits a broader distribution, with substantial density persisting in high-error regions (4–8 Å), reflecting greater prediction variability. Overall, STEGMN demonstrates significantly improved accuracy over ST-EGNN, with a more concentrated error distribution and lower mean error magnitude.

Table 3. Molecular trajectory prediction results on the AdK dataset. The “MSE ↓%” and “MAE ↓%” columns report the percentage reduction relative to EGNN (higher values indicate greater improvement). Bold values indicate the best result for each metric.

Model	MSE	MSE ↓%	MAE	MAE ↓%
EGNN [12]	2.837	—	1.018	—
AGL-STAN [38]	1.965	30.7	1.100	−8.1
ST-GNN [37]	2.774	2.2	1.196	−17.5
ST-GCN [8]	1.647	41.9	0.791	22.3
ST-EGNN [12]	1.598	43.7	0.770	24.4
STEGMN (Ours)	1.582	44.2	0.767	24.7

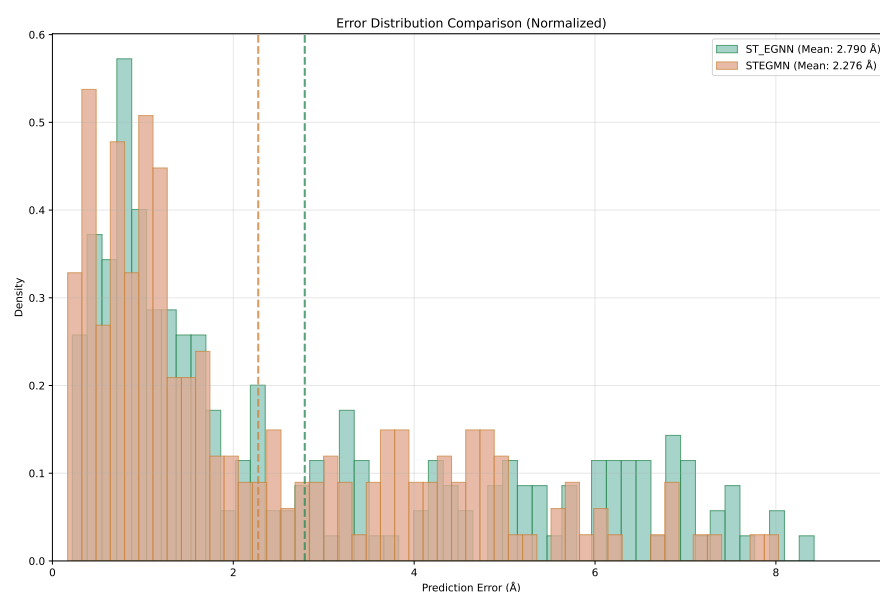


Figure 3. Comparison of error distributions between ST-EGNN and STEGMN (normalized). Dotted lines indicate the mean prediction errors for each model.

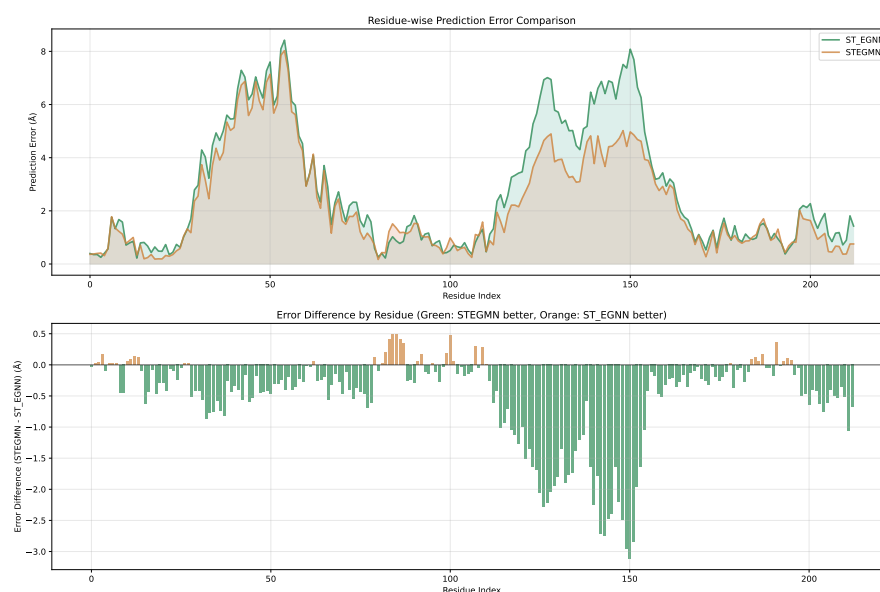


Figure 4. Comparative analysis of residue-level prediction errors between ST-EGNN and STEGMN. The top panel displays the distribution of prediction errors at each residue position (i.e., C_{α} atom positions) in the protein sequence for both the ST-EGNN and STEGMN models. The bottom panel shows the difference in errors at each residue position between the two models.

5.5. Ablation Study

To validate the contribution of the two core components proposed in this paper—the Temporal Mechanics Attention Layer (TMAL) and Temporal Mechanics Pooling Layer (TMPL)—to the overall model performance, we conducted a systematic ablation study. Specifically, on the rMD17 dataset, we evaluated the full STEGMN model by removing each of these two key components in turn.

- STEGMN -w/o tmal: This variant retains the full spatiotemporal architecture but replaces TMAL proposed in Section 4.2 with a standard, unconstrained equivariant temporal attention module. In this variant, queries and keys are still computed from node features at each time step. However, the values no longer employ angular accelerations governed by rigid constraints; instead, they adopt the equivariant coordinate update strategy from ST-EGNN: specifically, the relative position vectors (i.e., radial vectors) from the historical time steps 0 to $T - 1$ to the current time step $T - 1$ are used as attention values. These vectors are scaled by a learnable scalar function and directly added to the node coordinates for feature aggregation. While this variant mathematically preserves E(3)-equivariance, it fundamentally performs local coordinate offsets in Cartesian space, completely ignoring the rigidity constraints imposed by the generalized coordinate framework.
- STEGMN -w/o tmpl: This variant removes the TMPL introduced in Section 4.3. In this setting, the model no longer generates future conformations by learning to weight and aggregate historical angular velocities. Instead, we adopt a simple, Cartesian-residual-based linear weighting update strategy:

$$x(T) = \sum_i W_i \cdot (x(T-1) - x(i)) + x(T-1), \quad (19)$$

where W is a learnable weight vector, $x(T-1)$ denotes the coordinates of the most recent frame, and $x(i)$ represents the coordinates of historical frames $i = 0, \dots, T-1$. This operation is performed directly in Cartesian space—preserving E(3)-equivariance but completely disregarding the rigidity constraints imposed by the generalized coordinate framework—and attempts to fit coordinate trends in a purely data-driven manner.

- STEGMN full: The full version of STEGMN, incorporating both the Temporal Mechanics Attention Layer (TMAL) and the Temporal Mechanics Pooling Layer (TMPL).

Table 4 presents the validation errors of all variants on the rMD17 dataset. The results indicate that ablating either the TMAL or TMPL module consistently degrades prediction accuracy across all tested molecules. This confirms that the rigid geometric constraints enforced by both components are indispensable for maintaining physical consistency and ensuring robust spatiotemporal modeling.

Table 4. Ablation study of STEGMN on the rMD17 dataset (MSE in 10^{-3} , MAE in Å). Bold values indicate the best result for each metric.

Method	Metric	Asp	Bzn	Eth	Mal	Nap	Sal	Tol	Ura
STEGMN—w/o tmal	MSE	0.606	0.095	0.412	0.486	0.509	0.496	0.563	0.251
	MAE	0.019	0.008	0.015	0.017	0.018	0.017	0.019	0.011
STEGMN—w/o tmpl	MSE	0.588	0.106	0.383	0.378	0.684	0.567	0.592	0.441
	MAE	0.018	0.008	0.014	0.014	0.020	0.018	0.018	0.015
STEGMN full	MSE	0.372	0.086	0.309	0.258	0.403	0.368	0.348	0.249
	MAE	0.014	0.007	0.013	0.011	0.015	0.014	0.014	0.010

6. Discussion

The experimental results demonstrate that explicitly encoding rigid geometric constraints within an equivariant spatiotemporal architecture substantially improves trajectory prediction fidelity for molecular systems. Compared to baseline equivariant models that operate solely in Cartesian space (e.g., ST-EGNN), STEGMN achieves an average 40% reduction in prediction MSE relative to representative spatiotemporal graph models (ST-GNN, ST-GCN, and ST-EGNN) across eight small-molecule systems on the rMD17 benchmark, underscoring the value of generalized-coordinate formulations for preserving physical consistency during dynamic evolution.

From a broader perspective, STEGMN bridges two previously disjoint design paradigms: (1) equivariant GNNs that enforce symmetry priors but treat atoms as unconstrained point masses, and (2) mechanics-based networks that impose hard geometric constraints but lack explicit temporal modeling. By unifying these principles through angular-velocity-based attention and forward-kinematics pooling, our framework offers a template for incorporating domain-specific physical laws into deep spatiotemporal learners.

- Practical implementation considerations. Translating STEGMN into real-world molecular engineering workflows is facilitated by several inherent architectural strengths. First, the explicit enforcement of rigid geometric constraints and $E(3)$ -equivariance ensures physically plausible outputs without relying on extensive data augmentation or post hoc filtering, thereby reducing training overhead and inference instability. From a data preparation standpoint, the framework only requires standard molecular trajectory data with time-invariant topology. Node features can be as simple as atomic numbers or element embeddings, which are natively available in common file formats (e.g., PDB, SDF, XYZ). The primary preprocessing step involves constructing a rigid connectivity map based on chemical priors or force-field bond definitions, which can be automatically extracted using standard cheminformatics toolkits. While this design significantly streamlines practical application, deployment in large-scale or highly disordered systems may still face challenges related to computational overhead during long-horizon predictions, as well as the need for precise bond-topology annotation when handling flexible or reactive molecular motifs. Overall, STEGMN minimizes implementation complexity by hardcoding physical priors directly into the network, requiring only routine trajectory formatting to deliver reliable, physics-consistent predictions.
- Limitations and future directions. Current limitations include: (1) The explicit forward kinematics module, while guaranteeing strict geometric consistency, involves computationally intensive operations such as rotation matrix construction and coordinate transformation. Compared to direct element-wise updates in Cartesian space, this kinematic mapping incurs higher computational overhead during inference, resulting in increased latency. (2) Reliance on pre-defined rigid connectivity, which may not hold for highly flexible molecular motifs or reactive bond-breaking events, and limited scalability for large-scale molecular systems or extended prediction horizons due to the quadratic complexity of dense spatiotemporal attention, which increases memory and computational demands for systems with thousands of atoms or long trajectory sequences. Future work will explore: (i) adaptive constraint mechanisms for breakable bonds or angle-flexible structures to model chemical reactions; (ii) integration with generative frameworks (e.g., diffusion or flow-based models) for efficient sampling of physically plausible conformational spaces; and (iii) lightweight design strategies, such as adaptive sparse attention mechanisms and dynamic connectivity pruning [40], to selectively aggregate informative spatiotemporal dependencies and

substantially reduce computational overhead for large-scale molecular systems or long-horizon predictions.

7. Conclusions and Future Work

This work addresses the challenge that existing spatiotemporal graph neural networks struggle to explicitly preserve rigid geometric constraints in molecular dynamics trajectory prediction. We propose the Spatio-temporal Equivariant Graph Mechanics Network (STEGMN), which deeply integrates physical priors—namely, rigidity constraints and E(3)-equivariance—into the architecture’s design, rather than relying on implicit learning from data. Specifically, we introduce the Temporal Mechanics Attention Layer (TMAL), which uses rigidity-constrained angular accelerations as the values in attention-based information aggregation. This ensures that while capturing dynamic dependencies across time steps, interatomic geometric constraints are inherently maintained. Furthermore, we propose the Temporal Mechanics Pooling Layer (TMPL), which generates future states by learning to weight and aggregate historical angular velocities and applying forward kinematics. This temporal compression mechanism strictly preserves both E(3)-equivariance and the topological integrity of molecular structures.

Evaluated on two real-world datasets—small molecules (rMD17) and a protein system (AdK equilibrium)—STEGMN consistently outperforms strong equivariant baselines. Ablation studies further confirm that both TMAL and TMPL significantly contribute to overall performance. In future work, STEGMN can be extended to chemical reaction path prediction by incorporating dynamic constraint mechanisms, such as breakable bonds or flexible angle constraints, to effectively model more complex physical scenarios. Additionally, integrating this architecture with generative models like diffusion or flow-based models could enable efficient sampling of physically plausible molecular conformational spaces, accelerating drug discovery and materials design pipelines. Furthermore, we will explore lightweight design strategies, such as adaptive sparse attention mechanisms and dynamic connectivity pruning, to selectively aggregate informative spatiotemporal dependencies and substantially reduce computational overhead for large-scale molecular systems or long-horizon predictions.

Author Contributions: Original draft, Y.M.; data curation, Y.M.; software, Y.M.; implementation, Y.M.; investigation, Y.M.; formal analysis, Y.M.; visualization, Y.M.; conceptualization, Q.Q.; methodology, Q.Q.; funding acquisition, Q.Q.; project administration, Q.Q.; supervision, Q.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This work was sponsored by Advanced Materials-National Science and Technology Major Project (No. 2025ZD0620100), National Key Research and Development Program of China (No. 2023YFB4606200), Open Research Fund Yunnan Tin & Indium Laboratory (No. 202405AR340002-25EA02), Key Program of Science and Technology of Yunnan Province (No. 202302AB080020). The authors gratefully appreciate the anonymous reviewers for their valuable comments.

Data Availability Statement: The data and source code that support the findings are available at <https://github.com/miao2y/STEGMN> (accessed on 11 April 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Proof of Equivariance for STEGMN

Appendix A.1. TMAL

Since the equivariance of the spatial-domain model GMNLayer has been thoroughly established in GMN [17], this chapter focuses on proving the equivariance of the temporal module. Notably, the temporal module is involved only in the update of angular velocity. As a physical quantity characterizing the local rigid-body rotational state, angular velocity

is inherently translation-invariant; thus, its behavior under spatial translations need not be considered. Consequently, it suffices to verify that the angular velocity produced by the temporal module transforms equivariantly under rigid-body rotations—including both proper rotations and reflections—i.e., that it exhibits the correct transformation behavior under the action of the relevant symmetry group. That is, we need to prove that when the input node features and position vectors undergo an $E(3)$ transformation (i.e., a rotation $R \in O(3)$ and a translation $b \in \mathbb{R}^3$), the output of the constrained equivariant attention mechanism also transforms accordingly under the same equivariant rule. For angular velocity updates, this manifests as rotational and reflectional equivariance.

Consider an $E(3)$ transformation $g \in E(3)$, acting on the state of any particle i at any time step s , given by Equations (A1) and (A2):

$$g \cdot x_i(s) = Rx_i(s) + b \quad (\text{A1})$$

$$g \cdot v_i(s) = Rv_i(s) \quad (\text{A2})$$

Since the node feature $h_i(s)$ is designed as a scalar, it remains invariant under $E(3)$ transformations, i.e., $g \cdot h_i(s) = h_i(s)$. The computation of queries and keys depends solely on node features $h_i(t)$ and $h_i(s)$; therefore, the attention score α_i^{ts} remains unchanged under the action of g .

Regarding the equivariance of the value, according to Equation (5), the angular acceleration is computed as follows:

$$\ddot{\theta}_{ki}(s) = \frac{\sum_{i \in \mathcal{O}_k} x_{ki}(s) \times f_i(s)}{\sum_{i \in \mathcal{O}_k} \|x_{ki}(s)\|^2} \quad (\text{A3})$$

where $x_{ki}(s)$ denotes the Cartesian coordinate difference between the center of rigid structure $\mathcal{O}_k$ and node i at time s , i.e., the radial vector, and $f_i(s)$ represents the force acting on node i . Under transformation, the radial vector $x_{ki}(s)$ becomes

$$\begin{aligned} x_{ki}(s)' &= Rx_k(s) + b - (Rx_i(s) + b) \\ &= Rx_k(s) - Rx_i(s) \\ &= R x_{ki}(s) \end{aligned} \quad (\text{A4})$$

Consequently, the angular acceleration transforms equivariantly under the action of g , specifically as follows:

$$\begin{aligned} \Delta \ddot{\theta}_{ki}(t)' &= \sum_{s=0}^{T-1} \alpha_i^{ts} \cdot R \ddot{\theta}_{ki}(s) \\ &= R \sum_{s=0}^{T-1} \alpha_i^{ts} \cdot \ddot{\theta}_{ki}(s) \\ &= R \Delta \ddot{\theta}_{ki}(t). \end{aligned} \quad (\text{A5})$$

$$\begin{aligned} \ddot{\theta}_{ki}(t)'_{\text{temporal}} &= R \ddot{\theta}_{ki}(t)' + R \Delta \ddot{\theta}_{ki}(t) \\ &= R \ddot{\theta}_{ki}(t)_{\text{temporal}} \end{aligned} \quad (\text{A6})$$

Therefore, under $E(3)$ transformations, the angular update exhibits rotational and reflectional equivariance. When further applied to coordinates, it inherits full $E(3)$ -equivariance.

Appendix A.2. TMPL

Similarly, we need to prove that the angular velocity update is rotation and reflectional equivariant under $E(3)$ transformations. According to Equation (16), the angular velocity between two endpoints is computed as

$$\dot{\theta}_{ki}(s) = \dot{\theta}_{kj}(s) = \frac{(x_i(s) - x_j(s)) \times (v_i(s) - v_j(s))}{\|x_i(s) - x_j(s)\|^2}. \quad (\text{A7})$$

Under an $E(3)$ transformation, the transformed angular velocity becomes

$$\begin{aligned} \dot{\theta}_{ki}(s)' &= \dot{\theta}_{kj}(s)' = \frac{(Rx_i(s) + b - (Rx_j(s) + b)) \times (Rv_i(s) - Rv_j(s))}{\|Rx_i(s) + b - (Rx_j(s) + b)\|^2} \\ &= \frac{Rx_{ij}(s) \times R(v_i(s) - v_j(s))}{\|x_{ij}(s)\|^2} \\ &= R \dot{\theta}_{ki}(s) = R \dot{\theta}_{kj}(s). \end{aligned} \quad (\text{A8})$$

To compute the angular velocity at time T , we have

$$\begin{aligned} \dot{\theta}_{ki}(T)' &= \sum_{s=0}^{T-1} w_s \cdot R \dot{\theta}_{ki}(s) \\ &= R \sum_{s=0}^{T-1} w_s \cdot \dot{\theta}_{ki}(s) \\ &= R \dot{\theta}_{ki}(T). \end{aligned} \quad (\text{A9})$$

It follows that the angular velocity at time T is equivariant under $E(3)$ transformations, exhibiting both rotational and reflectional symmetry. Subsequently, we compute the Cartesian coordinates at time T via forward dynamics. First, we define the generalized Cartesian coordinate (the midpoint of a rigid connection) at time $T - 1$ as

$$q_k(T - 1) = \frac{x_i(T - 1) + x_j(T - 1)}{2}. \quad (\text{A10})$$

Under an $E(3)$ transformation, the generalized Cartesian coordinate transforms as

$$\begin{aligned} q_k(T - 1)' &= \frac{Rx_i(T - 1) + b + Rx_j(T - 1) + b}{2} \\ &= \frac{Rx_i(T - 1) + Rx_j(T - 1)}{2} + b \\ &= R q_k(T - 1) + b. \end{aligned} \quad (\text{A11})$$

Simultaneously, leveraging the property of rotation matrices, $\text{rot}(R\dot{\theta}) = R \cdot \text{rot}(\dot{\theta}) \cdot R^\top$, the coordinate update formula $x_i(T) = q_k(T - 1) + \text{rot}(\dot{\theta}_{ki}(T - 1)) \cdot x_{ki}(T - 1)$ transforms under an $E(3)$ transformation as follows:

$$\begin{aligned} x_i(T)' &= R q_k(T - 1) + b + R \cdot \text{rot}(\dot{\theta}_{ki}(T - 1)') \cdot R^\top \cdot R \cdot x_{ki}(T - 1) \\ &= R q_k(T - 1) + b + R \cdot \text{rot}(\dot{\theta}_{ki}(T - 1)') \cdot x_{ki}(T - 1) \\ &= R (q_k(T - 1) + \text{rot}(\dot{\theta}_{ki}(T - 1)') \cdot x_{ki}(T - 1)) + b \\ &= R x_i(T) + b. \end{aligned} \quad (\text{A12})$$

Thus, TMPL is $E(3)$ -equivariant.

Appendix B. Quantitative Analysis of Rigid Constraint Preservation

To concretely illustrate the failure modes of existing models regarding geometric constraint preservation, we define a Rigid Structure Deviation (RSD) metric that quantifies the average violation of pre-defined rigid bond lengths in predicted conformations. Formally, for a set of rigidly connected atom pairs $\mathcal{R} = \{(i, j)\}$ with ground-truth bond lengths $\{d_{ij}\}$, the RSD at prediction horizon T is computed as

$$\text{RSD}(T) = \frac{1}{|\mathcal{R}|} \sum_{(i,j) \in \mathcal{R}} \left| \|\hat{x}_i(T) - \hat{x}_j(T)\| - d_{ij} \right|, \quad (\text{A13})$$

where $\hat{x}_i(T)$ denotes the predicted 3D coordinate of atom i at time T , and d_{ij} is the reference bond length extracted from the ground-truth frame. This metric directly measures how well a model preserves local rigid geometry during trajectory forecasting. We evaluate RSD on the rMD17 test set for eight small molecules, comparing STEGMN against representative baselines. Results are summarized in Table A1. All values are reported in Å.

Table A1. Rigid Structure Deviation (RSD) on the rMD17 dataset (values in Å). Lower values indicate better preservation of rigid bond lengths. Subscripts denote training epochs for ST-EGNN; STEGMN enforces hard constraints via generalized coordinates, yielding errors at machine precision.

Method	Asp	Bzn	Eth	Mal	Nap	Sal	Tol	Ura
ST-EGNN ₁₀₀	0.048	0.020	0.031	0.046	0.038	0.035	0.035	0.034
ST-EGNN ₅₀₀	0.015	0.011	0.011	0.016	0.019	0.018	0.013	0.018
STEGMN ₁₀₀	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
STEGMN ₅₀₀	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The zero RSD values reported for STEGMN (in fact, all deviations are below 1×10^{-4} Å and rounded to 0.000 in Table A1) directly reflect the hard constraint enforcement mechanism inherent to our generalized-coordinate formulation. Unlike Cartesian-space models such as ST-EGNN, which directly regress atomic coordinates and must implicitly learn bond-length conservation through data-driven optimization, STEGMN fundamentally shifts the prediction target from Cartesian positions to angular velocities defined on rigid structural objects. During spatiotemporal feature aggregation, our architecture integrates historical dynamic information by modeling angular accelerations—which naturally encode rigid-body kinematic priors—and concurrently updates the rotational state of each rigid unit. The resulting angular velocities are then mapped to future Cartesian configurations via rotation-matrix-based forward kinematics. By construction, this pipeline ensures that atomic positions are evolved solely through rigid-body rotations, guaranteeing that inter-atomic distances within pre-defined rigid connections remain exactly constant throughout the prediction horizon, up to the numerical precision of floating-point arithmetic. The non-zero RSD values for ST-EGNN, even after extended training, underscore the inherent limitation of data-driven approaches: equivariant message passing in Cartesian space alone cannot fully eliminate unphysical bond-length drift.

References

1. Corso, G.; Stärk, H.; Jing, B.; Barzilay, R.; Jaakkola, T. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv* **2022**, arXiv:2210.01776.
2. Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H.; et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J. Med. Chem.* **2019**, *63*, 8749–8760. [CrossRef]
3. Schütt, K.; Kindermans, P.J.; Sauceda Felix, H.E.; Chmiela, S.; Tkatchenko, A.; Müller, K.R. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 992–1002.

4. Duan, C.; Du, Y.; Jia, H.; Kulik, H.J. Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model. *Nat. Comput. Sci.* **2023**, *3*, 1045–1055. [\[CrossRef\]](#)
5. Finzi, M.; Stanton, S.; Izmailov, P.; Wilson, A.G. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In Proceedings of the International Conference on Machine Learning, PMLR, Online, 13–18 July 2020; pp. 3165–3176.
6. Brandstetter, J.; Hesselink, R.; van der Pol, E.; Bekkers, E.J.; Welling, M. Geometric and physical quantities improve e (3) equivariant message passing. *arXiv* **2021**, arXiv:2110.02905.
7. Keriven, N.; Peyré, G. Universal invariant and equivariant graph neural networks. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 7092–7101.
8. Yu, B.; Yin, H.; Zhu, Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv* **2017**, arXiv:1709.04875.
9. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3848–3858. [\[CrossRef\]](#)
10. Vignac, C.; Loukas, A.; Frossard, P. Building powerful and equivariant graph neural networks with structural message-passing. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 14143–14155.
11. Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J.P.; Kornbluth, M.; Molinari, N.; Smidt, T.E.; Kozinsky, B. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13*, 2453. [\[CrossRef\]](#)
12. Satorras, V.G.; Hoogeboom, E.; Welling, M. E (n) equivariant graph neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Online, 18–24 July 2021; pp. 9323–9332.
13. Bevilacqua, B.; Frasca, F.; Lim, D.; Srinivasan, B.; Cai, C.; Balamurugan, G.; Bronstein, M.M.; Maron, H. Equivariant Subgraph Aggregation Networks. In Proceedings of the International Conference on Learning Representations, Online, 25–29 April 2022.
14. Munikoti, S.; Agarwal, D.; Das, L.; Halappanavar, M.; Natarajan, B. Challenges and opportunities in deep reinforcement learning with graph neural networks: A comprehensive review of algorithms and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 15051–15071. [\[CrossRef\]](#)
15. Ge, D.; Dong, Z.; Cheng, Y.; Wu, Y. An enhanced spatio-temporal constraints network for anomaly detection in multivariate time series. *Knowl.-Based Syst.* **2024**, *283*, 111169. [\[CrossRef\]](#)
16. Kim, J.; Park, J.; Kim, N.; Yu, Y.; Chang, K.; Woo, C.S.; Yang, S.; Kang, N. Physics-constrained graph neural networks for spatio-temporal prediction of drop impact on OLED display panels. *Expert Syst. Appl.* **2025**, *274*, 126907. [\[CrossRef\]](#)
17. Huang, W.; Han, J.; Rong, Y.; Xu, T.; Sun, F.; Huang, J. Equivariant Graph Mechanics Networks with Constraints. In Proceedings of the International Conference on Learning Representations, Online, 25–29 April 2022.
18. Kofinas, M.; Nagaraja, N.; Gavves, E. Roto-translated local coordinate frames for interacting dynamical systems. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 6417–6429.
19. Xu, C.; Tan, R.T.; Tan, Y.; Chen, S.; Wang, Y.G.; Wang, X.; Wang, Y. Eqmotion: Equivariant multi-agent motion prediction with invariant interaction reasoning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 1410–1420.
20. Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Chang, X.; Zhang, C. Connecting the dots: Multivariate time series forecasting with graph neural networks. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, San Diego, CA, USA, 6–10 July 2020; pp. 753–763.
21. Jiang, B.; Zhang, Z.; Lin, D.; Tang, J.; Luo, B. Semi-supervised learning with graph learning-convolutional networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11313–11320.
22. Ye, W.; Qian, Q. Forecasting the molecular interactions: A hypergraph-based neural network for molecular relational learning. *Knowl.-Based Syst.* **2024**, *300*, 112177. [\[CrossRef\]](#)
23. Gasteiger, J.; Groß, J.; Günnemann, S. Directional message passing for molecular graphs. *arXiv* **2020**, arXiv:2003.03123.
24. Godwin, J.; Schaarschmidt, M.; Gaunt, A.; Sanchez-Gonzalez, A.; Rubanova, Y.; Veličković, P.; Kirkpatrick, J.; Battaglia, P. Simple gnn regularisation for 3d molecular property prediction & beyond. *arXiv* **2021**, arXiv:2106.07971.
25. Xie, T.; Grossman, J.C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **2018**, *120*, 145301. [\[CrossRef\]](#)
26. Zhang, M.; Chen, Y. Link prediction based on graph neural networks. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 5171–5181.
27. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* **2017**, arXiv:1710.10903.
28. Ramakrishnan, R.; Dral, P.O.; Rupp, M.; Von Lilienfeld, O.A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, 140022. [\[CrossRef\]](#)
29. Chmiela, S.; Tkatchenko, A.; Sauceda, H.E.; Poltavsky, I.; Schütt, K.T.; Müller, K.R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3*, e1603015. [\[CrossRef\]](#)

30. Unke, O.T.; Meuwly, M. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693. [CrossRef] [PubMed]
31. Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv* **2018**, arXiv:1802.08219.
32. Liao, Y.L.; Smidt, T. Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs. In Proceedings of the International Conference on Learning Representations, Online, 1–5 May 2023.
33. Liao, Y.L.; Wood, B.; Das, A.; Smidt, T. EquiformerV2: Improved Equivariant Transformer for Scaling to Higher-Degree Representations. In Proceedings of the International Conference on Learning Representations (ICLR), Vienna, Austria, 7–11 May 2024.
34. Hutchinson, M.J.; Le Lan, C.; Zaidi, S.; Dupont, E.; Teh, Y.W.; Kim, H. Lietransformer: Equivariant self-attention for lie groups. In Proceedings of the International Conference on Machine Learning, PMLR, Online, 18–24 July 2021; pp. 4533–4543.
35. Fuchs, F.; Worrall, D.; Fischer, V.; Welling, M. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1970–1981.
36. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.
37. Gilmer, J.; Schoenholz, S.S.; Riley, P.F.; Vinyals, O.; Dahl, G.E. Neural message passing for quantum chemistry. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, NSW, Australia, 6–11 August 2017; pp. 1263–1272.
38. Sun, M.; Zhou, P.; Tian, H.; Liao, Y.; Xie, H. Spatial-temporal attention network for crime prediction with adaptive graph learning. In *Proceedings of the International Conference on Artificial Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 656–669.
39. Christensen, A.S.; von Lilienfeld, O.A. Revised MD17 Dataset (rMD17). Figshare. Dataset. 2020. Available online: https://figshare.com/articles/dataset/Revised_MD17_dataset_rMD17_/12672038/3 (accessed on 3 May 2026). [CrossRef]
40. Qu, K.; Xue, S.; Zheng, X.; Yan, D.; Cao, H. Learning dynamic inter-farm dependencies for wind power forecasting via adaptive sparse graph attention network. *Renew. Energy* **2025**, *258*, 124969. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.