

# MMI706

## Reinforcement Learning

METU  
Phase-I

18/03/2024

### 1. INTRODUCTION

Variable Geometry Truss (VGT) robots are highly modular, deformable, and controllable truss structures. The topology comprises two main components: nodes and members. Members are linear actuators with a single degree of freedom within constraints, while nodes are passive joints that form the mechanism. Despite trusses typically being rigid, VGTs are designed for locomotion, maneuvering around obstacles. Various robot topologies exist, ranging from simpler forms like tetrahedrons or octahedrons to more complex ones. Let's consider the octahedral structure as an example, consisting of twelve members and six nodes. Although seemingly less complex than a quadruped robot with the same number of actuators, locomotion in quadrupeds is often more intricate due to quasi-static balancing requirement. Optimization-based locomotion and path planning algorithms have been extensively researched, but many proposed algorithms are designed to specific conditions, potentially limiting their adaptability. Exploring different algorithms, particularly using reinforcement learning, can offer a broader exploration of locomotion possibilities.

### 2. METHODS

Our goal is to address locomotion challenges in VGTs using reinforcement learning. Reinforcement learning offers two significant advantages over classical methods. Firstly, its broad training approach enables a single training session to apply to a variety of robots. Secondly, modular agent methods allow tackling larger robot problems by assigning an agent to each member, effectively treating each member as an individual robot with a single degree of freedom. Reviewing existing literature reveals modular training methods, such as using different modules (like limbs and wheels) on a ground platform, resulting in effective locomotion for various robots. Additionally, employing a distributed coach approach, where each actuator has its own agent guided by a pseudo-agent, has shown effectiveness in snake robots with numerous joints. However, this approach's applicability to VGTs may vary due to differences in kinematic structure. Considering the task's complexity, we'll simplify aspects like limiting the range of topologies initially and assume a flat environment without obstacles. We'll conduct experiments using NVIDIA Isaac Gym, a standard choice for parallel training, leveraging the Proximal Policy Optimization (PPO) algorithm known for its effectiveness in similar tasks as shown in the two mentioned papers.

### 3. CONTRIBUTIONS & LIMITATIONS

There are several risks and limitations to consider in this work. Applying concepts from different robots requires a deep understanding of both methods, and modular exploration might face challenges due to the simplicity of discrete linear actuators. Training with different robots may not generalize locomotion aims, leading to convergence issues. Modularization efforts could increase input size and time consumption. Additionally, applying a cyclic gait mechanism observed in the snake robot locomotion, not applicable to VGTs, might yield unreproducible results. However, even in the face of method-independent failures, having a baseline locomotion achieved through PPO training ensures progress. Challenges such as implementing self-collision checks or requiring additional sensory data like ground reaction sensing may also arise, impacting locomotion strategy.