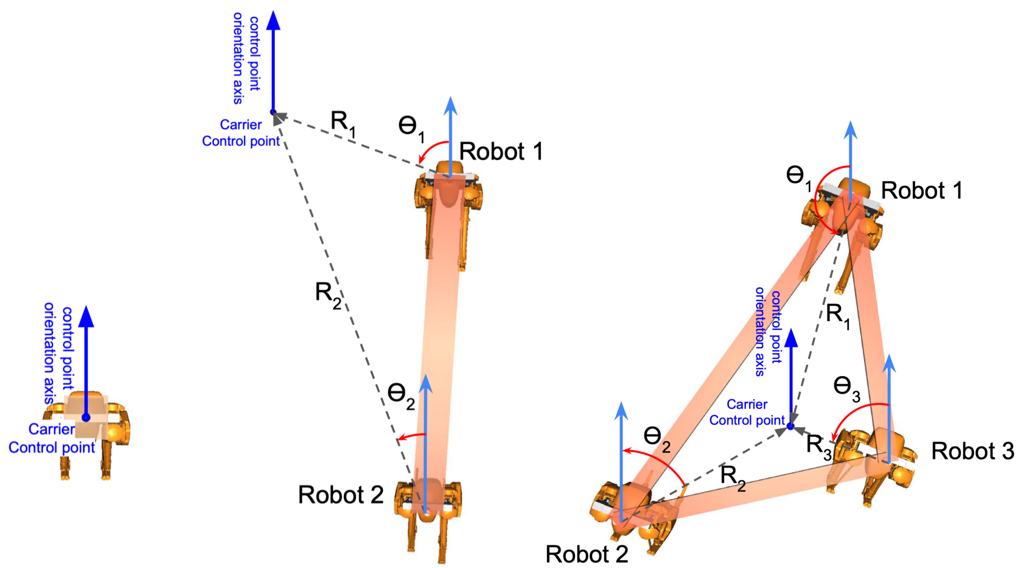
A Implementation details

In this section, we describe the implementation details of our decMBC approach, focusing on the configuration generation, and randomization of parameters used during training.

For generating N-robot configuration during training, where $N \in \{1, 2, 3\}$, we choose relative position p_i $\forall i \in \{1 \dots N\}$ with respect to the carrier control point which is a 2D polar coordinate indicated by (R_i, Θ_i). Figure 1 shows an example of robot configurations during training. In Figure 1a ($N = 1$), the carrier control point is located at the base of the robot, therefore relative position (R_1,Θ_1) is always zero. In Figure 1b ($N = 2$), we randomize carrier control point location up to 1 meter from any of the robots, and similarly, in Figure 1c ($N = 3$), we randomize carrier control point to be located anywhere enclosed within the triangular frame, which produces a range of relative position (R_i, Θ_i). Furthermore, in each of these configurations, we randomize the mass of the rigid bar connection as shown in Table 1.

The ranges of relative positions and masses seen during training generalize for various carrier configurations with an arbitrary number of robots and payloads as observed in the experiments.



(a) 1-robot configuration (b) 2-robot configuration (c) 3-robot configuration

Figure 1: Training configuration using 1, 2 and 3 robots

Parameter	Ranges
X velocity (V_x)	[-0.5, 2.0] m/s
Y velocity (V_y)	[-0.3, 0.3] m/s
Angular velocity (ω)	[-π/8, π/8] rad/s
Height (h)	[0.5, 0.8] m
R_i	[0, 3.5] m
Θ_i	[-π, π] rad
Commands Duration	[100, 450] timesteps
Force Duration	[50, 200] timesteps
Rigid bar masses	[0, 10] kg (1-robot configuration)
	[0, 20] kg (2-robot configuration)
	[0, 15] kg (3-robot configuration)

Table 1: Parameters ranges for training

B Reward details

Below, we outline the reward terms used in our implementation. The reward components are divided into two categories: Local rewards r_i^L are designed to produce stable locomotion for each robot, described in the local robot frame. The Global rewards r^G are designed for following carrier commands with the combined action of all robots in the control point frame.

Category	Reward Term	Definition	Weighting
r_i^L Local - Robot $i \in \{1, \ldots, N\}$	Feet airtime	$\begin{cases} 0.0 & \text{if } c_t = 0 \\ \sum_{f\in(l,r)} \lvert t^i_{f,air} - 0.35\rvert \cdot \mathbb{I}^i_{td,f} & \text{else} \end{cases}$	1.0
	Feet contact	$\begin{cases} (n_c^i == 2) + 0.5 * (n_c^i == 1) & \text{if } c_t = 0 \\ (n_c^i == 1) & \text{else} \end{cases}$	0.1
	Feet stance X	$\begin{cases} exp(-10 * \lvert feet^i_{l,x} - feet^i_{r,x}\rvert) & \text{if } c_t = 0 \\ exp(0) & \text{else} \end{cases}$	0.02
	Feet stance Y	$\begin{cases} exp(-5 * \lvert feet^i_{l,y} - feet^i_{r,y} - 0.385\rvert) & \text{if } c_t = 0 \\ exp(0) & \text{else} \end{cases}$	0.02
	Feet orientation	$exp(-30 * \sum_{f\in(l,r)} q_d(feet^i_{f,rpy}, base^i_{rpy}))$	0.15
	Relative yaw orientation	$-\lvert\theta^i_t\rvert/\pi$	0.5
	Joint force	$\begin{cases} exp(-0.2 * \lvert\lvert joint^i_{force,xy}\rvert\rvert) & \text{if } c_t = 0 \, and \, N > 1 \\ exp(0) & \text{else} \end{cases}$	0.1
	Base Height	$-\lvert base^i_z - h\rvert$	0.2
	Base acceleration	$exp(-0.01 * \sum \lvert base^i_{accel,xyz}\rvert)$	0.1
	Action difference	$exp(-8 * \sum \lvert a^i_t - a^i_{t-1}\rvert)$	0.1
	Torque	$exp(-\sum \lvert\tau^i\rvert/\tau^i_{max})$	0.05
r^G Global - Carrier	X,Y velocity	$exp(-2 * (\hat{V}_{xy} - V_{xy}))$	0.15, 0.1
	Orientation	$-q_d(O_{c,rpy}, \omega)$, $exp(-30 * q_d(O_{c,rpy}, \omega))$	2.0, 0.15

Table 2: Description of reward terms

Notations used in Table 2 for the reward description.

- N = Number of robots in the system
- c_t = Carrier command consisting of linear velocity, angular velocity, and base height. $c_t = (V_x, V_y, \omega, h)$
- $c_t = 0$ means a hold still command
- $t^i_{f,air}$ = air time for a particular feet
- $\mathbb{I}^i_{td,f}$ = boolean variable indicating a touchdown for a particular feet
- n^i_c = number of feet in contact with the ground
- $base^i$ = Robot pelvis global pose
- $feet^i$ = Pose of each foot in robot base frame
- $q_d(\cdot)$ = Quaternion distance function
- θ^i_t = Yaw orientation of each robot relative to the Carrier
- $joint^i_{force}$ = Force sensor reading on the ball joint connection
- a^i_t = Action taken by the policy

- τ^i = Motor torques
- O_c = Control point on the carrier
- $\hat{V}_{xy}$ = Current linear velocities of the control point

C Termination condition

An episode is terminated if any of the following conditions are met:

- The carrier's pitch or roll $O_{c,rp}$ is outside the range $[-30°, 30°]$.
- The robot's pelvis pitch or roll $base^i_{rp}$ is outside the range $[-30°, 30°]$.
- The robot's knee collides with the ground or with each other.
- The relative yaw orientation θ^i_t of each robot is outside the range $[-30°, 30°]$.
- The robot's pelvis height $base^i_z$ is outside the range $[0.5, 1.0]$
- The episode duration exceeds 500 timesteps (10 seconds).

D Dynamics Randomization

To facilitate transfer from simulation to the real robot, we employ dynamics randomization throughout all stages of the curriculum during training. This involves randomizing various physical parameters of the simulated environment, such as joint damping, link masses, center of mass positions, encoder noise, ground friction, terrain variations, and adding noise to states and policy rate. Table 3 lists the dynamics randomization ranges used.

Element	**Range**
Joint Damping	[-50, 250] $\%$
Body Mass	[-25, 25] $\%$
CoM Position	[-1, 1] $\%$
Friction	[-20, 20] $\%$
Encoder Noise	[-0.05, 0.05] rad
Ground Slope	[-0.05, 0.05] rad

Table 3: Dynamics randomization ranges

E PPO Hyperparameters

Table 4 lists the hyperparameters used for the PPO training.

Hyperparameter	**Value**
Learning Rate (Actor & Critic)	3e-4
PPO Clip Range	0.2
Number of Epochs	5
Batch Size (Episodes)	32
Discount Factor (Gamma)	0.95
GAE Lambda	1.0
Value Function Coefficient	0.5
Entropy Coefficient	0.01
Max Grad Norm	0.05
Buffer Size	60000

Table 4: PPO Hyperparameters