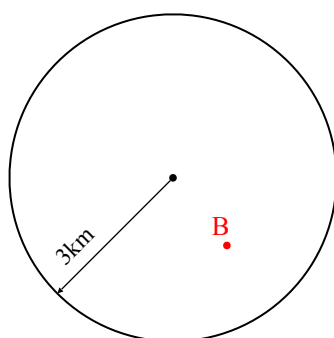


# F4v1 项目说明

该项目是在一个有限的圆柱体空间内通过一定策略操纵4个物体A对1个物体B进行探寻并追击。物体A与物体B的平面速度、加速度、垂直速度、活动范围（俯视）、深度、转向速度、视野（物体A专属）等信息根据客户需求进行了限制。

## 1 场景描述

场景描述可以细分为环境设定、单位设定、状态判定、操纵设定以及任务目标等五个部分。



### 1.环境设定

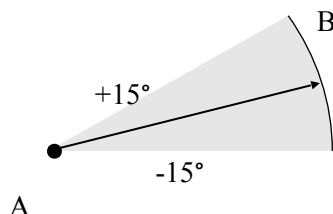
该环境半径为3km，深度范围为[-240,0]m。物体A与B均为质点。

### 2.单位设定

A物体的初速度数值与方向随机，初始水平位置为环境中心点，深度为-50m。其水平速度范围为[15.433,25.722]m/s，垂直速度为[-2,2]m/s。A的最大角度变化 $\theta' = (-0.0657v^2 + 3.5332v - 7.3176) / 57.3(\text{radius})$ ，最大加速度 $a = \theta' \cdot v$ 。

B物体的初速度与方向随机，初始水平位置与初始深度均为环境允许的随机数值。其平面速度范围为[0,12]m/s，垂直速度为[-1,1]m/s。B的最大角度变化 $\theta' = (-0.000185v^3 - 0.006391v^2 + 0.11501v) / 57.3(\text{radius})$ ，最大加速度 $a = \theta' \cdot v$ 。

### 3.状态判定



发现与锁定：当物体B处于物体A运动方向15°内范围且空间距离不超过2km时，物体A可以锁定B。其中，空间距离在1km范围内的锁定概率为100%，空间距离超过1km的锁定概率为10%。当某一A发现B后，所有物体A共享视野信息。

击中：当物体A与物体B的空间距离小于30m且垂直距离小于30m时，判定A成功击中B。

#### 4.操纵设定

操纵指令包括三个参数，**X**为物体的期望顺时针旋转角度（实际旋转角度需满足角度限制），**Y**为物体的期望水平速度（实际速度需满足水平速度限制），**Dv**为物体的期望垂直速度（实际速度需满足垂直速度限制）。

#### 5.任务目标

整个任务可分为**发现前**与**发现后**两个阶段。

第一阶段：令A尽快探索并对B进行锁定。

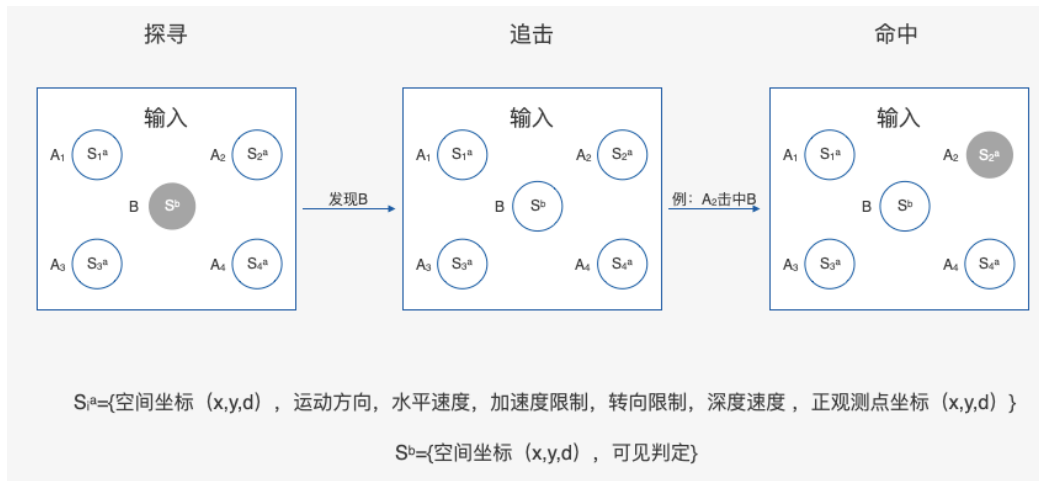
第二阶段：在某一物体A发现B的前提下，令所有的A尽快全部击中B。

## 2 算法设计

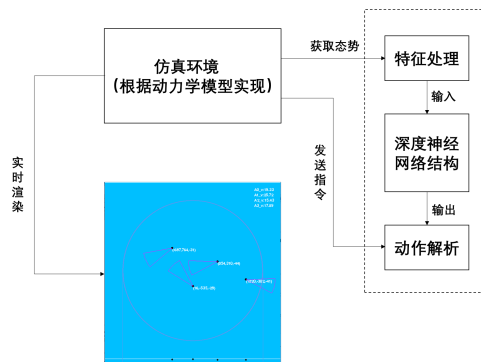
### 1.智能体设计

该项目以多智能体算法为基础，设计了四个基于深度神经网络构建的player分别控制物体A。

每个深度神经网络的输入包括当前控制物体A的空间坐标、水平速度、运动方向、垂直速度、速度限制值、生存判定等信息，以及以当前控制物体A作为基本点的其他A的相对空间坐标与基本信息。



在每一个时间戳下，智能体从仿真环境中获取当前的态势，并将其转换为深度神经网络的输入。深度神经网络根据当前的输入数据给出包括旋转角度、期望水平速度以及期望垂直速度等信息的输出数据。最后，我们将输出数据进行动作解析，并将解析后的结果发送给物体A，控制其在环境中做出相应动作。



### 2.奖励设计

奖励可分为**探索奖励**与**击中奖励**。在锁定前，智能体将根据探索到物体B与否获取奖励值；在发现物体B后，智能体根据探索到B的**时间长短**与**撞击速度**获取奖励值。目前的奖励设定A越早发现B，获取奖励越高；在发现B的前提下，A越早击中B，奖励越高。

### 3.智能体训练

该项目使用了PPO算法对智能体进行了训练，PPO是一种基于多智能体的深度强化学习算法。智能体通过收集物体A与环境进行交互的数据与对应奖励值对深度神经网络的参数进行持续更新。

## 3 效果说明

- a.黑色代表物体A，红色代表物体B，对应空间坐标见白色文字。
- b.右上方为物体A的实时速度，白色代表尚未击中B，红色代表已经击中B。
- c.智能体每5秒决策一次，共200步，即1000秒为一局。
- d.击中判定条件为物体A与物体B的空间距离为30m内且垂直距离为30m内。
- e.图像上方为俯视图，下方为侧视图，A的探测区域以扇形表示（半径为1km）。
- f.当某一A击中B时，图中对应点由黑色变为灰色。
- g.当A与B的平面距离较近而垂直距离较远时，A会绕B旋转并同时上浮或下潜运动，直至其满足击中判定。

