Final Model - SAC

This was designed to be run on a Tensorflow 2.10 enviornment for native GPU usage.

In [1]:
```
%load_ext autoreload
%autoreload 2

import tensorflow as tf
print("TensorFlow version:", tf.__version__)
print("tf.keras available:", hasattr(tf, "keras"))
print("Num GPUs Available: ", len(tf.config.list_physical_devices('GPU')))
```

```
TensorFlow version: 2.10.0
tf.keras available: True
Num GPUs Available:  1
```

In [2]:
```
from structure import DMC_structure
import numpy as np
import matplotlib.pyplot as plt
import os

from DMC_Env import DMC_Env

import logging
from datetime import datetime

from sac import SoftActorCritic, Actor
from replay_buffer import ReplayBuffer

tf.keras.backend.set_floatx('float32')

logging.basicConfig(level='INFO')
```

As per the format explained previously, the DMC chain is initialized below.

In [3]:
```
DMCarr = [[] for i in range(10)]
            # index, next, func, goal, input (T, P, Keq)
DMCarr[0] = [0, [0, 1], "DMC0", 400, [350, 5, 1]]
DMCarr[1] = [1, [2], "DMC1", 500, [350, 5, 1]]
DMCarr[2] = [2, [3, 5], "DMC2", 500, [350, 5, 1]]
DMCarr[3] = [3, [4], "DMC3", 500, [350, 5, 1]]
DMCarr[4] = [4, [5], "DMC4", 500, [350, 5, 1]]
DMCarr[5] = [5, [], "DMC5", 500, [350, 5, 1]]
DMCarr[6] = [6, [1, 6], "DMC6", 500, [350, 5, 1]]
DMCarr[7] = [7, [3, 7], "DMC7", 500, [350, 5, 1]]
DMCarr[8] = [8, [5, 8], "DMC8", 500, [350, 5, 1]]
DMCarr[9] = [9, [2, 3, 4, 9], "DMC9", 500, [350, 5, 1]]
# DMCarr[2] = [2, [], "Dummy", 0, [0, 0, 0]]

print("DMC array:", DMCarr)
struct = DMC_structure(DMCarr)
```

```
DMC array: [[0, [0, 1], 'DMC0', 400, [350, 5, 1]], [1, [2], 'DMC1', 500, [35
0, 5, 1]], [2, [3, 5], 'DMC2', 500, [350, 5, 1]], [3, [4], 'DMC3', 500, [350,
5, 1]], [4, [5], 'DMC4', 500, [350, 5, 1]], [5, [], 'DMC5', 500, [350, 5,
1]], [6, [1, 6], 'DMC6', 500, [350, 5, 1]], [7, [3, 7], 'DMC7', 500, [350, 5,
1]], [8, [5, 8], 'DMC8', 500, [350, 5, 1]], [9, [2, 3, 4, 9], 'DMC9', 500, [3
50, 5, 1]]]
```

Enviornment Setup

In [8]:
```
args = {
    'seed': 42,
    'render': False,
    'verbose': False,
    'batch_size': 128,
    'epochs': 50,
    'start_steps': 0,
    'model_path': './data/models/',
    'model_name': f'{str(datetime.utcnow().date())}-{str(datetime.utcnow().tim
e())}',
    'gamma': 0.99,
    'polyak': 0.995,
    'learning_rate': 0.001,
}
```

In [9]:
```
# Define DMC environment setup
env = DMC_Env(DMCarr)

state_space = env.observation_space.shape[0]
action_space = env.action_space.shape[0]

replay = ReplayBuffer(state_space, action_space)

log_dir = args['model_path'] + '/logs/' + datetime.utcnow().strftime("%Y%m%d-%
H%M%S")
writer = tf.summary.create_file_writer(log_dir)

sac = SoftActorCritic(action_space, writer,
                      learning_rate=args['learning_rate'],
                      gamma=args['gamma'],
                      polyak=args['polyak'])
```

In [20]:
```
n = 10 #of episodes to run
timestamp = datetime.utcnow().strftime("%Y-%m-%d-%H-%M-%S")
print("Current working directory:", os.getcwd())
print("Absolute save path:", os.path.abspath(os.path.join(
        args['model_path'], timestamp)))
```

```
Current working directory: c:\Users\dista\OneDrive - Georgia Institute of Tec
hnology\_Sem 10\CHBE 4803 AI for ChemE\Project\proc-control
Absolute save path: c:\Users\dista\OneDrive - Georgia Institute of Technology
\_Sem 10\CHBE 4803 AI for ChemE\Project\proc-control\data\models\2025-04-18-2
2-10-11
```

Training and saving loop:

In [ ]:
```
import os
from datetime import datetime

episode_rewards = []
global_step = 0
episode = 0
prev_avg_episode_reward = None

# Run for n episodes
for ep in range(n):
    current_state = env.reset()
    step = 1
    episode_reward = 0
    done = False

    while not done:
        if args['render']:
            env.render()

        # Choose action
        if global_step < args['start_steps']:
            action = env.action_space.sample()
        else:
            action = sac.sample_action(current_state)
            if np.isscalar(action):
                action = np.array([action])

        # Environment step
        next_state, reward, done, _ = env.step(action)
        episode_reward += reward
        end = 0 if done else 1

        # Optional logging
        if args.get('verbose', False):
            logging.info(f"Global step: {global_step}")
            logging.info(f"Current state: {current_state}")
            logging.info(f"Action: {action}")
            logging.info(f"Reward: {reward}")
            logging.info(f"Next state: {next_state}")
            logging.info(f"End flag: {end}")

        replay.store(current_state, action, reward, next_state, end)

        current_state = next_state
        step += 1
        global_step += 1

    # Training phase
    if replay.total_size > args['batch_size'] and global_step > args['start_st
eps']:
        for epoch in range(args['epochs']):
            batch = replay.fetch_sample(num_samples=args['batch_size'])
            critic1_loss, critic2_loss, actor_loss, alpha_loss = sac.train(*ba
tch)

            if args.get('verbose', False):
```

```
                print(f"Episode {ep}, Global step {global_step}, Epoch {epoc
h}:",
                      critic1_loss.numpy(), critic2_loss.numpy(),
                      actor_loss.numpy(), f"Episode Reward: {episode_reward}")

            sac.epoch_step += 1
            sac.update_weights()

    # Book-keeping
    episode_rewards.append(episode_reward)
    avg_episode_reward = np.mean(episode_rewards[-100:])

    print(f"Episode {ep+1} reward: {episode_reward}")
    print(f"Episode {ep+1} average reward (last 100): {avg_episode_reward}")

    if prev_avg_episode_reward is not None:
        print("Δ average reward:", avg_episode_reward - prev_avg_episode_rewar
d)
    prev_avg_episode_reward = avg_episode_reward

# ---------- SAVE AFTER TRAINING FINISHES ----------
timestamp = datetime.utcnow().strftime("%Y-%m-%d-%H-%M-%S")
save_dir = os.path.join(args['model_path'], timestamp)
os.makedirs(save_dir, exist_ok=True)

sac.policy.save_weights(os.path.join(save_dir, "policy.h5"))      # checkpoint
pair
np.save(os.path.join(save_dir, "alpha.npy"), sac.alpha.numpy())  # keep α if y
ou use it

print(f"\nModel saved to {save_dir}")
```

```
Episode 1 reward: -479.32556969718144
Episode 1 average reward (last 100): -479.32556969718144
Episode 2 reward: -496.03960596318564
Episode 2 average reward (last 100): -487.68258783018354
Δ average reward: -8.3570181330021
Episode 3 reward: -473.1511290236606
Episode 3 average reward (last 100): -482.8387682280092
Δ average reward: 4.843819602174335
Episode 4 reward: -437.702547327856
Episode 4 average reward (last 100): -471.55471300297086
Δ average reward: 11.284055225038344
Episode 5 reward: -435.7750240782949
Episode 5 average reward (last 100): -464.39877521803567
Δ average reward: 7.15593778493519
Episode 6 reward: -436.4576038102979
Episode 6 average reward (last 100): -459.741913316746
Δ average reward: 4.6568619012896875
Episode 7 reward: -443.170700564223
Episode 7 average reward (last 100): -457.3745972092427
Δ average reward: 2.3673161075032567
Episode 8 reward: -435.54420696920437
Episode 8 average reward (last 100): -454.64579842923797
Δ average reward: 2.7287987800047517
Episode 9 reward: -442.70183562307784
Episode 9 average reward (last 100): -453.31869145077576
Δ average reward: 1.3271069784622114
Episode 10 reward: -440.50839979736725
Episode 10 average reward (last 100): -452.0376622854349
Δ average reward: 1.2810291653408399

Model saved to ./data/models/2025-04-18-20-06-37
```

In [21]:
```
# only run if needed
timestamp = datetime.utcnow().strftime("%Y-%m-%d-%H-%M-%S")
save_dir = os.path.join(args['model_path'], timestamp)
os.makedirs(save_dir, exist_ok=True)

sac.policy.save_weights(os.path.join(save_dir, "policy.h5"))      # checkpoint
pair
np.save(os.path.join(save_dir, "alpha.npy"), sac.alpha.numpy())  # keep α if y
ou use it

print(f"\nModel saved to {save_dir}")
```

```
Model saved to ./data/models/2025-04-18-22-10-52
```

This code runs the model and give sample actions

In [27]:
```
import DMC_Play

# DMC_Play.main(['--verbose']) #pull most recent model
DMC_Play.main([
    "--model_root", "./data/models",
    "--model_name", "2025-04-18-22-10-52"
])
```

```
Action: [-0.8922216  -0.21486107 -0.77755433  0.22788236 -0.34593025 -0.57428
17
  0.07466309  0.8595521  -0.4163412  -0.8110762 ]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
```

```
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
```

```
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
```

```
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
```

```
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
```

```
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
```

```
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
```

```
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
```

```
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
```

```
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
```

```
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Action: [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.99999
9
  0.999999 -0.999999 -0.999999]
Episode reward: -435.4767408139788
Last Action [-0.999999 -0.999999 -0.999999 -0.999999  0.999999  0.999999  0.9
99999
  0.999999 -0.999999 -0.999999]
```

Advancements:

- running on GPU, but push to ICE in worst case. Every loop now takes just 30 seconds.
- Reward function is updated, with minimal performance improvements. It's still absurdly negative - which indicates that the actions picked and initial conditions are terrible.
- testing script shows that the action space is correctly defined

SAC3:

- now fixed the scaling in the action space tanh
- fixed the scaling in the reward function to not explode
- the reason why we have average ereward and episode reward is becuase we keep a running average to track how we are doing over time. we do see an update, and we do also see a convergence to a stable solution within 10 episodes.
- We should figure out a way to print the final model and that's it - recommend a stateless bandit.

Discussion:

- well.. it does converge even if it's terrible.
- stateless bandit as a "testing version"
- it's possible it's just the solutions the DMC's are optimized to