Supplementary Information for AcoustoReinforce: Multi-Particle Acoustophoretic Path Planning with Deep Reinforcement Learning

1 Pseudocode of MADDPG

The MADDPG implementation used to train the network policy.

```
ALGORITHM 1: MADDPG
Initialize critic networks $\mathbf{Q} = \{Q_1, Q_2, \ldots, Q_N\}$ with random parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_N\}$
Initialize actor networks $\boldsymbol{\pi} = \{\pi_1, \pi_2, \ldots, \pi_N\}$ with random parameters $\boldsymbol{\phi} = \{\phi_1, \phi_2, \ldots, \phi_N\}$
Initialize target networks $\boldsymbol{\theta}' \leftarrow \boldsymbol{\theta}$, $\boldsymbol{\phi}' \leftarrow \boldsymbol{\phi}$
Initialize replay buffer $\mathcal{B}$
Obtain an initial state $s$
Obtain initial observations $o_1, o_2, \ldots, o_N = \Omega(s)$
for timestep $t = 1$ to total timestep do
    for particle $i = 1$ to $N$ do
        Select action with an exploration noise $\sigma$: $a_i \leftarrow \pi_{\phi_i}(o_i) + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma)$
    end
    Execute the joint action $a = \{a_1, a_2, \ldots, a_N\}$
    Obtain reward $r$ and new state $s'$
    Obtain new observations $o'_1, o'_2, \ldots, o'_N = \Omega(s')$
    Store transition tuple $(s, a, r, s')$ in $\mathcal{B}$
    Update state $s \leftarrow s'$
    if $\mathcal{B}.size() \geq$ minimal size then
        Sample a mini-batch of $n$ transitions $(s, a, r, s')$ from $\mathcal{B}$
        for particle $i = 1$ to $N$ do
            $y \leftarrow r_i + \gamma Q_{\theta'_i}(s', a'_1, \ldots, a'_j, \ldots, a'_N)|_{a'_j = \pi_{\phi'_j}(o'_j)}$
            Update the critic: $\theta_i \leftarrow \theta_i + \frac{1}{n}\sum(y - Q_{\theta_i}(s, a))^2$
            Update the actor $\phi_i$ by the deterministic policy gradient:
                $$\nabla_{\phi_i} J(\phi_i) = \frac{1}{n} \sum \nabla_{\phi_i} \pi_{\phi_i}(o_i) \nabla_{a_i} Q_{\theta_i}(s, a_1, \ldots, a_i, \ldots, a_N)|_{a_i = \pi_{\phi_i}(o_i)}$$
        end
        Update target networks:
        $\boldsymbol{\theta}' \leftarrow \tau\boldsymbol{\theta} + (1 - \tau)\boldsymbol{\theta}'$
        $\boldsymbol{\phi}' \leftarrow \tau\boldsymbol{\phi} + (1 - \tau)\boldsymbol{\phi}'$
    end
end
```

2 The Hyperparameters in MADDPG

The hyperparameters used in MADDPG.

Parameter	**Value**
Total timestep	5e5
Replay buffer size	5e5
Minimum buffer size	2.5e4
Batch size	512
Exploration noise θ	0.3
Discount factor γ	0.8
Soft update τ	0.01
Actor learning rate	$1e-4$
Critic learning rate	$4e-4$

Table 1: The hyperparameters of policy training.

3 Simulation-based Evaluation

Table 2: Success rate and computation time of different path planning methods in simulated evaluation.

Metrics	Method	4	6	8	10
Success Rate	AcoustoReinforce	**0.9980**	**0.9980**	**0.9820**	**0.9830**
	S2M2	0.9810	0.9650	0.9280	0.8920
	CBS	0.9880	0.9890	0.9720	0.9370
Runtime	AcoustoReinforce	5.550 ± 1.210	5.822 ± 0.2953	6.226 ± 1.545	7.192 ± 0.9869
	S2M2	**0.3913** ± 0.3962	**0.5893** ± 0.6686	**0.8648** ± 0.9953	**1.263** ± 1.268
	CBS	11.29 ± 51.90	11.88 ± 51.03	20.64 ± 70.41	30.10 ± 91.65

All evaluations were carried out on a Windows 11 machine with an AMD Ryzen 9 4900H CPU and 16 GB of RAM. Simulation results demonstrate that AcoustoReinforce average runtime ($5.55 - 7.19s$) is longer than that of S2M2 ($0.39 - 1.26s$), it remains significantly lower than that of CBS ($11.29 - 30.10s$). Moreover, AcoustoReinforce exhibits a smaller standard deviation in runtime, indicating greater stability.

4 Real-World Evaluation with Different Solvers

Table 3: Stability rate comparison for 8 particles under different V_{max} values. For each of the three planners, 100 solutions were tested using both the Naive and TWGS solvers. AcoustoReinforce performs Gor'kov optimizations under each solver and consistently outperforms both baseline methods in real-world experiments.

Hologram Solver	Path Planner	Velocities (m/s)			
		0.2	0.15	0.1	0.05
Naive	CBS	0.0700	0.3600	0.6800	0.8400
	S2M2	0.0400	0.3800	0.6600	0.8900
	AcoustoReinforce	**0.3000**	**0.5700**	**0.8400**	**0.9600**
TWGS	CBS	0.2600	0.5800	0.7800	0.9400
	S2M2	0.2000	0.6000	0.8100	0.9600
	AcoustoReinforce	**0.4600**	**0.7400**	**0.9100**	**0.9900**