Deep Reinforcement Learning based Super Twisting Controller for Liquid Slosh Control Problem

Ashish Kumar Shakya* Kshitij Bithel G. N. Pillai*** Sohom Chakrabarty ******

**Indian Institute of Technology, Roorkee, Haridwar, India (e-mail: akumarshakya@ee.iitr.ac.in).*
*** Indian Institute of Technology, Roorkee, Haridwar, India (e-mail: kbithel@ee.iitr.ac.in)*
**** Indian Institute of Technology, Roorkee, Haridwar, India (e-mail: sohom.chakrabarty@ee.iitr.ac.in*
***** Indian Institute of Technology, Roorkee, Haridwar, India (e-mail: gn.pillai@ee.iitr.ac.in))*

Abstract: Deep Reinforcement Learning (DRL) based parameter optimization of super twisting control (STC) for the liquid slosh control problem in a moving vehicle is proposed in this paper. The slosh control problem, including the vehicle dynamics, represents an under-actuated nonlinear dynamical system. The slosh phenomenon is modeled by a simple pendulum on a cart and STC had been designed for the system when the vehicle motion is in a straight line. In this paper, a DRL framework is designed for the first time to tune the STC parameters in order to deliver near optimal performance. The effectiveness of this proposed learning-based approach for STC design for the slosh control problem is validated in a Python simulation environment and compared to the simple STC design without the learning.

Keywords: Deep reinforcement learning, super twisting control, lateral slosh, under-actuated system.

1. Introduction

Sloshing is commonly defined as the movement of a free liquid surface within its container. During the translational or rotational accelerations of the liquid containers, a substantial volume of liquid tends to move unrestrained in the containers, generating the sloshing problem that can lead to system failures. The sloshing problem frequently occurs in partially filled containers in a variety of applications, including oil and liquefied natural gas storage tankers, liquid cargo carriers, liquid rocket fuel tanks, melted metal handling in steel plants, the beverage industry, and attitude or trajectory maneuvers in spacecraft (Abramson, 1966). It has a significant negative impact on the performance of industrial processes and is hard to eradicate.

Liquid fuel sloshing often has a considerable impact on the motion of partially filled spacecraft in the presence of attitude or trajectory maneuvers and can even cause instability (Nichkawde, 2004; Vreeburg, 2005). Fuel slosh and its effects on spacecraft dynamics have been studied extensively in recent decades (Peterson and Crawley, 1989; Baozeng et al., 2016).

Hence, it is essential to analyze and precisely characterize the sloshing phenomenon, as well as to establish, identify, and experimentally evaluate mathematical models of slosh that may be employed control development.

The motion of the liquid occurs in different forms based on the nature of the functional force, the container geometry, etc. As a result, different sloshing phenomena arise (Abramson, 1966; Ibrahim et al., 2001), such as lateral, rotational, swirling, or even chaotic, quasi-periodic sloshing.

The combined dynamics of slosh and vehicle form a nonlinear under-actuated dynamical system. Underactuated systems (Spong, 1998) are those in which the number of configurable variables to be controlled exceeds the number of actuators available to do so. When Petit and Rouchon (2002) used the fluid dynamic approach to tackle the slosh control problem, they ran into several control problems. This approach also requires the real-time computation of complex equations such as Navier-Stokes equations that describe fluid motion. This can be achieved using a Computational Fluid Dynamics (CFD) technique, but it is computationally costly and uncontrollable (Abramson, 1966; Ibrahim et al., 2001).

A nonlinear and complicated mathematical model can be utilized to represent the sloshing dynamics (Peterson and Crawley, 1989), but such dynamics becomes too challenging for the controller design. This necessitates the development of simpler mathematical models for slosh in order to save computational time and expense while providing controllable models. Equivalent mechanical models are effective in this situation as they simplify fluid dynamics equations by presuming oscillatory point masses and rigid bodies, making the control model easier. To represent the sloshing phenomenon, spring-mass damper and pendulum models are commonly used. Moving mass in these models is used to represent the sloshing mass of the liquid. Abramson (1966) and Ibrahim et al. (2001) discussed the pendulum model for slosh phenomena. The use of a simple pendulum model to represent lateral slosh, which is a type of linear and planar slosh, has been extensively accepted and documented in the literature (Abramson, 1966; Ibrahim et al., 2001).

Many scientists have sought to find solutions to the difficult issues that sloshing dynamics pose. Different passive control techniques like baffles (Abramson, 1966; Ibrahim et al., 2001) are reported to control the sloshing effects in launch vehicles specifically and in other applications alike. However, it increases the system's weight and, as a result, the cost, making it less desirable. Researchers have been increasingly interested in active control solutions for slosh suppression over the last two decades (Yano and Terashima, 2001; Gandhi and Duggal, 2009; Bandyopadhyay et al., 2009b; Thakar et al., 2012; Thakar et al., 2017a).

Some have used various control schemes such as PID control (Sira-Ramírez and Fliess, 2002), sliding mode control (Bandyopadhyay et al., 2009a, 2009b; Thakar et al., 2013; Kurode et al., 2013), H∞ control (Yano and Terashima, 2001), adaptive nonlinear dynamic inversion control (Weerdt et al., 2008), linear quadratic regulator, linear quadratic Gaussian control (De Souza and De Souza, 2014), Lyapunov-based feedback control (Reyhanoglu and Hervas, 2013). Because basic mechanical models are approximate models, they necessarily produce model uncertainties and approximated nonlinearities, requiring the use of robust control design. SMC is a sophisticated and efficient robust control method (Decarlo et al., 1988; Utkin, 1977), which provides resilience against matching uncertainties and disturbances and has been used extensively to solve the slosh control problem. Super-twisting control (STC) based on second order sliding mode philosophy was employed in Thakar et al., (2017b) for slosh suppression which reduced the problematic chattering effect of traditional first order SMC.

1.1 Motivation

- Control systems have a deep, comprehensive, and foundational knowledge established over the last six decades, with a significant emphasis on decision-making in uncertain conditions. Despite advances in several subfields of control engineering, considerable work has to be done to effectively address control of complex dynamical systems in the face of rapid environmental change and high degrees of uncertainty.
- One of the main goals of Artificial Intelligence (AI) is to create systems that can learn and think by themselves and can plan activities on their own to complete a given task. We can aspire to build much more efficient and adaptable control systems by exploiting the latest developments and advances in artificial intelligence (AI), such as machine learning (ML), deep learning (DL) and reinforcement learning (RL). To do so, we have to specify particular goals that are currently unattainable using conventional control methodologies but could be reached using AI developments.
- The study and design of control systems have traditionally relied on precise mathematical models of the system with well-understood uncertainties. RL approaches, on the other hand, try to directly learn models and control actions from data and experiments. Clearly, there is little possibility for RL in areas where thorough classical control-oriented models are viable and have already been constructed. However, in areas where such detailed mathematical models do not exist, or performance targets are defined at a high level, or the degree of uncertainty is substantially higher with unverified sources, or the control objectives have high diversity, a much larger opportunity arises.
- Traditional control strategies frequently require the domain expertise of a control engineer for parameter tuning and other operations. For these mathematically challenging tasks, RL can be utilized to find optimum or near optimum solutions.

1.2 Main Contribution

- A model-free, data-driven, and self-learning DRL based STC agent is designed for the first time for slosh container system under diverse operating conditions.
- By interacting with a simulated Python environment of the slosh container system, the proposed agent is trained using Deep Deterministic Policy Gradient (DDPG) algorithm (Lillicrap et al., 2016). The proposed DRL agent learns the controller parameters from scratch to solve the slosh control problem.
- For the slosh control problem, the performance of DRL-based STC is compared with conventional STC (Thakar et al., 2017b). The comparison is based on the second and infinity norms of the control input and the slosh angle.

1.3 Paper Outline

The outline of the paper is as follows: The slosh control problem statement is described in Section 2, along with the slosh container system dynamics. The proposed DRL based STC framework and strategy are discussed in Section 3. The simulation results are compared in Sections 4. Finally, Section 5 draws some conclusions.

2. PROBLEM STATEMENT

When a liquid-filled container is moved from one place to another, the exerted force tends to stimulate the liquid and induce sloshing. The aim of the control task in this paper is to drive a partially filled liquid container in a straight path from its starting position to the destination position with minimal sloshing by a correctly predicted control input.

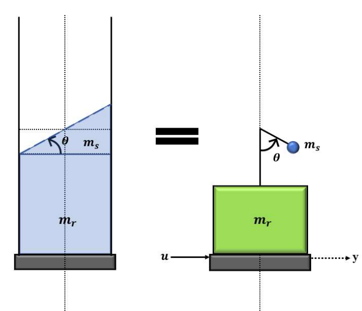


Fig.1. Pendulum analogy for lateral slosh control problem

The lateral slosh with container dynamics model, which was discussed in Bandyopadhyay et al. (2009b) and Kurode, (2009), is revisited in this paper. A simple pendulum mathematical model is employed to model the lateral slosh in a moving container system, as shown in Fig. 1. There are two degrees of freedom in this lateral slosh container system dynamics: displacement of container (y) and pendulum angle or slosh angle (θ). Unlike y, the slosh angle has no direct input, resulting in an underactuated system. As indicated in Fig. 1, the Liquid Container is given a horizontal controlling force u and is anticipated to travel laterally by y distance.

The system's dynamical equations in y and θ, derived using Euler Lagrange's formulation (Bandyopadhyay et al., 2009b), are as follows:

$$M\ddot{y} + m_s l cos\theta\ddot{\theta} - m_s l\dot{\theta}^2 sin\theta = u + d \quad (1)$$
$$m_s l cos\theta\ddot{y} + m_s l^2\ddot{\theta} + c\dot{\theta} + m_s g l sin\theta = 0 \quad (2)$$

Here, M is the total mass of the system in kg, which comprises container base plate mass (m_b), container mass (m_c), slosh mass (m_s) and non-sloshing liquid mass ((m_l). The external disturbance of the input channel is denoted by d. TABLE I contains information on the various system parameters used in the system dynamics model (Bandyopadhyay et al., 2009b).

TABLE I. PARAMETERS FOR THE SLOSH CONTAINER SYSTEM

Parameters	*Value*	*Unit*
Slosh mass (m_s)	1.32	kg
Non-sloshing liquid mass ((m_l)	6.0	kg
Liquid container mass (m_c)	2.0	kg
Base plate mass (m_b)	1.5	kg
Total rigid mass (m_r)	9.7	kg
Pendulum length (l)	0.052126	m
Gravitational acceleration (g)	9.8	m/s^2
Viscous damping coefficient (c)	3.0490e-4	kgm^2/sec

3. CONTROL DEVELOPMENT

For this control problem, we need to design a controller that provides the control force input u, allowing us to move the container to the intended position along a straight path with the least amount of slosh. The control input u for the system dynamics equations (1) and (2) can be developed using DRL based STC to obtain the desired results.

3.1 Super Twisting Control (STC)

For the linear sliding surface and super twisting control discussed in Thakar et al. (2017b), a DRL based parameter tuning approach is proposed in this paper, in order to obtain performance improvement. The dynamics of the error variables for this slosh container system can be represented as

$$\varepsilon_y = y - y_d, \quad \varepsilon_\theta = \theta - \theta_d, \dot{\varepsilon_y} = \dot{y} - \dot{y}_d, \ \dot{\varepsilon_\theta} = \dot{\theta} - \dot{\theta}_d \quad (3)$$

Here subscript 'd' is representing the desired value of the variable. For this problem the desired container position is $y_d = 175$ mm and θ_d and $\dot{\theta}_d$ should be equal to zero for slosh minimization. The error dynamics state vector can be represented as

$$[\delta_1 \quad \delta_2 \quad \delta_3 \quad \delta_4]^T = [\varepsilon_y \quad \dot{\varepsilon_y} \quad \varepsilon_\theta \quad \dot{\varepsilon_\theta}]^T \quad (4)$$

Considering system outputs $\delta_1$and δ_2, the following is an evident linear sliding surface.

$$\rho = c_1\delta_2 + c_2\delta_1 \quad (5)$$

where c_1 and c_2 are sliding surface parameters.

For finite time convergence of system trajectories to a second order sliding set, the super twisting control developed in Thakar et al. (2017b) is used:

$$v = -k_1|\rho|^{\frac{1}{2}} sign(\rho) - \int_0^t k_2 sign(\rho)\, dt \quad (6)$$

Based on the system dynamics and v, the designed control input in Thakar et al. (2017b) is given below:

$$u = (c_1 a_1)^{-1}(v - \tau) \quad (7)$$

where $a_1 = \frac{1}{M - m_s cos^2\delta_3}$ and $\tau = \frac{c_2}{c_1}(\rho - c_2\delta_1)$

Now the goal is to use DRL to fine-tune the STC parameters c_2, k_1, and k_2. Without loss of generality, $c_1 = 1$ can be chosen. In Thakar et al. (2017b), these parameters are designed only satisfying Lyapunov stability condition, and no methodology for tuning these parameters to obtain optimal performance is provided.

3.2 The Basic concept of Reinforcement Learning

RL is a type of ML that studies how AI agents should respond to sequential decision-making problems in a given environment to maximize the cumulative reward (Sutton and Barto, 2018). Along with supervised and unsupervised learning, RL is one of three core ML paradigms. Generally, RL agents start with no prior knowledge of the environment and learn by trial and error, which is slow but effective. The basic RL block diagram is shown in Fig. 2.

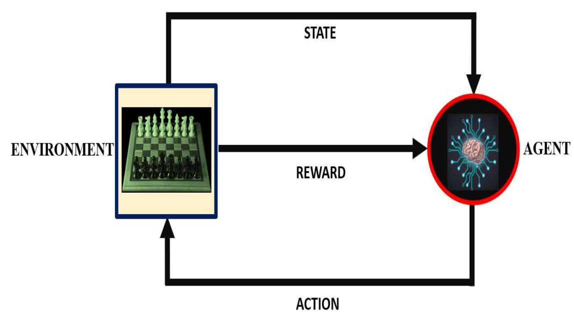


Fig.2. Reinforcement Learning block diagram

In the RL framework, the learner is known as the agent, and the surroundings around the learner are known as the environment. The present condition of the environment is known as the state in RL literature, and RL agents execute actions depending on the state and reward signals. To evaluate how advantageous it is for the agent to be in a specific state or to execute an action in that state, the majority of RL algorithms employ the estimated state value function and the state action value function. In RL, value functions are used to estimate the optimal action selection strategy or policy. These value functions provide a representation of expected future rewards in an indirect manner.

Traditional tabular-based RL algorithms, such as Monte Carlo and Temporal Difference Learning (Sutton and Barto, 2018), have several limitations. The "curse of dimensionality” affects these methods significantly, which means that the computational needs increase exponentially as the number of state or action variables increases. So, we need some more efficient RL algorithms for continuous state and action spaces.

DNN can be used as a function approximator to address this issue by approximating the value functions in DRL methods. DRL algorithms take advantage of DNN's powerful feature extraction capabilities, which eliminates the need for manual feature extraction. For some of the most complicated sequential decision-making problems, these advanced DRL agents can even predict control actions based on raw visual inputs (Mnih et al., 2015). DRL agents have achieved superhuman performance in complex board games such as Go (Silver et al., 2017), Chess, and Shogi.

DRL had been used to create robust autonomous controllers that can learn optimal control actions for complicated control engineering problems from scratch without the need for human expertise (Duan et al., 2020). DRL is currently an active research area, and tremendous progress has been made in terms of advancing the field and applying it in various challenging decision-making domains. We can use DRL to design new architectures for robust, intelligent, and adaptive controllers that can work across a wide range of application domains while enhancing performance and ensuring safety.

3.3 Proposed framework for DRL based STC

When the liquid container is moving in a straight path, the objective of this DRL-based STC agent is to provide effective control action u such that sloshing is minimized. The framework of the proposed DRL-based STC agent is shown in Fig. 3.

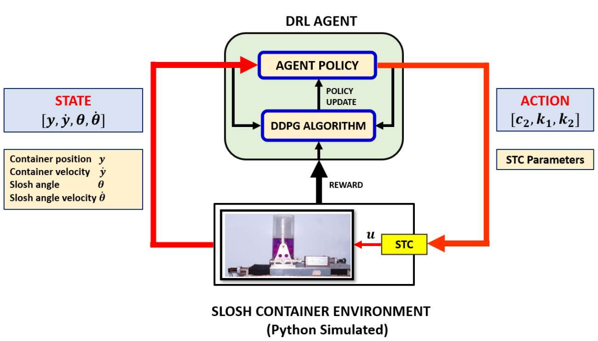


Fig.3. Framework for DRL based STC

This DRL agent was trained using past experience and rewards by interacting with the slosh container system (in a Python simulated environment). In this architecture, the state vector $[y,\ \dot{y},\ \theta,\ \dot{\theta}]$ is used as the state signal for the DRL agent. Based on the initial state, the DRL agent predicts a set of actions which are the controller parameters $[c_2,\ k_1,\ k_2]$ for the STC and calculates the reward function for the entire duration of each simulation episode run of 3 seconds.

The reward function utilized for training the DRL agent is given below.

$$reward\ function = -[4 * |y - y_d| + 0.3 * |\theta - \theta_d| + 0.3|\dot{\theta} - \dot{\theta}_d|] \qquad (8)$$

The above-mentioned reward function and its parameters were designed based on the learning performance of many trial runs.

Using Deep Deterministic Policy Gradient (DDPG) algorithm, this DRL agent was trained to estimate the suboptimal values of controller parameters $[c_2,\ k_1,\ k_2]$. Because the control input u is a function of these controller parameters, the DRL agent is indirectly predicting the sub-optimal values of control action u for slosh minimization.

The DDPG algorithm (Lillicrap et al., 2016), a value-based DRL algorithm, is used to solve our slosh control problem. DDPG is a deterministic policy gradient-based model-free actor-critic (Konda and Tsitsiklis, 1999) approach. The DDPG method expanded the basic idea of the DQN algorithm (Mnih et al., 2015) to continuous state and action spaces, and have solved over 20 simulated physical control problems using high-dimensional sensory input that used the same hyperparameters and DNN architecture (Lillicrap et al., 2016). Below is a description of the DDPG algorithm.

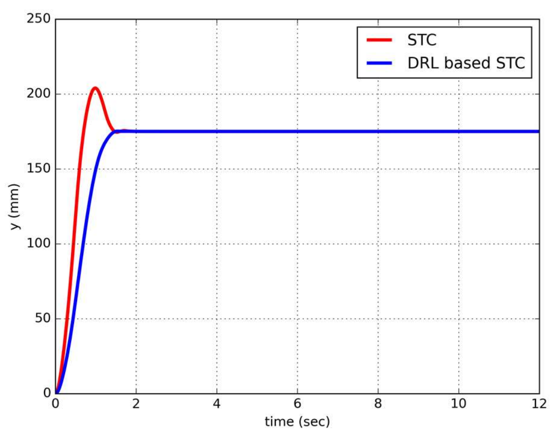
DDPG Algorithm

Initialize: critic $Q(x, u; \omega_Q)$ and actor $\pi(x; \Theta_\pi)$ with random weights ω_Q and Θ_π
: target Q' and π' with weights $\omega_{Q'} \leftarrow \omega_Q, \Theta_{\pi'} \leftarrow \Theta_\pi$
Initialize replay memory $\mathcal{D}$
for all episodes $= 1,2 \ldots\ldots\ldots N$
Initialize a random process $\mathcal{N}$ for action exploration
Receive starting state x_1
For all steps in the episode $k = 1,2 \ldots\ldots\ldots K$
Select action $u_k = \pi(x_k; \Theta_\pi) + \mathcal{N}_k$ according to the current policy and exploration noise
Execute action u_k and observe the reward r_k and state x_{k+1}
Store transition (x_k, u_k, r_k, x_{k+1}) in $\mathcal{D}$
Sample a random minibatch of M transitions (x_i, u_i, r_i, x_{i+1}) from $\mathcal{D}$
Set $h_i = r_i + \gamma Q'(x_{i+1}, \pi'(x_{i+1}; \Theta_{\pi'}); \omega_{Q'})$
Update critic by minimizing the loss: $L = \frac{1}{M}\sum_i \left(h_i - Q(x_i, u_i; \omega_Q)\right)^2$
Update the actor policy using the sampled policy gradient:
$\nabla_{\Theta^\pi} J \approx \frac{1}{M}\sum_i \nabla_u Q(x, u; \omega_Q)\big|_{x=x_i, u=\pi(x_i)} \nabla_{\Theta^\pi}\pi(x; \Theta_\pi)|_{x_i}$
Update target networks:
$\omega_{Q'} \leftarrow \tau\omega_Q + (1-\tau)\omega_{Q'}$
$\Theta_{\pi'} \leftarrow \tau\Theta_\pi + (1-\tau)\Theta_{\pi'}$
end for
end for

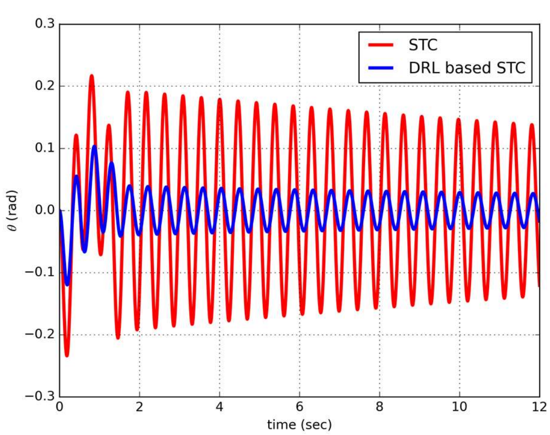
4. SIMULATION RESULTS

A slosh container environment is simulated in Python using the mathematical model provided in Thakar et al. (2017b) for slosh dynamics in straight line motion of liquid container. For improved learning, the initial states of the DRL agent were chosen at random from a practical range of all the state variables for each simulation episode. Therefore, the suggested learning-based controller is expected to be robust against variations in operating conditions due to disturbances or uncertainties. An NVIDIA GeForce GTX 1080 machine with 12GB of RAM and 16 CPU cores is used for the entire processing. The chosen values of STC parameters in Thakar et al. (2017b) are $c_2 = 7,\ k_1 = 1.77$ and $k_2 = 4.08$. Our DRL agent was tested for STC parameter predictions for 10 test runs after 100 thousand training episodes. The predicted values of the STC parameters by the DRL agent were nearly identical for all test runs. The largest percentage change discovered in the DRL agent's predictions is 3.94 %. The DRL agent's performance in terms of slosh angle and control effort was quite satisfactory after 50 thousand training episodes, however the largest percentage change in the STC parameters was substantially larger. Based on reward function given in (8) the best possible values of the STC parameters after 100 thousand training episodes were obtained as $c_2 = 2.8872,\ k_1 = 1.1993$ and $k_2 = 0.4696$. To replicate the steady container in the beginning of the simulation, the initial values of all state variables $y,\ \dot{y},\ \theta$ and $\dot{\theta}$ are fixed at zero for all test runs.

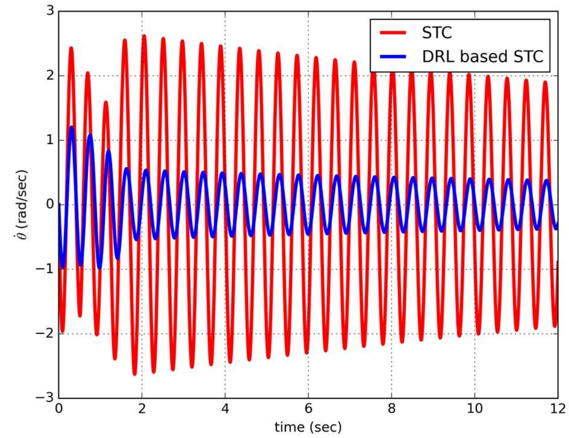
Simulation comparison results for our proposed DRL-based STC and STC discussed in Thakar et al. (2017b) are shown in Figs. 4a-d. TABLE II compares the required control effort and slosh suppression results for both approaches. All these results are compared for a 12-second test run. Our DRL-based STC outperforms the conventional STC design in terms of control effort requirements ($\|u\|_2$ and $\|u\|_\infty$), as shown in Table II. In comparison to the approach presented in Thakar et al. (2017b), the values of $\|\theta\|_2$ and $\|\theta\|_\infty$ also show that it regulates sloshing in a better way.



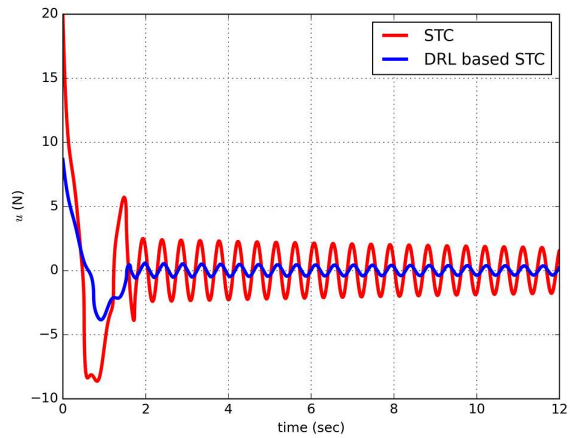
(a)



(b)



(c)



(d)

Fig.4. Simulation results (a) Slosh container position y, (b) Slosh angle θ, (c) Slosh angle velocity $\dot{\theta}$, (d) Control effort u

TABLE II. RESULTS COMPARISON

Control	$\|\theta\|_2$	$\|\theta\|_\infty$	$\|u\|_2$	$\|u\|_\infty$
STC, Thakar et al. (2017b)	40.59	0.22	1025.40	19.97
DRL based STC	10.41	0.10	426.14	8.66

For the sake of performance evaluation, the DRL agent's training was stopped. Otherwise, it could continue to improve its learning by adjusting the DNN weights parameters during continuous on-line execution to adapt to changing operational conditions of the environment.

5. CONCLUSIONS AND DISCUSSION

This paper proposes for the first time a DRL-based super twisting controller that can optimize its parameters on its own, without the need of human intervention. This DRL-based architecture is designed to execute control actions automatically to ensure container movement with minimal slosh. When compared to the existing STC method, the suggested framework performs better in a Python simulated environment. When the environment conditions change, the controller can readjust the control actions to optimize total reward. The proposed DRL-based STC framework offers a huge possibility to go beyond present adaptive control frameworks and paradigms, achieving considerably higher degrees of performance.

References

Abramson, H. N. (1966). The dynamic behaviour of liquids in moving containers with applications to space vehicle technology, *National Aeronautics and Space Administration Tech. Rep. NASA*, Washington, D.C. USA, SP-106.

Bandyopadhyay, B., Kurode, S. and Gandhi, P. S. (2009a). Sliding mode control for slosh-free motion-A class of underactuated system, *International Journal of Advanced Mechatronic Systems*, 1, 203–213.

Bandyopadhyay, B., Gandhi, P. S. and Kurode, S. (2009b). Sliding mode observer based sliding mode controller for slosh-free motion through PID scheme, *IEEE Transactions on Industrial Electronics*, 56(9), 3432–3442.

Baozeng, Y., Wenjun, W. and Yulong, Y. (2016). Modeling and coupling dynamics of the spacecraft with multiple propellant tanks, *AIAA*, 54, 3608–3618.

Decarlo, R. A., Zak, S. H. and Matthews, G. P. (1988). Variable Structure Control of Nonlinear Multivariable Systems: A Tutorial, *Proc. IEEE*, 76, 212–232.

Duan, J., Shi, D., Diao, R., Li, H., Wang, Z., Zhang, B., Bian, D. and Yi, Z. (2020). Deep-Reinforcement-Learning-Based Autonomous Voltage Control for Power Grid Operations, *IEEE Transactions on Power Systems*, 35(1), 814–817.

Gandhi, P. S. and Duggal, A. (2009). Active stabilization of lateral and rotary slosh in cylindrical tanks, in *IEEE International Conf. ICIT*,1–6.

Ibrahim, R. A., Pilipchuk, V. N. and Ikeda, T. (2001). Recent advances in liquid sloshing dynamics, *Applied Mechanics Reviews*, 54(2), 133–199.

Konda, V. R. and Tsitsiklis, J. N. (1999). Actor-critic algorithms, in *Advances in Neural Information Processing Systems*,12, 1008–1014.

Kurode, S. (2009). *Sliding Mode Control for slosh-free motion using state and output feedback: An observer based and fast output sampling approach*, PhD Thesis, IIT Bombay.

Kurode, S., Spurgeon, S. K., Bandyopadhyay, B. and Gandhi, P. S. (2013). Sliding mode control for slosh-free motion using a nonlinear sliding surface, *IEEE/ASME Trans. Mechatronics*, 18(2),714–724.

Peterson, L. D., Crawley, E. F. and R. J. H. (1989). Nonlinear fluid slosh coupled to the dynamics of a spacecraft, *AIAA Journal*, 27(9), 1230–1240.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N. Erez, T., Tassa, Y., Silver, D. and Wierstra, D. (2016). Continuous control with deep reinforcement learning, *4th International Conference on Learning Representations*, ICLR 2016.

Spong, M.W. (1998). Underactuated Mechanical Systems, *Control Problems in Robotics and Automation*, Springer, Berlin, Heidelberg, 230,135–150.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A., Veness, J., Bellemare, M., Graves, A., Riedmiller, M., Fidjeland, A., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. and Hassabis, D. (2015). Human-level control through deep reinforcement learning, *Nature*, 518, 529–533.

Nichkawde, C., Harish, P. M. and Ananthkrishnan, N. (2004). Stability analysis of a multibody system model for coupled slosh-vehicle dynamics, *Journal of Sound and Vibration*, 275(3), 1069–1083.

Petit, N. and Rouchon, P. (2002). Dynamics and solutions to some control problems for water-tank systems, *IEEE Trans. Automatic Control*, 47, 594–609.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning An Introduction*, Cambridge, MA, USA: MIT Press.

Reyhanoglu, M. and Hervas, J. R. (2013). Robotically controlled sloshing suppression in point-to-point liquid container transfer, *J. Vib. Control*, 19(14), 2137–2144.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., Driessche, G. van den, Graepel, T. and Hassabis. D. (2017). Mastering the game of Go without human knowledge, *Nature*, 550(7676), 354–359.

Sira-Ramírez, H. and Fliess, M. (2002). A flatness based generalized PI control approach to liquid sloshing regulation in a moving container, *Proceedings of the American Control Conference*, 4(February 2002), 2909–2914.

De Souza, A. G. and De Souza, L. C. G. (2014). Satellite Attitude Control System Design Taking into Account the Fuel Slosh and Flexible Dynamics, *Mathematical Problems in Engineering*, 2014, 1–8.

Thakar, P. S., Bandyopadhyay, B., Gandhi, P. S. and Kurode, S. (2012). Robust control of rotary slosh using integral sliding modes, in *12th International Workshop on Variable Structure Systems*, Mumbai, India, 440–445.

Thakar, P. S., Trivedi, P. K., Bandyopadhyay, B. and Gandhi, P. S. (2017a). A New Nonlinear Control for Asymptotic Stabilization of a Class of Underactuated Systems: An Implementation to Slosh-Container Problem, *IEEE/ASME Transactions on Mechatronics*, 22(2), 1082–1092.

Thakar, P. S., Bandyopadhyay, B. and Gandhi, P. (2017b). Improved output-feedback second order sliding mode control design with implementation for underactuated slosh-container system having confined track length, *IET Control Theory and Applications*, 11(8), 1316–1323.

Thakar, P. S., Bandyopadhyay, B. and Gandhi, P. S. (2013). Sliding mode control for an underactuated slosh-container system using non-linear model, *International Journal of Advanced Mechatronic Systems*, 5(5), 335–344.

Utkin, V. I. (1977). Variable Structure System with Sliding Modes, *IEEE Transactions on Automatic Control*, AC-22, 212–222.

Vreeburg, J. P. (2005). Spacecraft maneuvers and slosh control, *IEEE Control Systems*, 25, 12–16.

Weerdt, E., Kampen, E., Gemert, D., Chu, Q. P. and Mulder, J. A. (2008). Adaptive nonlinear dynamic inversion for spacecraft attitude control with fuel sloshing, in *AIAA Guidance, Navigation and Control Conference and Exhibit*, 1–24.

Yano, K. and Terashima, K. (2001). Robust liquid container transfer control for complete sloshing suppression, *IEEE Trans. Control Systems Technology*, 9, 483–493.