International Journal of Intelligent Robotics and Applications
https://doi.org/10.1007/s41315-024-00412-4

REGULAR PAPER

Experimental evaluation of commercial quadruped robots: stability and performance in non-inertial environments

Stephen Misenti[1] · Brendan Hertel[1] · Bowen Weng[2] · Ryan Donald[1] · Advait Jawaji[3] · Magnus-Tryggvi Kosoko-Thoroddsen[3] · J. Gregory Trafton[4] · Adam Norton[1] · Reza Azadeh[1] · Yan Gu[3]

Received: 16 September 2024 / Accepted: 11 December 2024

Abstract

This paper presents an experimental evaluation of the stability and performance of commercial quadrupedal robots, specifically the Ghost Robotics Vision 60 and Boston Dynamics Spot, under dynamic ground conditions typical of non-inertial naval environments. Our study systematically assesses these robots in controlled laboratory settings and real-world scenarios aboard the M80 Stiletto, a naval prototype vessel. To rigorously test the robots' stability, balance, and path-following capabilities, we analyze primary metrics including the distance of the center of mass to the support polygon boundary, body-position tracking accuracy, trunk orientation steadiness, and joint torque profiles under dynamic disturbances from ground accelerations and rotations. The results reveal that while both robots demonstrate operational capabilities, significant challenges remain. Vision 60, in particular, exhibits superior stability, balance, and lower peak torque when handling substantial ground motions compared to Spot. However, both robots struggle with accurate body-position tracking under aggressive ground motions. These findings highlight the limitations of current commercial quadruped robots in operating within non-inertial environments and underscore the urgent requirement for further research and development in robot planning and control.

Keywords Legged robots · Dynamic environments · Performance evaluation

✉ Yan Gu
yangu@purdue.edu

Stephen Misenti
stephen_misenti@student.uml.edu

Brendan Hertel
brendan_hertel@student.uml.edu

Bowen Weng
bweng@iastate.edu

Ryan Donald
ryan_donald@student.uml.edu

Advait Jawaji
ajawaji@purdue.edu

Magnus-Tryggvi Kosoko-Thoroddsen
mkosokot@purdue.edu

J. Gregory Trafton
greg.j.trafton.civ@us.navy.mil

Adam Norton
adam_norton@uml.edu

Reza Azadeh
reza@cs.uml.edu

[1] University of Massachusetts Lowell, One University Avenue, Lowell, MA 01854, USA

[2] Iowa State University, 2434 Osborn Drive, Ames, IA 50011, USA

[3] Purdue University, 585 Purdue Mall, West Lafayette, IN 47907, USA

[4] Naval Research Laboratory, 4555 Overlook Avenue, SW, Washington, DC 20375, USA

Published online: 07 February 2025

1 Introduction

1.1 Related work

(1) Robot testing and benchmarking: Legged robots are increasingly recognized for their potential in executing a broad spectrum of real-world tasks such as warehouse automation, home assistance, package delivery, search and rescue, and inspection and maintenance. The stability and performance of legged robots in dynamic environments are crucial to ensuring their operational efficacy and safety.

Stability in this context refers to a legged robot's ability to sustain intended movements-such as walking, standing, and object manipulation-without falling over (Westervelt et al. 2007; Zhao and Gu 2020). Additionally, an essential performance measure for legged robots is robustness, defined as the ability to maintain planned behaviors amidst environmental uncertainties (Xiong and Ames 2022; Dai et al. 2022; Gibson et al. 2022; Caron et al. 2019). Recent years have seen a surge in interest towards systematically evaluating the effectiveness of legged robotic systems, especially quadrupeds. This highlights the growing need to develop and refine metrics and methodologies that accurately assess the capabilities and limitations of these advanced robotic platforms in complex, real-world scenarios.

To evaluate the stability and performance of control and planning approaches for legged robots, various validation methods have been utilized. Techniques such as hand-pushing, foot-kicking, and the application of forces through means including rope-dragging and hockey sticks poking have been common (Radosavovic et al. 2023). Notably, Chen et al. highlighted the use of a yoga ball and swing-ball impacts as dynamic tools to simulate sudden interactions with moving objects, offering a realistic spectrum of challenges a robot might encounter in operational settings (Chen et al. 2023).

Despite these diverse testing methods, there remains a pressing need for more standardized approaches to systematically and comprehensively assess the stability and performance of legged robots. Recent studies, such as those by Weng et al., have begun addressing this gap. They employed a linear impactor to standardize disturbance rejection tests for legged robot locomotion, establishing a preliminary framework for standardized testing protocols (Weng et al. 2024, 2023, 2022).

Performance metrics for legged robots have significantly advanced, incorporating methodologies that quantitatively assess their capabilities. Aller et al. (2019) have expanded the state of the art in humanoid robot locomotion metrics, enhancing how these measurements contribute to improved design and control strategies.

Adversarial disturbance rejection tests show considerable variability, ranging from standardized tests employing mechanical apparatus to more unconventional methods that pose unexpected challenges to robots. Shi et al. (2024) conducted a comprehensive analysis of adversarial attacks on learning-based controllers of quadrupedal locomotion.

The role of benchmarking in robotic research cannot be overstated, as various initiatives strive to establish performance standards across different platforms (Norton et al. 2022, 2023). Torricelli and Pons (2019) explored the Eurobench framework, designed to equip robots for real-world functionalities by providing standardized testbeds that mimic common human environments. Similarly, Castano et al. (2022) examined dynamic balancing controllers for humanoid robots, providing insights into the effectiveness of these controllers under various simulated conditions.

(2) Robot locomotion in non-inertial environments: Legged locomotion in non-inertial environments-settings that experience acceleration or rotation in the inertial frame-is increasingly recognized as a critical capability for robotic systems. Common examples of such environments include moving ships, airplanes, and ground vehicles. Legged robots that can navigate these environments reliably can enhance real-world applications such as firefighting, inspection, and maintenance on moving public transportation vehicles, ships, airplanes, and oil platforms.

A defining feature of these non-inertial environments is the temporally persistent and continuous acceleration or rotation of the ground, often described as dynamic rigid surfaces (DRS) (Iqbal 2023). The time-varying nature of ground motions leads to fundamentally different dynamics for legged locomotion compared to static surfaces, creating a substantial open problem of legged robot control as existing control approaches typically assume static grounds (Westervelt et al. 2007; Zhao and Gu 2020).

In response to these challenges, recent research has focused on improving the stability and robustness of robots during such ground movements. Efforts include dynamic modeling (Iqbal et al. 2023; Iqbal and Gu 2021), state estimation (Gao et al. 2022; He et al. 2024), real-time motion generation (Iqbal et al. 2023; Gao et al. 2023), robust controller design (Iqbal 2023; Gao 2023; Donald et al. 2024), and stability analysis (Iqbal et al. 2020). These studies have treated surface movements explicitly to adapt legged locomotion control approaches to DRS within non-inertial environments.

Despite these advances, a substantial gap remains in the formal benchmarking and evaluation of these control systems in non-inertial environments. While individual research efforts have demonstrated efficacy on isolated robots and ground motions, there is a lack of standardized testing frameworks or evaluation protocols that can comprehensively assess and compare the performance of different

control strategies under a consistent set of dynamic ground conditions. This absence of benchmarking standards means that while theoretical and experimental advancements continue to evolve, their practical applications and comparisons across different platforms and environments cannot be systematically validated. Establishing robust benchmarking protocols is essential for progressing towards generally applicable solutions that ensure the stability and satisfactory performance of legged robots in non-inertial environments.

1.2 Contributions

This study represents one of the earliest systematic experimental evaluations of the stability and performance of commercial quadrupedal robots in non-inertial environments. The primary contributions of this research are outlined below:

(a) We develop and implement a comprehensive testing framework that assesses quadrupedal robots under a variety of dynamic ground conditions, both in laboratory simulations and real-world naval scenarios. This framework encompasses typical robotic movements and ground motions designed to replicate diverse sea wave conditions and ship dynamics, providing a thorough evaluation of the robots' capabilities and limitations.
(b) This research generates both quantitative and qualitative data on key performance metrics such as stability, the center of mass (CoM) distance to the support polygon boundary, joint torque profiles, and body-position trajectory accuracy during ground movements. This analysis reveals the operational challenges that quadruped robots face in maintaining stability and performing locomotion tasks on moving platforms. The dataset from this study has been made publicly available for broader use and can be accessed online.[1]
(c) By testing two leading models of commercial quadrupedal robots-the Ghost Robotics Vision 60 and the Boston Dynamics Spot-this study gives insights into their relative strengths and weaknesses. Such comparative analysis aids in deeper understanding of each robot's suitability for specific maritime tasks.
(d) The results from our experiments identify significant areas where current commercial quadrupedal robots underperform, especially in managing the continuous and dynamic ground motions encountered in non-inertial environments such as those onboard ships.

2 System description

The experimental setup for this study included several critical components. Field tests were conducted aboard the M80 Stiletto, a prototype naval vessel utilized by the Office of Naval Research, which was docked at a naval station on the U.S. eastern coast. For controlled laboratory experiments, a Motek M-Gait treadmill was used to simulate ship motions, capable of executing pre-programmed pitching and swaying movements. The experiments involved two commercially available quadrupedal robots, each equipped with a robotic arm. The subsequent sections provide a detailed overview of the hardware systems employed in our experiments.

2.1 Robots

As depicted in Fig. 1, the Boston Dynamics' Spot (Fig. 1a) and Ghost Robotics' Vision 60 (Fig. 1b) were selected for experimentation. Spot was chosen for its robust balance and mobility for static terrains, as well as for its user-friendly capabilities. Vision 60 was selected for its military orientation and payload capacities. Figure 2 illustrates the configuration of Spot's leg joints; Vision 60 employs a similar naming convention for its joints. To ensure safety during in-lab experiments, custom safety harnesses were designed and utilized for each robot, as shown in Fig. 1.

(1) Ghost Robotics Vision 60: The Ghost Robotics Vision 60 is a semi-autonomous Quadrupedal Unmanned Ground Vehicle (Q-UGV) designed for defense, military, and commercial use. Table 1 outlines the key specifications advertised for the Vision 60 system. The robot's modular design enables the addition of various payloads, including robotic arms, sensors, and communication devices. In our experiments, a Vision 60 version 5.1C was employed, equipped with a 7-DoF Kinova Gen2 robotic manipulator that features a three-finger KG-3 gripper. The manipulator was securely mounted using a custom-designed aluminum plate.

(2) Boston Dynamics Spot: Boston Dynamics Spot is a versatile quadrupedal unmanned ground vehicle (Q-UGV) that has been utilized across various industries for tasks including inspection, surveying, and hazard detection. For our experiments, Spot was equipped with the Spot Arm, a 6-DoF robotic arm capable of lifting up to 11 kg and dragging as much as 25 kg.

Table 2 shows the advertised key specifications of Spot (Boston Dynamics 2023). Notably, the mass and inertial properties of Spot have not been disclosed by Boston Dynamics, making these parameters inaccessible without extensive testing (G. 2023).

[1] https://github.com/purdue-tracelab/quadruped_assessment.

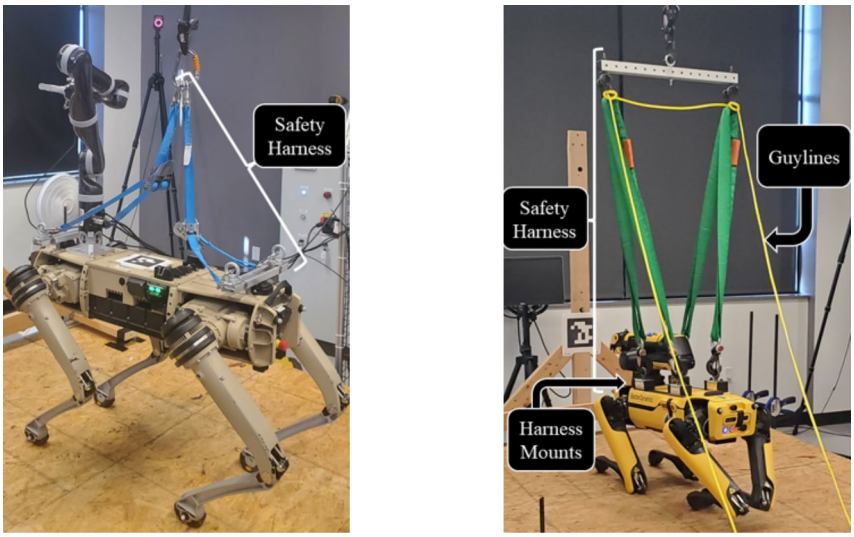
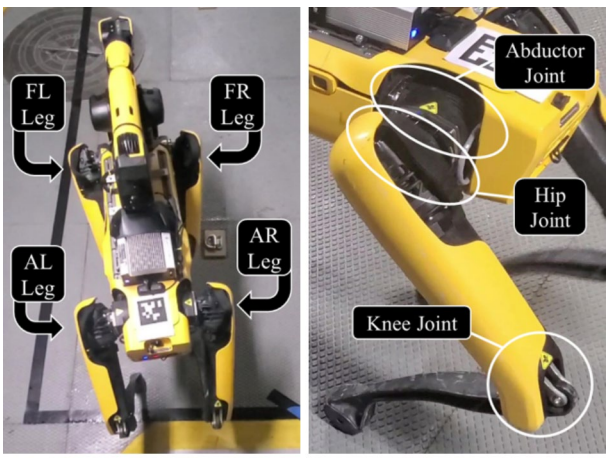


Fig. 1 Vision 60 and Spot in their safety harnesses on the wooden treadmill platform

(a) Vision 60 in harness.

(b) Spot in harness.



(a) Leg locations.

(b) Leg joints.

Fig. 2 Illustration of Spot's leg joints. In subplot **a**, the "FL," "FR," "AL," and "AR" legs are the front-left, front-right, aft-left, and aft-right legs, respectively. All three joints of each leg in subplot **b** are revolute and directly actuated. This illustration also applies to Vision 60's leg joints

Table 1 Key features of Ghost Robotics Vision 60

Features	
Weight	51 kg
Ingress, Temp.	IP67, – 45 to 55 °C
Top speed	3 m/s
Max. distance	10 km
Max. power runtime	3 hr
CPU/GPU	NVIDIA Xavier
Max. payload	10 kg

Table 2 Key specifications of Boston Dynamics Spot

Features	Specifications	Features	Specifications
Net mass	32.7 kg	Battery capacity	564 Wh
Ingress, Temp	IP54, -20–55 °C	Average runtime	90 min
Max. speed	1.6 m/s	Standby time	180 min
Max. slope	± 30°	Recharge time	60 min
Max. step height	300 mm	Max payload mass	14 kg

2.2 Dynamic ground simulation

(1) Dynamic treadmill: To simulate the accelerating rigid ground (i.e., DRS) commonly found in real-world non-inertial environments such as naval ships rocked by waves or airplanes experiencing turbulence, we utilized a Motek M-Gait modular treadmill (B.V. 2024). This treadmill is capable of simulating rotary ground motions, and is housed at the New England Robotics Validation and Experimentation (NERVE) Center at the University of Massachusetts Lowell (UML).

The treadmill can execute pre-programmed pitch and sway movements at varying angles and speeds. Specifically, it can achieve pitch angles up to ± 10° and sway movements up to ±0.05 m, offering a controlled environment to assess the stability and performance of quadruped robots. Further specifications of the treadmill are detailed in Table 3 (Brennan 2023).

To accommodate the robots' movements and ensure safety, the treadmill's handrails were removed, and a custom wooden platform was designed and attached by NERVE center

Table 3 Specifications of the Motek M-Gait treadmill

Features	Specifications
Walking surface size (L × W)	200 cm × 100 cm
Base size (L× W)	225 cm × 182 cm
Max. speed	18 km/hr
Max. acceleration	3 m/s^2
Load capacity	5000 N

engineers. Additionally, a safety cage was installed to separate the operators from the treadmill during all in-lab experiments. Figure 3 illustrates the treadmill in three configurations: without the custom wooden platform, with the platform, and with both the platform and the safety cage installed.

(2) Moving vessel: Field data were collected aboard the M80 Stiletto, a Navy prototype vessel docked at a naval station on the U.S. eastern coast (Fig. 4). This vessel is sometimes utilized by naval engineers, researchers, and scientists as a versatile platform for real-world data collection.

3 Experimental setup

This section details the experimental setup for both the controlled in-lab experiments and the field tests conducted with the Vision 60 and Boston Dynamics Spot robots. These experiments encompass various types of robot and ground movements to assess the robots' performance and stability under different conditions.

3.1 In-lab experimental setup

The controlled in-lab experiments were carried out at the NERVE Center at UML. These experiments aimed to evaluate the robots' performance under repeatable and precisely controlled conditions. The setup involved the Motek M-Gait treadmill, which simulated dynamic ground conditions akin to those experienced on naval vessels.

For safety and consistency across all experiment trials, the treadmill was outfitted with a custom wooden platform and surrounded by a safety cage, as depicted in Fig. 3.

We conducted multiple locomotion tasks with both the Vision 60 and Boston Dynamics Spot robots to test their performance under different programmed treadmill motions, summarized as follows:

(1) Treadmill motion types: To thoroughly evaluate the robots' performance, three distinct treadmill motion profiles were utilized, each consisting of combined pitch and sway movements. "Treadmill Motion 1" and "Treadmill Motion 2" are characterized by single sinusoidal waveforms. In contrast, "Treadmill Motion 3" consists of a composite of sinusoidal time functions derived from field experiments.

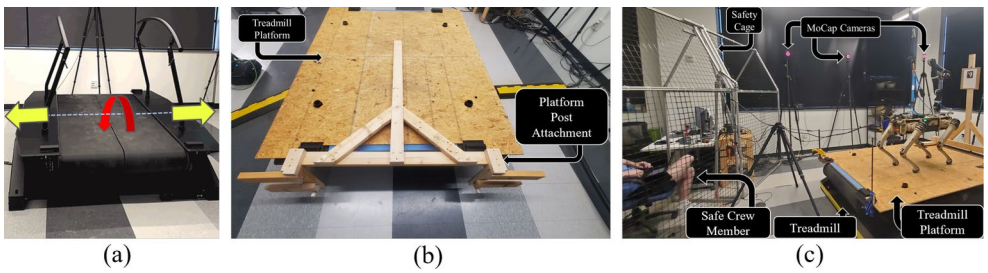


Fig. 3 Motek M-Gait treadmill: **a** without the customized treadmill platform; **b** with the treadmill platform installed; and **c** with the customized safety cage for operator safety protection. In subplot **a**, the yellow arrows show the treadmill's sway direction, while the red arrow illustrates the pitch direction about the rotation axis (i.e., the dashed line) (color figure online)

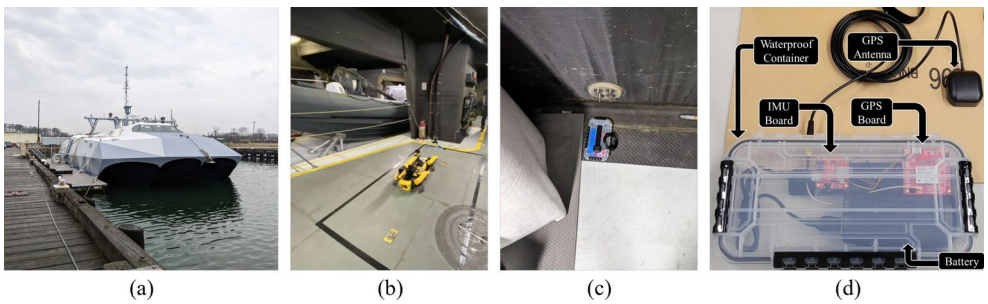


Fig. 4 Illustration of the M80 Stiletto vessel used for the field tests. **a** Stiletto docked at a naval station during the day of testing as captured by Paul Robinette (UML) with permission from NRL. **b** Rectangular safe walking region inside the ship, defined by the black tapes on the ship's floor. **c** Customized sensor box rigidly attached to the ship. **d** Components of the sensor box

The mathematical expressions for these motion profiles are provided below, specifying the "Pitch" angles in degrees, "Sway" displacements in meters, motion frequencies in radians per second, and time t in seconds:

(i) "Treadmill Motion 1": Pitch $= 3\sin(1.5t)$ and Sway $= 0.05\sin(2t)$.
(ii) "Treadmill Motion 2": Pitch $= 8\sin(2.4t)$ and Sway $= 0.05\sin(2.4t)$.
(iii) "Treadmill Motion 3":

$$\begin{aligned}\text{Pitch} = \frac{10}{0.054574}\Big(&0.01773\sin(1.891 + 0.547t)\\ &+ 0.00903\sin(1.293 + 0.859t)\\ &+ 0.00845\sin(2.568 + 1.016t)\\ &+ 0.00760\sin(2.399 + 1.328t)\\ &+ 0.00572\sin(1.904 + 1.641t)\\ &+ 0.00489\sin(2.249 + 2.031t)\\ &+ 0.00411\sin(2.635 + 2.422t)\\ &+ 0.00401\sin(2.742 + 2.812t)\\ &+ 0.00328\sin(2.285 + 3.125t)\\ &+ 0.00350\sin(2.763 + 3.516t)\\ &+ 0.00335\sin(2.900 + 3.906t)\\ &+ 0.00294\sin(-2.995 + 4.297t)\\ &+ 0.00310\sin(2.846 + 4.609t)\\ &+ 0.00239\sin(2.452 + 4.922t)\Big)\end{aligned}$$

and

$$\begin{aligned}\text{Sway} = \frac{-0.05}{0.077063}\Big(&0.02907\cos(-1.488 + 0.547t)\\ &+ 0.01623\cos(-1.171 + 0.938t)\\ &+ 0.01350\cos(-0.888 + 1.328t)\\ &+ 0.00987\cos(-1.285 + 1.641t)\\ &+ 0.00729\cos(-1.155 + 2.031t)\\ &+ 0.00700\cos(-0.700 + 2.422t)\\ &+ 0.00631\cos(-0.417 + 2.812t)\\ &+ 0.00553\cos(-0.713 + 3.125t)\\ &+ 0.00543\cos(-0.504 + 3.516t)\\ &+ 0.00529\cos(-0.270 + 3.906t)\\ &+ 0.00476\cos(-0.547 + 4.219t)\\ &+ 0.00484\cos(-0.293 + 4.609t)\\ &+ 0.00390\cos(-0.559 + 4.922t)\Big).\end{aligned}$$

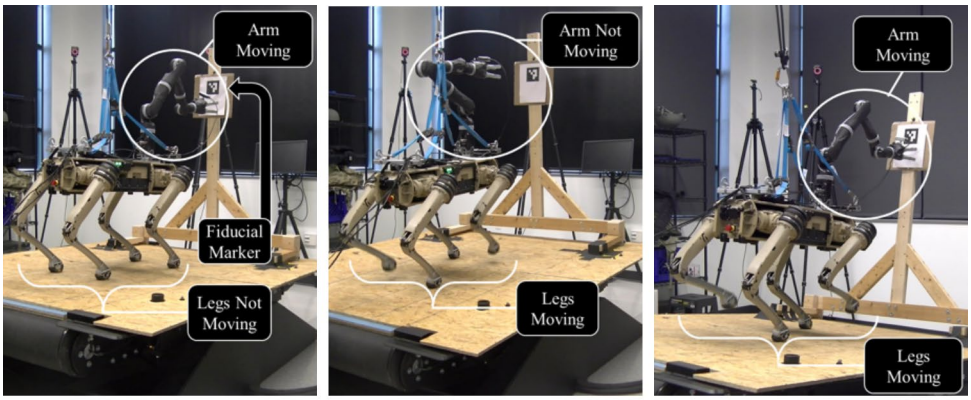
To derive the "Treadmill Motion 3" from field experiments, we used the acceleration data returned by the IMU attached to the ship (see Sec. 3.2) for details about sensor setup during field tests). The cumulative trapezoidal integration was used on the acceleration data to obtain the displacement data. Since only the pitch and sway directions of the treadmill accelerations can be directly pre-programmed, the motion data of the horizontal directions are used to form an equation for the swaying motion, while the difference between the vertical displacements of the two IMUs attached to the front and the back of the ship was used to calculate the pitch equation. Then, by applying fast fourier transform, we identified the amplitudes and frequencies of the components of the displacement data. We selected the components with the highest amplitude peaks to form a filtered ship motion profile. Finally, due to the physical limits of the treadmill motion, we normalized and shifted the filtered ship motion profile to obtain "Treadmill Motion 3" that respects the movement limits of the treadmill.

To highlight the influence of ground motion on the robots' stability and performance, the "Baseline Treadmill Motion" was also tested. This motion profile was trivial, corresponding to the stationary treadmill, and served as a control scenario to benchmark the effects of dynamic ground movements on the robots.

(2) Robot movement types: The robots were tasked with different movement activities, as illustrated in Fig. 5. These tasks were designed to assess the robots' stability and functional capabilities under different conditions.

Experiments on both the Spot and Vision 60 robots were systematically conducted across four sets of experiments. Each set included trials under three different treadmill motions, with the exception of the baseline set, which involved no treadmill movement. The four sets of experiments conducted are labeled as: "Baseline," "Just Body DRS," "Just Arm DRS," and "Both DRS." The "DRS" label in each set denotes the incorporation of DRS motions to challenge the robots' stability and performance. Below is a summary of these robot motion types:

(i) "Baseline": This set had the quadruped robots walking in place without any treadmill motion, serving as a baseline for comparison.
(ii) "Just Arm DRS": In this set, the robots used their arms to track a fiducial marker attached to the treadmill platform post (illustrated in Fig. 5), while their legs made minimal joint motions necessary to maintain balance.
(iii) "Just Body DRS": In contrast to the "Just Arm DRS," this set focused on the robots' legs, which were actively moving to maintain the desired position and orientation of the robot's body, while the arm joints remained mostly stationary.
(iv) "Both DRS": This comprehensive set involved both the robotic arms and legs. The arms were engaged



(a) "Just Arm DRS" (b) "Just Body DRS" (c) "Both DRS"

Fig. 5 Illustrations of three robot movement types during controlled in-lab experiments

in tracking a fiducial mark, while the legs worked to maintain a steady pose of the robot's trunk.

(3) Sensor setup: An optical motion capture system was employed to accurately measure the three-dimensional (3-D) positions of reflective markers attached to selected points on the robots and the treadmill. A total of 14 markers were utilized in each trial. These included a marker on each corner of the treadmill platform, one near the fiducial marker, one on the wrist of the end-effector, two on the front corners of the robot trunks, two on the rear corners, and one on the outer side of each of the four toes. These specific marker placements facilitated the precise determination of the ground-truth positions and orientations of the robots and treadmill within the inertial motion capture frame.

To capture these marker positions, eight infrared Motion Analysis Kestrel 4200 cameras were deployed (M. A. Inc. 2024a). The recorded data were then post-processed using Motion Analysis' Cortex software (M. A. Inc. 2024b).

3.2 Field experiment setup

With collaboration from researchers at UML's NERVE Center, Navy Research Laboratory (NRL), Naval Undersea Warfare Center, Naval Surface Warfare Center, and Purdue University, the field experiments for this study were conducted aboard the M80 Stiletto vessel at a U.S. naval station.

For the field experiments, Spot was provided by the NRL. We chose to test Spot alone without Vision 60 in the shipboard environments for two primary reasons. Firstly, our field tests on the M80 Stiletto were constrained by a limited time window due to the financial and personnel resources required, as well as weather conditions. Secondly, as this was, to our best knowledge, one of the first systematic evaluations involving a legged robot on a ship exposed to sea waves, we believed that focusing on collecting ample and unique data from one of the world's most advanced legged robot platforms would be both valuable and impactful. Future work will focus on testing multiple robot platforms in shipboard environments.

The testing area on the ship was clearly marked with black tapes to define boundaries, as illustrated in Fig. 4b. During the tests, Spot utilized its proprietary control system, optimized for enhanced mobility in challenging environments. Additionally, the "Moving Platform Mode," a feature developed by the NRL to improve stability on non-inertial surfaces, was activated to enhance Spot's capabilities of handling dynamic ship-like movements.

A range of robot and ship motions were tested to thoroughly evaluate Spot's performance under dynamic conditions. An experienced joystick operator from the NRL provided high-level movement commands to Spot, such as moving forward, backward, turning, and adjusting speed. These commands enabled the operator to assist Spot in maintaining balance and accurately following its intended trajectory.

(1) Sea state and ship motion types: The field experiments conducted aboard the ship utilized motion profiles based on Sea State Code 1 conditions (0~10 cm wave height), reflecting relatively calm seas with noticeable wave motions. The following ship motion profiles were rigorously tested:

(i) "At Pier": This scenario represents the calmest condition, with the ship docked at the pier and engines powered off. Minor waves impacted the hull, but the overall movement of the ship was minimal. These tests serve as a baseline to compare robot performance in a stable setting.
(ii) "Underway": This profile describes the ship leaving the pier and advancing into the sea at approximately 10 knots (18.5 km/h). The movement included moderate hull swaying and pitching due to low-amplitude waves.

(iii) "Sea State": Upon reaching a speed of 20 knots (37 km/h) in open waters, the wave motion intensified. This condition simulated moderate sea motion, causing more pronounced pitching and rolling as the ship navigated through larger waves.
(iv) "S-Curves": In this intense scenario, the ship executed S-curve maneuvers at up to 20 knots (37 km/h), inducing aggressive lateral swaying. These maneuvers mimicked the dynamic effects of a ship executing sharp turns at high speeds, challenging the robots with sudden shifts in lateral forces.

The varied ship motions provided non-inertial environments with increasing levels of disturbance, challenging the robots to maintain stability and execute locomotion tasks under dynamic conditions. Data collected from these motions informed the synthetic movements programmed into the treadmill for controlled in-lab experiments (i.e., "Treadmill Motion 3").

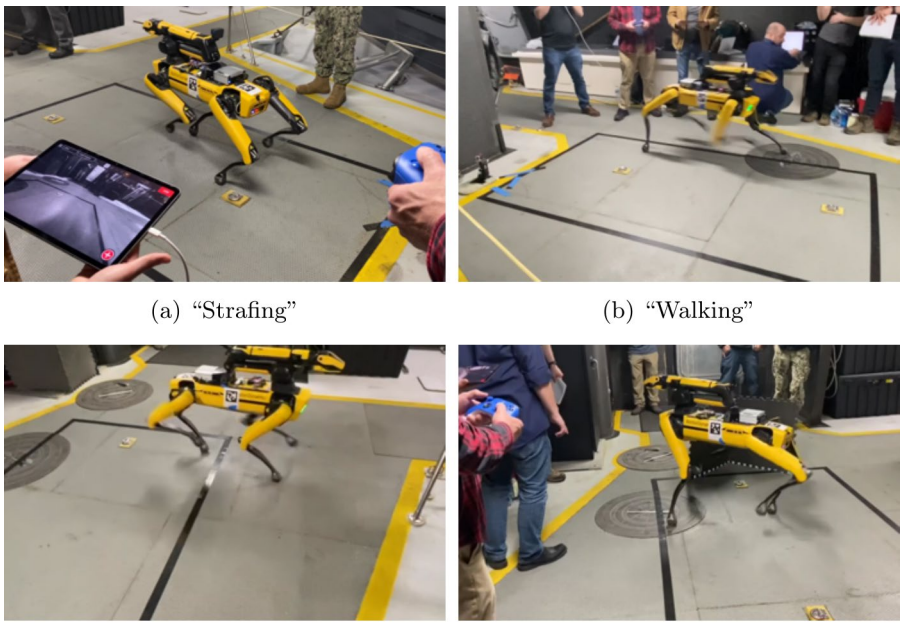
(2) Robot movement types: during the field experiments aboard the M80 Stiletto, a variety of representative robot motions were rigorously tested, outlined as follows:

(i) "Sitting": This motion involved Spot sitting powered on, but with the motors idle.
(ii) "Standing": These experiments consisted of Spot standing still with no joystick inputs, serving as the baseline for comparison with other activities.
(iii) "Walking in Place": Spot performed stepping motions in place to assess changes in its stability and performance related to ship movement without changing the its position within the testing area.
(iv) "Walking": These tests required Spot to walk back and forth from the rear to the front of the testing area.
(v) "Port-Starboard Walking": these experiments had Spot walk back and forth between the ship's port (i.e., left side of the vessel when facing the front) and starboard (i.e., right side).
(vi) "Strafing": Facing the front of the ship, Spot moved laterally from the port to the starboard.
(vii) "Knee-Knocker": This challenging motion had Spot navigating through knee-knockers to evaluate its stair-climbing capabilities.
(viii) "Bottle Pick-up": Simple loco-manipulation tests were conducted to determine Spot's ability to pick up a bottle amidst ship motions.

Figures 6 and 7 illustrate various movement tasks performed by Spot inside the ship during the field tests.

(3) Sensor setup: We used four inertial measurement units (IMUs) equipped with GPS to record the ship's pose and motion. These IMUs captured the wave-induced rotary and translational motions of the ship that the robots had to navigate during the trials. These sensors were contained in waterproof boxes (see Fig. 4c), and rigidly attached to the ship (Fig. 4d).

Six GoPro cameras were strategically placed inside the ship to capture video data of the robot's movements from various angles. For this testing phase, we chose to use GoPro



(a) "Strafing" (b) "Walking"

(c) "Port-Starboard Walking" (d) "Walking-in-Place"

Fig. 6 Examples of Spot's movement types within the shipboard testing environment

(a) (b) (c)

Fig. 7 Images of Spot moving inside the ship captured by GoPro cameras. **a** Combination of robot "Walking" motion with "Underway" ship movement, during which the robot fell over to the ground as shown in the image. **b** Combination of the robot's "Bottle Pick-up" motion and the "Underway" ship movement, where the robot successfully completed the task. **c** Robot "Knee-knocker" task

cameras instead of conventional optical motion capture systems, such as those from Vicon and Optitrack, because these systems typically require cameras to be mounted on stationary bases with minimal movement or vibrations. However, the ship used for our tests did not meet this requirement due to its significant movement during the experiments. In contrast, GoPro cameras are more resilient to base movements and vibrations and have been extensively used in human sports applications.

The cameras were mounted on the ceiling, deck floor, and rails to record both top-down and side views of the robot during the tests. Figure 7 showcases images captured by these cameras from different perspectives.

(4) Data synchronization: The data returned by different sensors were synchronized using different methods. For the GoPro cameras mounted on the ship, we displayed GPS time on each camera during field tests and used this displayed time to synchronize data recording across all external cameras during post-analysis. For external IMUs mounted on the ship, GPS data from each IMU were used to synchronize data across all IMUs. Spot's internal sensor data, including joint angles, velocity, torques, and foot-ground contact mode (on or off), were synchronized automatically through Spot's internal software.

Using these synchronization methods, data from the GoPro cameras and external IMUs were synchronized via GPS time. These were not synchronized with Spot's internal data, as the proposed analysis does not require synchronization between these datasets. However, synchronization between the two sets could be achieved by identifying distinct robot behaviors during the tests that both the GoPro cameras and Spot's internal sensors could capture. For example, at the start of the field experiments, the robot's initial standing motion caused a notable, sudden change in joint torque, captured by Spot's internal sensors, and this initial standing event was also recorded by the GoPro cameras.

3.3 Assessment measures and parameters

(1) Evaluation measures: We developed and implemented a set of quantifiable measures to assess the stability and performance of legged locomotion during ground acceleration:

Stability: Stability is an essential performance metric for legged robots, indicating how well they can stand, walk, and maneuver in environments with non-inertial frames such as moving naval vessels, airplanes, and trains. In such environments, robots continuously face perturbations from the accelerating or rotating ground, which may prevent timely recovery. Stability is measured by the robot's ability to remain upright and sustain intended behaviors, such as maintaining a stationary posture or following the desired path. In this study, a robot is considered unstable if it fails to maintain these behaviors and falls towards the ground.

Balance: Alongside stability, balance assesses whether a robot can keep its CoM within the support polygon. A legged robot's support polygon is defined as the convex hull formed by the ground contact points of the robot's feet, as illustrated in Fig. 8. For locomotion on regular static ground, this polygon represents the area within which the robot's CoM must remain to maintain quasi-static balance. The key metric here is the minimum distance from the CoM to the edges of the support polygon. The greater this distance, the less likely the robot may be to fall, providing a robust indicator of balance.

Body-position tracking accuracy: This measure is often used for tasks involving navigation and obstacle avoidance. We evaluated the accuracy of the robot's body position by comparing its actual position to the intended position, which was a fixed point on the dynamic ground for "Sitting," "Standing," and "Walking in Place" robot movements, or a continuous three-dimensional (3-D) path within the non-inertial environment for path following tasks including "Strafing," "Walking," and "Port-Starboard Walking" robot motions.

Trunk orientation steadiness: For tasks requiring precise manipulations, such as inspection or maintenance,

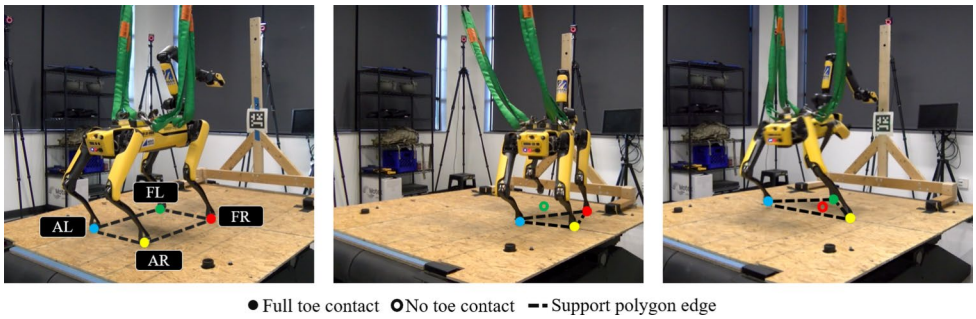


Fig. 8 Snapshot examples of the support polygon boundary for Spot on a moving treadmill

maintaining a steady trunk is critical. Trunk steadiness was evaluated based on the deviation of the robot's trunk pitch and roll angles from the horizontal plane. Excessive trunk movement can impair the performance of manipulator arms by transferring vibrations to the end-effector.

Joint effort exertion levels: Effective disturbance response and stability maintenance require rapid and adequate joint torque responses. Moreover, for hardware longevity, keeping joint torque exertion within manufacturer-specified limits is desirable. This study analyzed the robot's joint torque profiles against these limits to assess effort exertion levels during various tasks.

(2) Assessment data: Experiment data was collected using both external sensors (such as IMUs and GoPro cameras attached to the ship) and onboard sensors within the robots (e.g., encoders and IMUs). The joint data, joystick input values, and toe contact information were gathered using the robots' onboard sensors, while the robot's body position, trunk orientation, and support polygon were recorded using cameras (GoPro cameras for field tests and motion capture cameras for in-lab experiments).

The key data utilized to assess the effectiveness of the robot systems in addressing non-inertial environments are summarized as follows:

Treadmill position and orientation data during in-lab tests: The position of motion capture markers placed on the treadmill was used to obtain the ground-truth motion of the treadmill. This was necessary to verify the accuracy of the treadmill's motions, especially when operating near its motion limits.

Joint torque data: Joint torques were directly read from the robots' proprietary software. This data was compared across different testing trials and against motor limits to determine the effort required for the robots to maintain stability and remain upright under the continuous and unknown perturbations caused by the ground movements.

Contact data: Toe contact data, indicating whether a toe is in contact with the ground or airborne, was used to determine which legs were providing support at any given moment, crucial for identifying the support polygon's position.

CoM position data: The robots' CoM position was critical for evaluating balance. It was measured by the displacement of the CoM relative to the support polygon boundary during locomotion, using motion capture data from reflective markers attached to the robot as well as the inertial and geometric properties of the robots provided by the manufacturer.

Body position data: The robot's body position in the non-inertial environment is defined by the 3-D position of its approximate geometric center, calculated from the average position coordinates of the trunk's corners obtained based on the motion capture data. The body position data was instrumental in assessing the robot's deviation from its intended path within the dynamic environment.

Trunk orientation data: Trunk orientation, calculated from external camera data, was used to evaluate how well the robots maintained a steady orientation under the perturbations caused by ground motion.

Support polygon data: The support polygon data, required for balance assessment, was derived by combining camera data, joint angle data, and identified support toes from toe contact data.

4 Testing execution

This section describes the execution of test procedures for both the in-lab and field experiments.

4.1 Controlled in-lab experiments testing procedure

To ensure consistency and efficiency, a specific procedure was followed for the controlled in-lab experiments on Vision 60 and Spot:

Robot preparation: Before each trial, both robots were powered on, and initial calibration was conducted to align sensors correctly and establish baseline stability. The Kinova Gen2 arm was fitted to Vision 60, and the Spot Arm to Spot. Desired postures (e.g., extended or folded arm) and motion profiles (e.g., standing or walking in place) were pre-programmed into the robots' controllers. Each robot was then positioned at the center of the treadmill.

Treadmill motion execution: The treadmill was programmed to execute the predefined surface motions as described earlier. Each test incrementally increased the surface motion intensity to simulate progressively challenging conditions. Motion profiles were maintained for a set duration (e.g., 3–5 min) to allow adequate data collection and observation of the robots' long-term stability and performance under the specified conditions.

Data collection and post-processing: Data from all required external and onboard sensors were synchronized and continuously collected throughout each trial. Post-trial, data were reviewed to confirm proper recording and to identify any anomalies or unexpected behaviors. Subsequent processing and analysis focused on key performance metrics, as outlined in Sec. 3.3.

Safety constraints: In the controlled laboratory environment, comprehensive safety measures were enforced to protect both personnel and equipment. Safety precautions included a cage around the testing area, described in Sec. 2.2, and emergency stop buttons on the Motek M-Gait treadmill that could be easily accessed to halt all motions immediately in case of instability or robot failure. Motion sensors were installed to prevent close proximity to the moving treadmill. Robots were secured to overhead support frames with elastic safety lines to prevent falls from the treadmill during intense movements. Additionally, both Vision 60 and Spot were programmed with torque and velocity limits on their actuators to automatically prevent mechanical overload.

4.2 Field experiments testing procedure

The field experiments were conducted aboard the M80 Stiletto to test Spot's performance in real-world non-inertial environments. The ship's motion, caused by waves and maneuvers, created dynamic conditions that were more unpredictable and complex than the controlled in-lab tests. The following steps outline the testing procedure for the field experiments:

Ship preparation and equipment setup: The testing area on the deck of the M80 Stiletto was marked with black tapes to define the boundaries of the robot's movement zone, as shown in Fig. 4b. This area provided enough space for forward and lateral walking tests. Six GoPro cameras were placed around the testing area to record the robots' movements from different angles. Cameras were placed on the ceiling, deck floor, and rails to capture top-down and side views of the robots during the tests.

Sensor synchronization: GPS data was logged to ensure accurate time synchronization across all the cameras and onboard robot data.

Robot preparation: Prior to the experiment trials, Spot was powered on and underwent a calibration process using its onboard sensors. The robot was positioned at the center of the designated testing area, initially set to remain stationary to establish a baseline for stability assessment. Spot's proprietary control system, along with the "Moving Platform Mode" designed for stabilizing the robot on shifting grounds, was activated to optimize performance in the dynamic ship environment.

Execution of ship motions: The ship's motion was varied throughout the trials, starting with the "At Pier" condition (where the ship was docked) and progressing to more dynamic conditions such as "Underway" (with the ship moving at 10 knots) and "Sea State" (at 20 knots). The most aggressive trials involved the ship performing sharp "S-curve" maneuvers. Each motion profile was maintained for a set period of 3–6 min to observe how the robots adapted to different levels of ship motion intensity.

Execution of robot movements: The robots were commanded to perform the movement types as specified in Sec. 3.2. To avoid the Spot robot from moving overly close to nearby operators and to command the robot to perform the intended locomotion tasks, the robot took high-level commands, such as moving forward and backward, and faster or slower, provided from a joystick operated by the test team.

Data collection and post-processing: The external and robot on-board sensors were synchronized using the GPS data, and continuously recorded data throughout the trials. After each trial, the video footage and sensor data were reviewed to ensure proper synchronization and to identify any significant events, such as robot falls or instability. Data were processed to analyze key performance metrics specified earlier.

Safety constraints: During the field experiments aboard the M80 Stiletto, several safety precautions were taken to ensure the protection of both the robot and the testing personnel. Designated areas for robot testing were clearly marked on the deck to prevent personnel from approaching the test area overly closely, and personnel operated the robots remotely from a safe distance using joystick controls. These precautions were necessary due to the high torque spikes observed in the robot's joints during rapid lateral ship movements, which posed potential risks of mechanical failure or uncontrolled movements. The use of joystick control also prevented the robot from moving excessively close to operators, obstacles, or equipment in the shipboard environment.

5 Results and analysis

5.1 Summary of statistics and key insights

Throughout this study, comprehensive experiments were conducted, both in a controlled laboratory setting and aboard the M80 Stiletto naval vessel. The number of trials and the

Table 4 Number of trials per experimental set during in-lab tests

Experiment type	Treadmill motion	Number of trials	
		Vision 60	Spot
Baseline	No motion	5	5
Just Arm DRS	Motion 1	5	5
	Motion 2	5	5
	Motion 3	5	5
Just Body DRS	Motion 1	10	5
	Motion 2	10	8
	Motion 3	10	6
Both DRS	Motion 1	10	10
	Motion 2	10	5
	Motion 3	10	10

corresponding range of conditions provide an extensive dataset for analyzing the performance of both Vision 60 and Spot under dynamic surface motions.

A total of 80 trials were performed across various treadmill settings for each of Vision 60 and Spot during controlled in-lab experiments, as listed in Table 4. Each combination of robot movement types and treadmill motion types was tested in 5–10 trials. These trials encompassed motion profiles ranging from mild swaying to aggressive pitching and lateral movements, as detailed in Sec. 3. Each robot underwent testing for at least 60 s across various tasks. However, some experiments with Spot were cut short due to the instability of the robot system, leading to unsafe conditions. Consequently, the number of trials for Spot differed from those for Vision 60. Out of the total 80 tests, 64 trials were successfully conducted with Spot, while it failed in the remaining 16 trials.

A total of 42 trials were conducted aboard the M80 Stiletto to assess performance during various ship motions. These trials were categorized into different types of ship movement, including "At Pier," "Underway," "Sea State," and "S-Curves." Each trial lasted between 2 and 5 min.

Overall, Vision 60 performed more consistently across controlled in-lab experiments. Compared to Spot, it exhibited lower peak joint torques, more accurate body-position tracking, and better balance management, particularly during aggressive treadmill motions. Compared to Vision 60, Spot struggled more with maintaining stability under aggressive treadmill motions, and its joint torques were consistently higher. These insights set the stage for a more detailed analysis of the robots' performance, including joint torque comparisons, body-position tracking results, and balance maintenance across different conditions.

Between the controlled in-lab experiments and the shipboard tests, we observed that the robot failed more frequently in the lab than in the field. However, during the walking-in-place locomotion task without the arm deployed, the range of Spot's position on the ship was significantly larger than on the treadmill.

5.2 Observation 1: Limitations of current commercial systems in shipboard environments

Our experiments reveal that current commercial quadruped systems, such as Vision 60 and Spot, are not fully capable of performing locomotion and manipulation tasks effectively in dynamic shipboard environments. A critical factor contributing to this limitation is the inability of these robots to effectively reject the disturbances induced by the unknown ground motions (Fig. 9).

(1) Stability: Overall, Spot experienced more frequent failures in the lab than in the field. Out of the 80 in-lab trials, Spot failed in 16 trials. In contrast, out of the total of 42 field trials, Spot failed once and fell to the ground during the "Walking" robot motion and the "Underway" ship movement, as shown in Fig. 4a. Figure 10 displays the snapshots of Spot struggling to maintain walking stability on the treadmill, while Fig. 11 shows a trial where Spot was able to sustain stable walking. The difference in failure rates between lab and field tests may be attributed to the greater intensity of the treadmill motion that caused most of Spot's failures (i.e., "Treadmill Motion 2") compared to the ship motion, which is approximately represented by "Treadmill Motion 3." "Treadmill Motion 2" had larger amplitudes and higher frequencies in both pitch and sway directions than "Treadmill Motion 3." This comparison suggests that more aggressive ground motions, such as those simulated by "Treadmill Motion 2," can be more likely to induce instability in the robot.

In contrast, Vision 60 maintained stable walking across all controlled in-lab tests without falling over to the surface of the treadmill. Figure 12 illustrates Vision 60 walking on the treadmill with its end-effector commanded to track the fiducial marker rigidly attached to the treadmill post.

(2) Distance between CoM and support polygon boundary: The balance of a quadrupedal robot is heavily influenced by the relationship between the CoM and the support polygon, as mentioned earlier. Figures 13 and 14 show the box plots of the average values of the CoM distance from the support polygon boundary for Spot and Vision

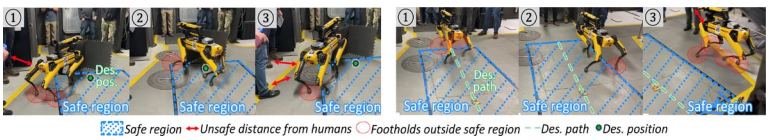


Fig. 9 Snapshots of Spot struggling to maintain balance inside the moving Stiletto ship

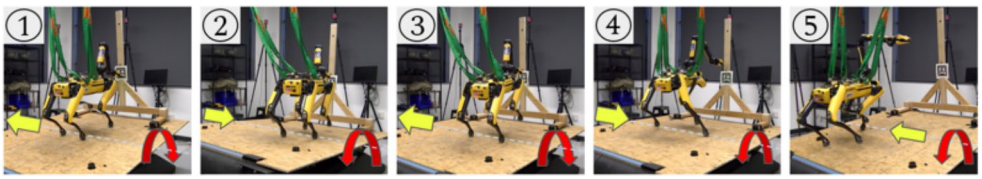


Fig. 10 Snapshots of Spot struggling to maintain balance on an accelerating and rotating treadmill. The yellow and red arrows show the treadmill's sway and pitch directions, respectively (color figure online)

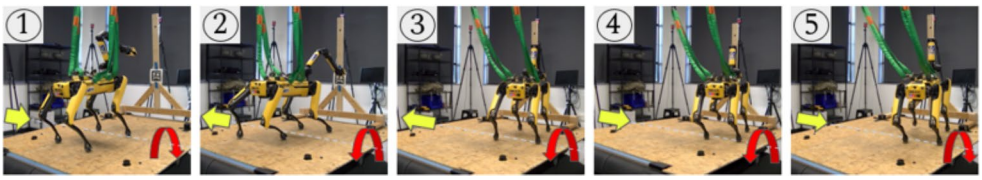


Fig. 11 Snapshots of Spot walking on treadmill without falling. The yellow and red arrows indicate the treadmill's sway and pitch directions, respectively (color figure online)

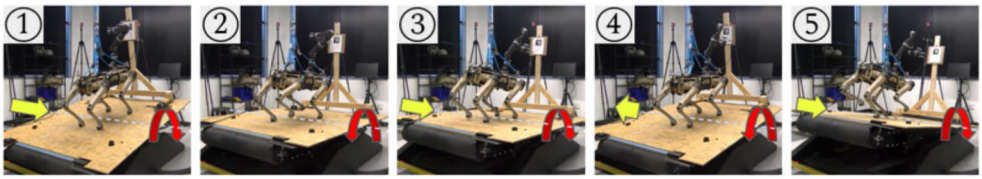


Fig. 12 Snapshots of V60 walking on treadmill without falling. The yellow and red arrows indicate the treadmill's sway and pitch directions, respectively (color figure online)

60, respectively. For the box plots, the smaller the distance between the CoM and the support polygon boundary, the more likely the robot system might lose balance. If the distance reaches 0, it means that the CoM has reached the support polygon's boundary, indicating that the robot is likely close to falling over.

The box plots shown in Figs. 13 and 14 are generated using the MATLAB function 'boxplot.' The whiskers in the box plots represent the data that lies outside the 25th and 75th percentiles but are not considered outliers. Let W denote the multiplier whisker. The value of W determines which data points are considered outliers, based on the following criteria: Data Point> $Q3 + W(Q3 - Q1)$ or Data Point> $Q1 - W(Q3 - Q1)$, where $Q1$ and $Q3$ are the 25th and 75th percentiles of the data, respectively. With $W = 0$, there would be no whiskers, and every point outside $Q1$ and $Q3$ would be considered as an outlier. The default setting of W in MATLAB is $W = 1.5$, which corresponds to approximately 99.3% coverage if the data are normally distributed. For all the box plots in this paper, we used a whisker multiplier value of $W = 15$ to further reduce the number of outliers compared to the default setting.

Vision 60 vs. Spot: Overall, in most experiments, Vision 60 performed better than Spot in maintaining a larger CoM distance from the support polygon edges. From the "baseline" box plots in Figs. 13 and 14, which represent tests with no treadmill motion, we observe that Vision 60's median CoM distance from the support polygon edge was approximately 72 mm, with a small interquartile range (IQR)-the range between the 25th and 75th percentiles. Spot 60 maintained a similar median distance for its baseline case, also with limited variability. This suggests both robots maintained a similarly balanced gait when the treadmill was stationary.

The remaining box plots in the two figures correspond to tests with treadmill motion. These plots indicate that Spot's median CoM distances from the support polygon boundary ranged between 57 and 73 mm, while Vision 60's median values ranged between 72 and 78 mm. Additionally, Vision 60's IQR values ranged between 1 and 5 mm, while Spot's IQR ranged between 5 and 17 mm. This comparison under moving treadmill conditions highlights that, compared to Spot, Vision 60 demonstrated a larger and more consistent

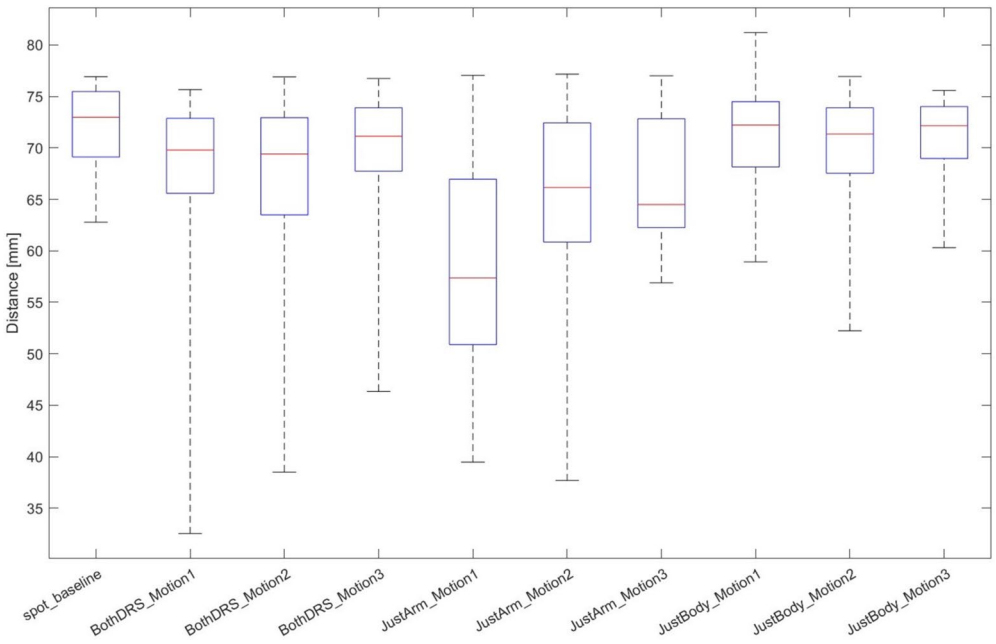


Fig. 13 Spot's CoM distance from the support polygon boundary during in-lab experiments. The *x*-axis indicates the robot and treadmill motion types as described in Sec. 3.1 (e.g., "BothDRS_Motion3" refers to the "Both DRS" robot motion during "Treadmill Motion 3")

balance margin across various motion types during in-lab experiments.

Furthermore, to more clearly show the CoM distance comparison between the two robots, Table 5 summarizes the average and standard deviation of the CoM distance for each test case shown in the box plots in Figs. 13 and 14. Comparing the average CoM distance between Vision 60 and Spot across all cases, we can see that the two robots kept a similar distance when the treadmill is stationary (i.e., under the baseline case), which is also revealed in the box plots. However, across all the rest test cases during treadmill motions, Vision 60 consistently achieved a longer CoM

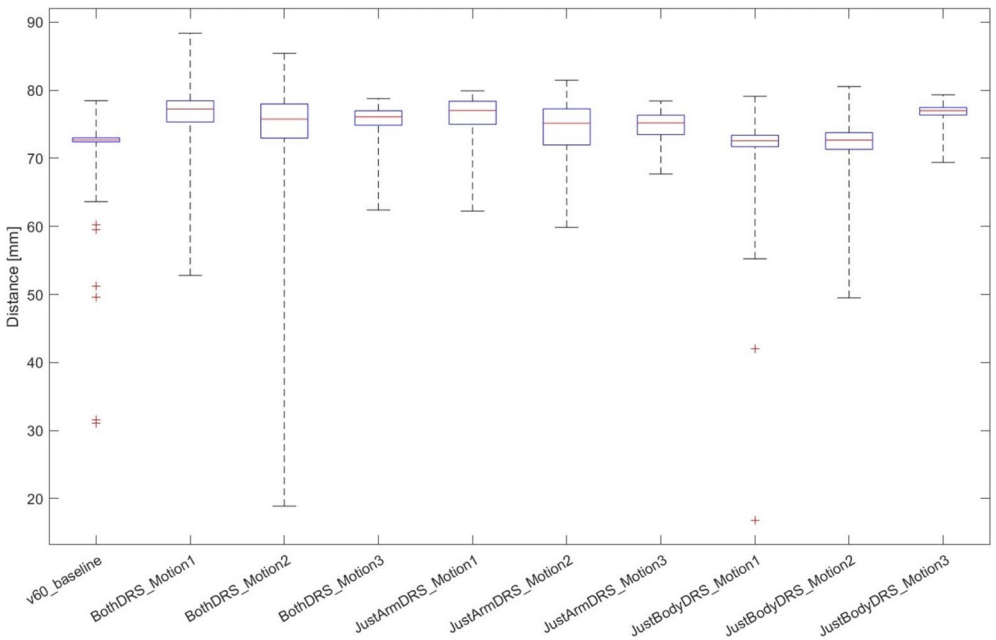


Fig. 14 Vision 60's CoM distance from the support polygon boundary during in-lab experiments. The *x*-axis indicates the robot and treadmill motion types as described in Sec. 3.1 (e.g., "BothDRS_Motion3" refers to the "Both DRS" robot motion during "Treadmill Motion 3")

Table 5 Comparison of CoM distance from the support polygon edge during in-lab experiments

Test name	Spot: average (standard deviation) [cm]	Vision 60: average (standard deviation) [cm]
Baseline	72.37 (3.22)	72.71 (0.61)
BothDRS_Motion 1	68.35 (5.93)	76.60 (2.40)
BothDRS_Motion 2	67.50 (7.00)	75.14 (3.29)
BothDRS_Motion 3	70.42 (4.29)	75.70 (1.77)
JustArmDRS_Motion 1	59.02 (10.08)	76.40 (2.49)
JustArmDRS_Motion 2	66.09 (6.73)	74.43 (3.56)
JustArmDRS_Motion 3	67.06 (5.41)	74.86 (1.86)
JustBodyDRS_Motion 1	70.91 (4.49)	72.51 (1.33)
JustBodyDRS_Motion 2	70.24 (4.69)	72.49 (2.18)
JustBodyDRS_Motion 3	71.29 (3.09)	76.80 (0.97)

distance from the support polygon edges, by a margin ranging from 1.6 to 17.4 cm and with an average difference of 7.1 cm. Also, Vision 60's standard deviation of the CoM distance for each case is consistently and significantly smaller than Spot, by a factor ranging between 1.9 and 16.4 cm, indicating that Vision 60's balance maintaining capability is more consistent than Spot.

Impact of treadmill motions on balance: Those figures also illustrate that more abrupt treadmill motions may cause more extreme variations in the CoM distance from the support polygon. It would appear that Vision 60 had better control over its CoM than Spot during the majority of experiments, except for the "Both" robot movement during "Treadmill Motion 2." However, most of Spot's failed experiment trials also occurred during the experiment set with the "Both" robot motion and "Treadmill Motion 2." Both robot systems performed the best during "Treadmill Motion 3" compared to the other two treadmill motions across all experiment types. This is potentially due to the relatively mild average motion of "Treadmill Motion 3" among the three treadmill motions tested.

(3) Body trajectory tracking accuracy: The accuracy of body-position trajectory is a key indicator of the robot's stability and ability to perform precise path following tasks.

Vision 60 vs. Spot: During the in-lab experiments, Vision 60 demonstrated a higher body-position trajectory accuracy compared to Spot, as shown in the box plots in Figs. 15 and 18. During the in-lab tests, Spot was programmed to place the end-effector right in front of the fiducial marker attached to the treadmill post. The location of the fiducial marker was sensed by Spot's trunk camera. The "Baseline" x- and y-displacements of Spot's body were between -100 and 150 mm and -150 and 100 mm, respectively, while the largest absolute deviations in x- and y-directions were both around 600 mm for the rest of the test trials. In contrast,

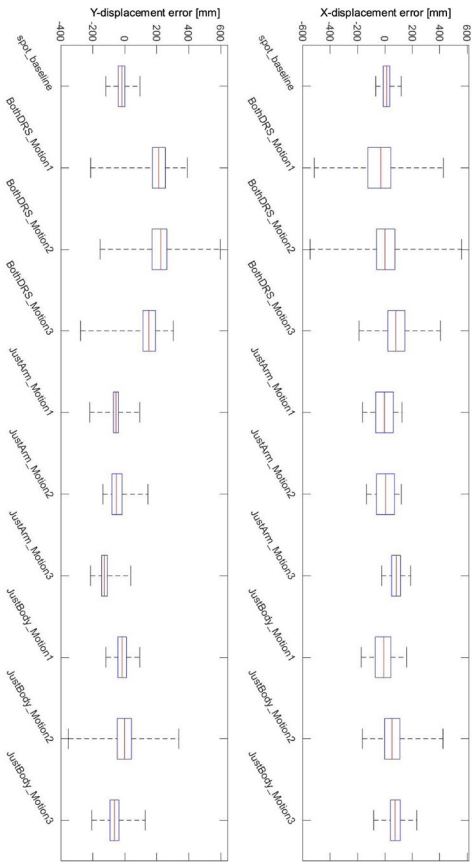


Fig. 15 Spot's body-position errors during in-lab experiments. The x-axis labels indicate the robot and treadmill motion types as described in Sec. 3.1 (e.g., "BothDRS_Motion3" refers to the "Both DRS" robot motion during "Treadmill Motion 3")

Vision 60's maximum absolute deviations in the x- and y-directions were 300 mm and 350 mm, respectively, which were both substantially smaller than Spot.

Figures 16, 17, 19, and 20 highlight the significantly greater body-position drift of Spot on the treadmill compared to Vision 60. The intended performance goal for the experiments shown in these figures was for each robot to step in place on the treadmill, keeping its trunk/body at a fixed location relative to the treadmill and maintaining its wrist position to accurately follow the marker attached to the treadmill's post. The robot's desired fixed body position is its initial location, marked as a red dot in each figure.

Figures 16 and 17 clearly show that Spot's trunk deviated significantly from its desired (initial) body position, with maximum deviations of approximately 50 cm and 30 cm in the x- and y-directions, respectively. This level of body-position tracking accuracy is inadequate for real-world tasks that require the robot to maintain a steady location on a moving platform (e.g., a ship) to perform manipulation tasks.

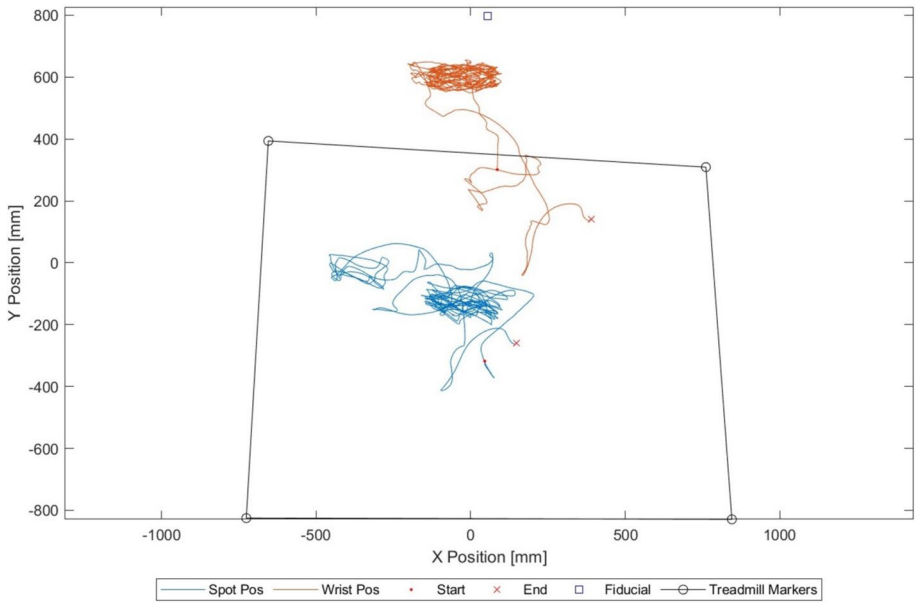


Fig. 16 Spot's in-lab tests with the largest x-displacement standard deviation. The legend label "wrist pos" refers to the position of the robot's wrist

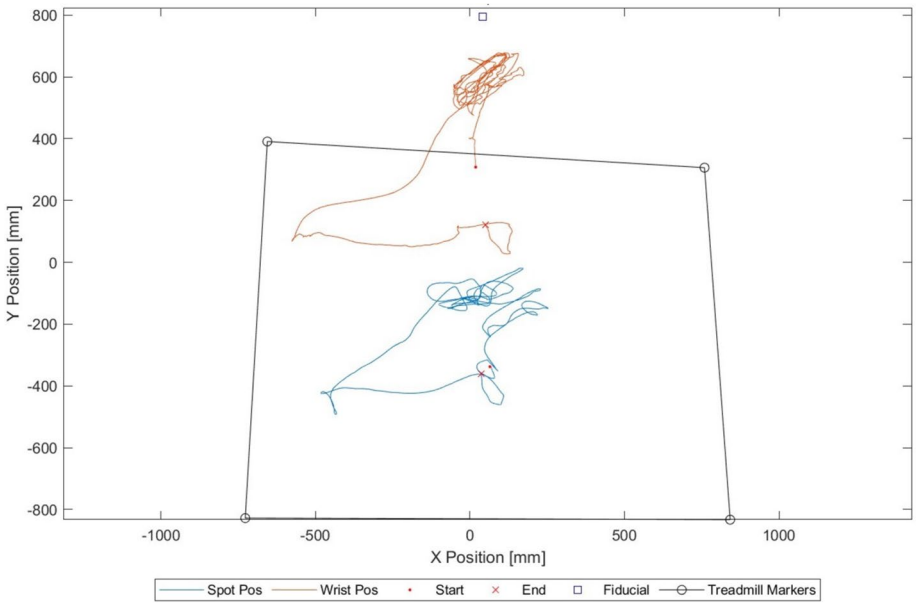


Fig. 17 Spot's in-lab tests with the largest y-displacement stdev. The legend label "wrist pos" refers to the position of the robot's wrist

Additionally, the wrist trajectories in these plots reveal a lack of tracking accuracy. Since the marker that the wrist should follow is fixed to the treadmill, the wrist position should ideally remain stationary in the plots. Instead, the wrist shows similarly poor tracking accuracy as the body position. Due to the kinematic coupling between the wrist and body, the wrist position is directly influenced by the body position. Therefore, high body-position tracking accuracy is essential for precise hand-position tracking on a moving surface. Inaccuracies in body-position tracking can propagate to other parts of the robot (Fig. 18).

Figures 19 and 20 present the body and wrist positions for the Vision 60 robot. Comparing these figures with Figs. 15 and 16, we observe that Vision 60's body-position tracking error is notably smaller than Spot's during in-lab experiments. Specifically, the maximum body-position deviations from the start point are approximately 30 cm and 20 cm in the x- and y-directions, respectively. This comparison highlights that Vision 60 achieves more consistent and accurate body-position tracking than Spot during in-lab experiments.

In-lab vs. field test results: Both in-lab and field tests revealed that more aggressive ground motion leads to greater body-position drift. For the in-lab tests, "Treadmill Motion 2" was the most intense, while in the field, the "S-Curve" motion was the most aggressive. The largest

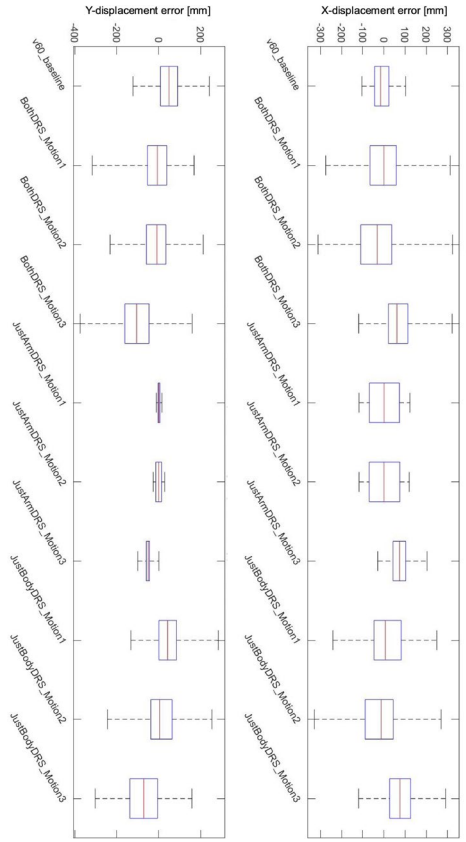


Fig. 18 Vision 60's body-position errors during controlled in-lab experiments. The *x*-axis indicates the robot and treadmill motion types as described in Sec. 3.1 (e.g., "BothDRS_Motion3" refers to the "Both DRS" robot motion during "Treadmill Motion 3")

body-position drift during walking tasks occurred under these conditions.

The walking-in-place (WiP) motion without arm movement was the most comparable task for Spot between lab and field tests. As shown in Fig. 21, during the shipboard "WiP*-P" tests, with the ship at the pier experiencing minimal motion, Spot's *x*-position range was approximately 15 cm, comparable to the 20 cm range in the in-lab baseline test. However, under the most aggressive ship motion, the "S-Curve," Spot's maximum position range exceeded 150 cm in the *x*-direction and 120 cm in the *y*-direction, as shown in Fig. 21. In contrast, the maximum range for in-lab tests occurred during "Treadmill Motion 2," with a position range of 60 cm in both directions, as shown in Fig. 14. Despite having a smaller range of drift during "Treadmill Motion 2," Spot experienced more frequent failures during this condition than in any shipboard test.

This discrepancy could be due to the fact that, in some in-lab tests, Spot quickly diverged and was stopped for safety before reaching excessively high body velocity and acceleration. This may explain why Spot failed more frequently in the lab but exhibited a smaller *y*-range of body position compared to the field tests. Additionally, the *x* and *y* body-position ranges in the field tests were measured using GoPro cameras, which may have limited accuracy.

Influence of ship motion on tracking accuracy: Figure 21 depicts Spot's actual body-position trajectories during the field tests. The abbreviations used in the two figures are explained in Table 6. During the field tests, it was observed that the body-position tracking accuracy generally degraded as the ship motion worsened, with and without the arm deployed, in both the forward and lateral directions. For instance, during the "Walking in Place" experiments, the "At

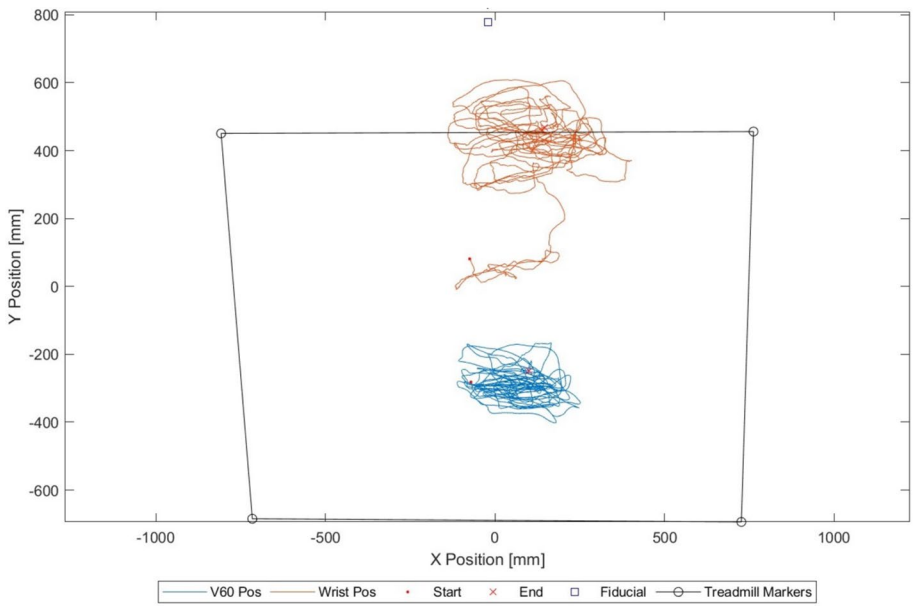


Fig. 19 Vision 60 in-lab experiment with the largest *x*-displacement standard deviation. The legend label "wrist pos" refers to the position of the robot's wrist

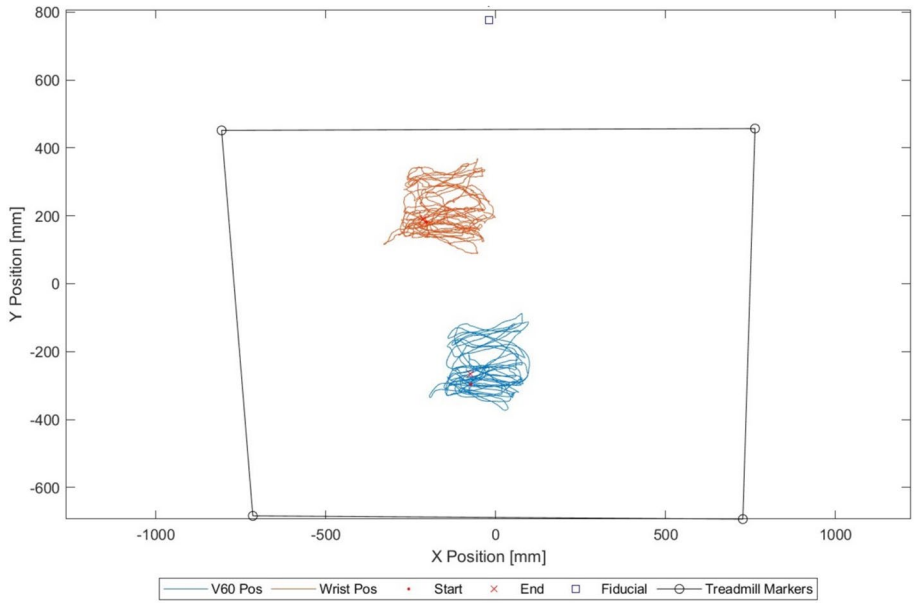


Fig. 20 Vision 60 in-lab experiment with the largest *y*-displacement standard deviation. The legend label "wrist pos" refers to the position of the robot's wrist

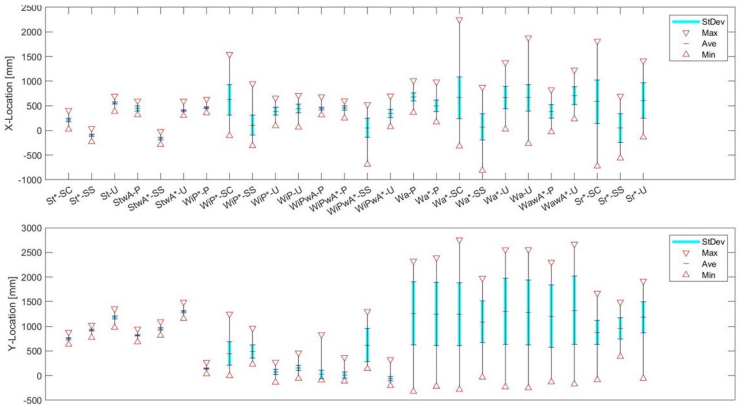


Fig. 21 Spot's body positions in the forward *x*- and lateral *y*-directions during field experiments as recorded from a GoPro camera. The *x*-axis label of the plots indicates the robot motion and ship motion types as described in Table 6

Pier" tests had the smallest position range, the "Underway" experiment was larger, then the "Sea State" range was larger, and the "S-Curves" had the largest. Figure 9 illustrates the scenarios where Spot failed to reliably track the desired body trajectory inside a moving ship. In subplot a, the ship motion was "S-Curve" with a maximum ship speed at 20 knots, during which the robot was not able to perform walking in place at a fixed location inside the ship. In subplot b, the ship was moving along a box path during the "Underway" ship motion with a maximum speed of 10 knots, where the robot failed to walk along the long center line of the rectangular walking area.

Summary: The data clearly indicate that both Vision 60 and Spot struggled to follow the desired body-position trajectories during aggressive ground motions. This lack of tracking accuracy can compromise their ability to reliably perform real-world inspection and manipulation tasks, especially under extreme lateral or pitching motions of the ground. While Vision 60 performed better in terms of body-position tracking accuracy, neither system was fully capable of maintaining sufficiently low tracking errors, highlighting the need for more advanced control algorithms and mechanical designs for shipboard operations.

(4) Trunk orientation steadiness: Overall, in controlled in-lab environments, Vision 60 and Spot exhibited similar trunk roll and pitch steadiness with respect to the inertial world frame. Also, compared to the "Baseline" experiments, the trunk orientation varied more substantially for both robots during the rest of the experiments, indicating the adverse effects of ground motions on the trunk steadiness for both robots.

Figures 22 and 23 depict the box plots of the two robots during all in-lab experiment trials excluding the failure trials of Spot where the robot movements were terminated due to aggressive behaviors caused by loss of balance. Compared to Spot, Vision 60 had a slightly larger range of the average roll

Table 6 Spot's Stiletto plot abbreviations for Fig. 21

Abbreviation	Meaning
St	Standing
WiP	Walking-in-place
Wa	Walking
Sr	Strafing
wA	With Arm
*	Walking platform mode on
-p	At Pier
-U	Underway
-SS	Sea State
-SC	S-Curves

angle values, but slightly smaller ranges of the average pitch and yaw angle values. However, Vision 60 had a smaller range of overall values for roll, pitch, and yaw angles. For both systems, "Treadmill Motion 2" had some of the largest ranges of overall and average values, due to the most aggressive ground motions among the three tested treadmill motions. It was observed that Vision 60 typically tended to orient its trunk as close as possible to be normal to the direction of gravity even when ascending or descending ramps. Also, Vision 60's step frequency tended to increase when ascending ramps and decrease when descending ramps. This behavior was observed during our reported experiments where the treadmill was essentially a swaying ramp with a varying pitch angle.

Specifically, across all experiments, the average values of Spot's trunk orientation ranged from −0.025 to 0.05 rads, −0.05 to 0.1 rads, and −0.2 to 0.25 rads for roll, pitch, and yaw, respectively. Across all experiments, the average values of Vision 60 ranged from −0.075 to 0.025 rads, −0.075 to 0.05 rads, and −0.1 to 0.25 rads for roll, pitch, and yaw, respectively.

5.3 Observation 2: Joint torque comparison: Vision 60 versus Spot

Joint torque is a critical metric for evaluating how well a robot can maintain stability and perform locomotion tasks, particularly in dynamic environments. It reflects the amount of effort exerted by the robot's actuators to maintain balance, compensate for surface motion, and execute gait or manipulation tasks.

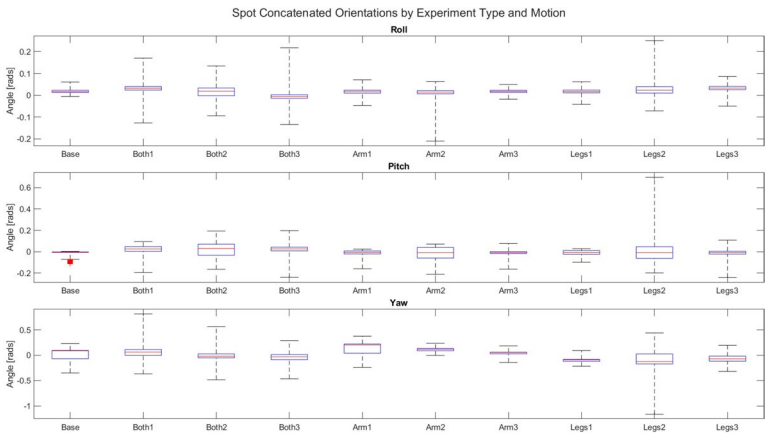


Fig. 22 Spot's trunk orientations during controlled in-lab experiments. The *x*-axis indicates the robot and treadmill motion types as described in Sec. 3.1. The words "Legs," "Arm," and "Both" refer to the "Just Body DRS," "Just Arm DRS," and "Both DRS" robot motions, respectively. The numbers "1," "2," and "3" represent "Treadmill Motion 1," "Treadmill Motion 2," and "Treadmill Motion 3," respectively

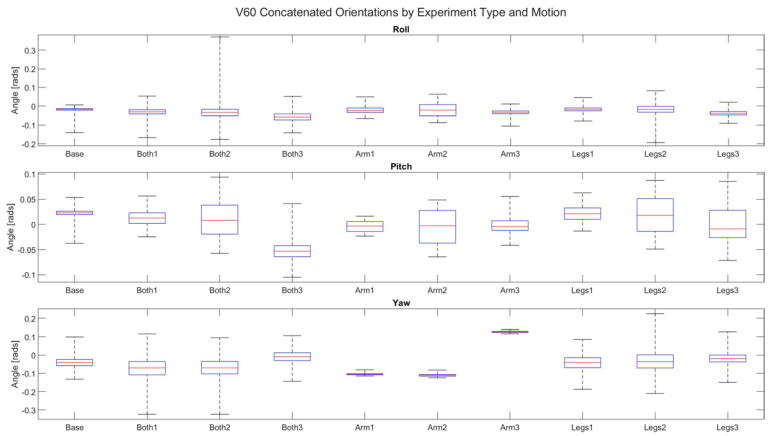


Fig. 23 Vision 60's trunk orientations during controlled in-lab experiments. The *x*-axis indicates the robot and treadmill motion types as described in Sec. 3.1. The words "Legs," "Arm," and "Both" refer to the "Just Body DRS," "Just Arm DRS," and "Both DRS" robot motions, respectively. The numbers "1," "2," and "3" represent "Treadmill Motion 1," "Treadmill Motion 2," and "Treadmill Motion 3," respectively

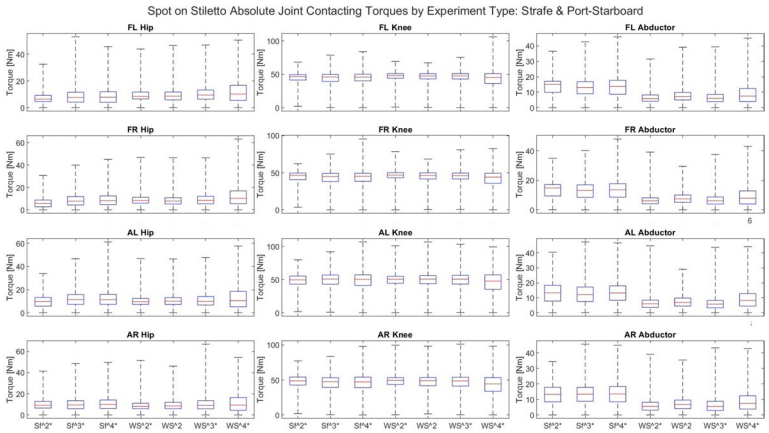


Fig. 24 Joint torques from Spot's "Port-Starboard" field experiments. The *x*-axis labels indicate the robot and ship motion types. "Sf" and "WS" refer to the "Strafing" and "Port-Starboard" robot motions, respectively. "2," "3," and "4" represent "Underway," "Sea State," and "S-Curve" ship motions, respectively. The symbol "*" indicates the "Moving Platform Mode" is on

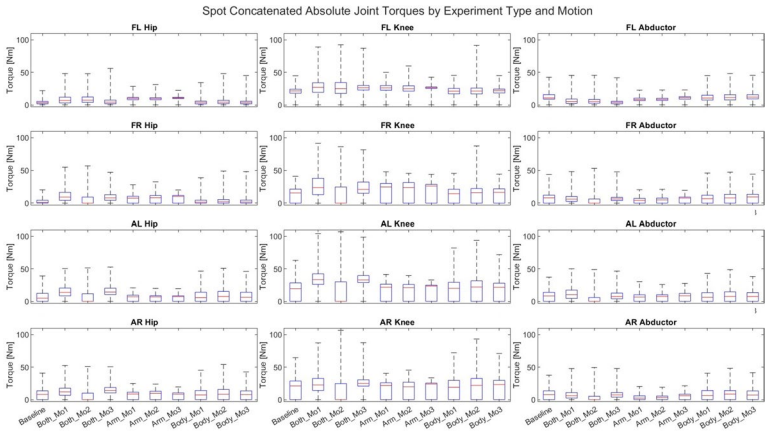


Fig. 25 Spot's joint torques from in-lab experiments. The *x*-axis indicates the robot and treadmill motion types as described in Sec. 3.1. The words "Body," "Arm," and "Both" refer to the "Just Body DRS," "Just Arm DRS," and "Both DRS" robot motions, respectively. The labels "Mo1," "Mo2," and "Mo3" represent "Treadmill Motion 1," "Treadmill Motion 2," and "Treadmill Motion 3," respectively

In both Vision 60 and Spot, the leg joints comprise the hip, knee, and abductor joints, as shown in Fig. 2a, and are responsible for maintaining stability and position tracking accuracy. These joints are subjected to varying levels of torques based on the intensity of the surface motions and the complexity of the locomotion tasks. By analyzing the joint torques, we can gain insights into how efficiently each robot compensates for disturbances and how much mechanical strain each robot experiences under dynamic conditions.

The box plots of Spot's and Vision 60's joint torques across all in-lab experiment trials are shown in Figs. 25 and 26, as well as all field tests for Spot in Fig. 24.

(1) In-lab joint torque analysis: Vision 60 exhibited a more controlled and consistent torque response across the experiments, compared to Spot. Vision 60 experienced lower maximum torques and a smaller range of torques than Spot, but slightly higher average values. This can especially be seen in the knee values; Spot had longer upper whiskers while Vision 60 had longer lower whiskers. Spot's "Walking in Place" knee torque averages were between 5 and 40 Nm, while Vision 60's were between 30 and 50 Nm. Vision 60 was heavier than Spot, which may explain Vision 60's higher torque averages, but Spot's higher torque maximums may be attributed to lower performance in maintaining stability. With higher maximum values of the actual torque signals commanded by a controller, the chance of motor failure from either robot increases, which in turn increases the chance of losing stability. This may explain why Spot experienced failed tests during experiments that used the aggressive "Treadmill Motion 2," which induced the most substantial disturbances among the three tested treadmill motion profiles.

(2) In-lab vs. field test results: Compared to the field experiments, Spot had lower overall knee torque values and higher abductor torque values during the in-lab experiments. The higher knee torque values during field experiments could be due to the wave swells and varying g-forces that were experienced by the robot during ship motions, but were significantly smaller during the in-lab experiments due to the robots experiencing much smaller linear vertical ground motions. The higher abductor torque values during in-lab experiments, as compared to the field experiments, could likely be attributed to the increased side-side ground movements that Spot endured during the in-lab experiments.

There were trends that were seen in the torque data from both field and in-lab experiments. The knee joints

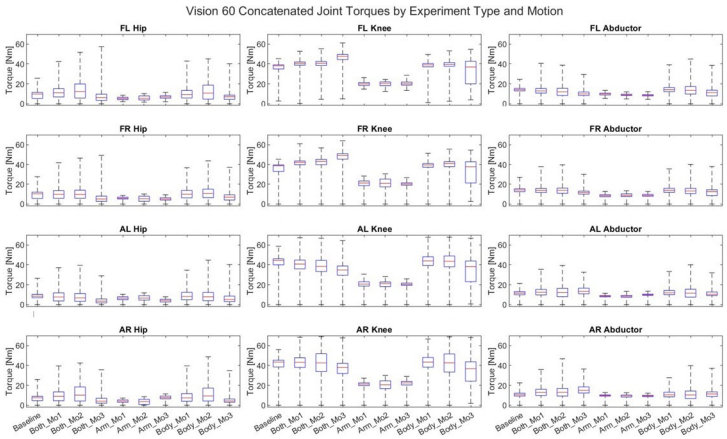


Fig. 26 Vision 60's absolute joint torques from in-lab experiments. The *x*-axis indicates the robot and treadmill motion types as described in Sec. 3.1. The words "Body," "Arm," and "Both" refer to the "Just Body DRS," "Just Arm DRS," and "Both DRS" robot motions, respectively. The labels "Mo1," "Mo2," and "Mo3" represent "Treadmill Motion 1," "Treadmill Motion 2," and "Treadmill Motion 3," respectively

consistently experienced the highest torques among all joints. This phenomenon was likely due to the knee joints bearing the brunt of mechanical stress during locomotion, especially in dynamic activities such as those simulated on treadmills. Furthermore, the aft leg joints registered higher torques compared to the front leg joints.

(3) Impact of arm deployment on torque exertion level: Similar to Spot, during the controlled in-lab tests on Vision 60, the robot was evaluated both with and without its arms deployed. The torque data from these tests revealed that the use of the arm resulted in higher hip torque values in the front leg joints and increased knee torque values in the aft leg joints. This pattern suggests that deploying the arm alters the robot's dynamics, increasing load and stress on specific joints to compensate for the shifted CoM and the additional kinetic forces during arm movement. Interestingly, despite these variations in specific joint torques, the overall torque values and averages across different test sets were similar whether the arm was deployed or stowed. This consistency is likely due to Vision 60's considerably heavier total mass compared to the mass of its arm. Although the arms of both Vision 60 and Spot have similar masses (approximately 5 kg), Vision 60's total weight is notably greater than Spot's. The relatively lighter arm in comparison to the heavier body may explain why the arm's motion had only a mild effect on Vision 60's behavior, whether the arm was attached or not.

(4) Effects of torque limit violation on robot stability: Exceeding the peak torque limits, even briefly, poses challenges for robot control. It can lead to motor overheating, reduced control precision, and jerky or unstable robot movements. To illustrate how higher torques impact Spot's performance compared to Vision 60, we calculated the percentage of time during which at least one of the robot's joint torques exceeded 80% and 100% of the peak joint-motor torque limits over each experiment's entire test period. The results show that Vision 60 did not even exceed the 80% peak torque limits in any test trial. In contrast, Spot exceeded 80% of the peak torque limits in seven test cases:

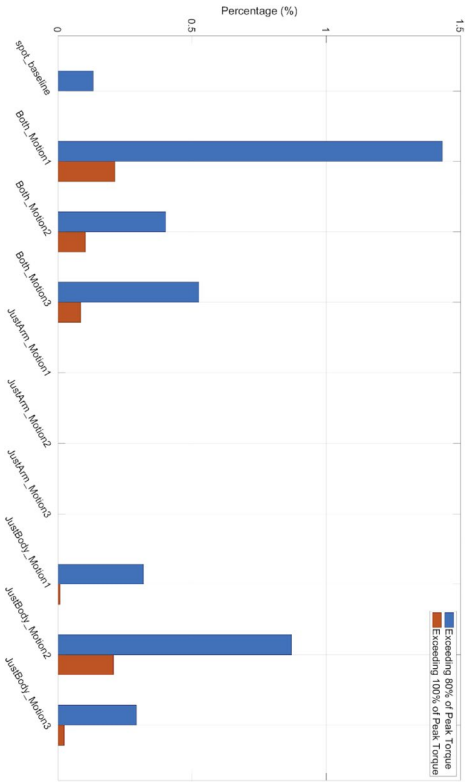


Fig. 27 Percentage of the entire time period of each experiment case during which at least one of Spot's joint torques exceeds 80% (blue) or 100% (red) of the peak joint-motor torque limits (color figure online)

"spot_baseline," "Both_Motion1," "Both_Motion2," "Both_Motion3,""JustBody_Motion1," "JustBody_Motion2," and "JustBody_Motion3," as summarized in Fig. 27. The associated time percentages of exceeding the 80% torque limit ranged from 0.13% to 1.43%, respectively. Except for the baseline case of "spot_baseline," all the other six cases also showed nonzero percentages of time exceeding the 100% peak torque limits.

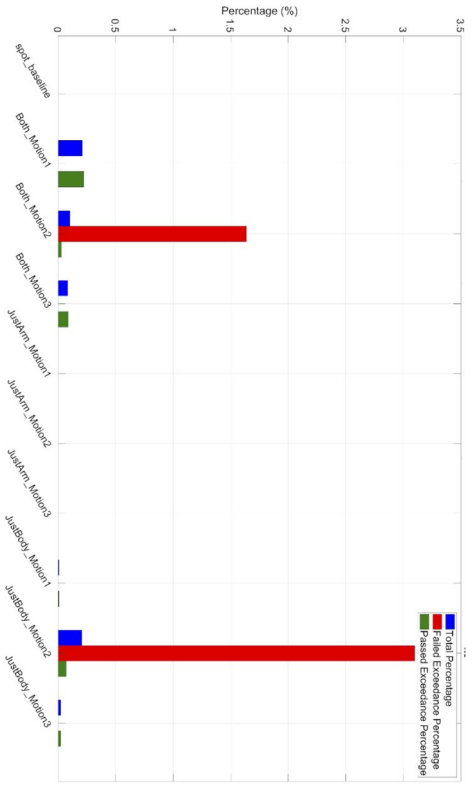


Fig. 28 Percentage of the entire time period of each experiment case during which at least one of Spot's joint torques exceeds the peak torque limits over: all tests within each experiment case (blue), all failure tests within each experiment case (red), and all success tests within each experiment case (green) (color figure online)

In Fig. 27, we also observe that in cases where only the arm moved while all four legs maintained contact with the moving treadmill (i.e., "JustArm_Motion1," "JustArm_Motion2," and "JustArm_Motion3"), the robot did not exceed the torque limits. This may be because the larger support polygon in these scenarios makes it easier for the controller to maintain balance. By contrast, in tests where Spot took steps (i.e., the six cases listed earlier), the time percentages of torque limit violations were all nonzero, with the highest percentages occurring during the most aggressive treadmill motion, "Treadmill Motion 2."

Moreover, comparing torque limit violation percentages across all cases, we see that failure cases have significantly higher percentages of torque limit violations than success cases, as shown in Fig. 28. This suggests that the more frequent torque limit violations may have contributed to Spot's higher failure rate compared to Vision 60 during in-lab tests.

5.4 Discussions

(1) Primary causes of the identified performance gaps: The unsatisfactory performance of Vision 60 and Spot can be fundamentally attributed to the underlying assumption in their control systems that the ground remains stationary in the Earth's inertial frame. Although the proprietary details of these control systems are not disclosed to the research community, the assumption of a static ground has long been prevalent in the control of ground robots, including legged robots. This assumption influences dynamics modeling, state estimation, control, and motion planning. Without explicit consideration, surface motion acts as persistent and continuous perturbations to the robot's control systems, making it difficult for existing approaches to manage these perturbations effectively.

The precise effects of surface motion on the controller can be understood through dynamic models of a legged robot walking on a rigid surface that is either significantly heavy or rigidly actuated (e.g., vessels or airplanes). Our previous work (Iqbal et al., 2020) proposed such a dynamics model. During a continuous phase where some legs maintain ground contact while the others swing in the air, the robot's dynamics can be modeled using Lagrange's method, resulting in second-order, nonlinear differential equations. This model indicates that the holonomic constraints formed at the robot-ground contact area are explicitly time-varying, rather than time-invariant, which fundamentally changes the time-variance property of the robot's dynamics associated with locomotion on static terrains. Without explicit treatment, surface motion induces significant modeling errors, leading to performance degradation in controllers designed based on static ground assumptions. Beyond continuous-phase dynamics, surface motion also impacts other components of the robot's dynamics. Specifically, the discrete dynamics, which describe how the robot's state evolves across leg-landing events, become explicitly time-varying, contrasting with the time-invariant nature of robot dynamics during walking on static ground. Furthermore, the switching surface, which is a hypersurface describing the condition of transition between the continuous phases and discrete events, also shifts from being time-invariant to explicitly time-varying during surface motion.

(2) Effects of CoM height on stability and performance: A lower CoM height may be advantageous for stability and overall performance during ground motion. For legged robots, a lower CoM typically corresponds to more heavily bent knees. Our prior findings (Gu et al. 2015) indicate that, under the same joint torque limits, more heavily bent knees can lead to higher dynamic CoM manipulability. This means that the robot has a greater range of possible

CoM accelerations it can achieve immediately, without violating ground-contact constraints (e.g., avoiding foot slippage and rolling). Enhanced CoM manipulability can improve disturbance rejection, as it can allow the robot to react more effectively to maintain balance and optimize performance.

During the reported tests, Vision 60 indeed maintained a lower average CoM height compared to Spot, with both robots controlled by their respective proprietary systems. This difference likely contributed to Vision 60's better balance performance compared to Spot. Specifically, Spot's average walking height was 61 cm, while V60's was approximately 42 cm.

(3) Future research directions on legged robot control: While some control approaches that explicitly address ground motion for reliable locomotion performance have been proposed, as summarized in the introduction, the locomotion control problem remains largely underexplored for non-inertial environments. First, to manage unknown surface motions, surface motion estimation could potentially be integrated with robust control methods (e.g., Iqbal et al. 2023) to enhance locomotion robustness without inducing excessive joint torques. Second, for data-driven approaches, incorporating surface motion into algorithm training could improve the algorithm's robustness to varying surface conditions. Finally, another key open question is state estimation in environments such as indoor shipboard settings where sensors such as GPS may not function well and there are no fixed landmarks attached to a stationary ground.

(4) Future research directions on legged robot testing: Future investigations will aim to broaden the scope of our field testing for legged robots by incorporating a diverse array of ships. Since the disturbances experienced on smaller vessels differ significantly from those on larger ships in identical ocean conditions, this diversity will allow for a more comprehensive assessment of robotic stability and functionality across different maritime platforms. Furthermore, enhancing the data collection setup during these field tests would be meaningful. Improvements will include upgrading our camera systems to capture higher quality data on robot trunk position, as well as acquiring more precise measurements of the CoM and support polygon. Although setting up motion capture cameras on a ship presents challenges, we intend to use high-resolution cameras equipped with wide-angle lenses, replacing the fisheye cameras previously employed, to ensure superior data integrity and reliability.

6 Conclusion

This study provided an in-depth evaluation of commercial quadruped robots, specifically the Ghost Robotics Vision 60 and Boston Dynamics Spot, in dynamic naval environments. Our findings highlighted the robots' capabilities and limitations in maintaining locomotion stability and performance under challenging conditions. Despite the robots' potential, consistent performance was lacking in highly dynamic scenarios. This points to the need for advanced control strategies in order to improve their operational reliability in unpredictable non-inertial settings involving accelerating or rotating ground. Another important future research direction is to enhance standardized benchmarking for legged robots in various real-world non-inertial environments with comprehensive, high-precision data collection.

Acknowledgements This research has been partly supported by the U.S. Office of Naval Research (N00014-21-1-2582, N00014-23-1-2744, N00014-24-1-2028, N00014-24-1-2634).

Author contributions S. M.: Methodology, Experimentation, Data Analysis, Generation of Open-Source Dataset, and Paper Writing.B. H.: Experimentation.B. W.: Methodology, Data Analysis, Paper Writing and Editing. R.D.: Experimentation. A. J.: Data Analysis, Generation of Figures, and Generation of Open-Source Dataset. M. K.: Generation of Figures. J. G. T.: Experimentation, Project Administration. A. N.: Methodology, Project Administration, and Funding Acquisition. R. A.: Methodology, Data Analysis, Paper Writing and Editing, Supervision, Project Administration, and Funding Acquisition. Y. G.: Methodology, Data Analysis, Paper Writing and Editing, Supervision, Project Administration, and Funding Acquisition.

Data availability The dataset from this study has been made publicly available for broader use and can be accessed online at https://github.com/purdue-tracelab/quadruped_assessment.

Declarations

Conflict of interest The authors declare no competing interests.

References

Aller, F., Pinto-Fernandez, D., Torricelli, D., Pons, J.L., Mombaur, K.: From the state of the art of assessment metrics toward novel concepts for humanoid robot locomotion benchmarking. IEEE Robot. Autom. Lett. **5**(2), 914–920 (2019)

B.V., M.M.: M-Gait: The High-Tech Modular Treadmill System. https://www.motekmedical.com/solution/m-gait/ (2024). Accessed 14 July 2024

BostonDynamics: Spot—The Agile Mobile Robot. https://bostondynamics.com/products/spot/ (2023). Accessed 9 Nov 2023

Brennan, S.: In Focus: Motek. https://logemas.com/in-focus-motek/ (2023). Accessed 13 July 2024

Caron, S., Escande, A., Lanari, L., Mallein, B.: Capturability-based pattern generation for walking with variable height. IEEE Trans. Robot. **36**(2), 517–536 (2019)

Castano, J.A., Humphreys, J., Mingo Hoffman, E., Fernández Talavera, N., Rodriguez Sanchez, M.C., Zhou, C.: Benchmarking dynamic balancing controllers for humanoid robots. Robotics **11**(5), 114 (2022)

Chen, H., Hong, Z., Yang, S., Wensing, P.M., Zhang, W.: Quadruped capturability and push recovery via a switched-systems characterization of dynamic balance. IEEE Trans. Robot. **39**, 2111–2130 (2023)

Dai, M., Xiong, X., Ames, A.: Bipedal walking on constrained footholds: momentum regulation via vertical com control. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 10435–10441 (2022)

Donald, R., Hertel, B., Misenti, S., Yan, G., Azadeh, R.: An adaptive framework for manipulator skill reproduction in dynamic environments. In: Proceedings of International Conference on Ubiquitous Robots, pp. 498–503 (2024)

G., M.: Boston Dynamics Case Subject: "Spot CoM using urdf"'. https://support.bostondynamics.com/s/case/5004X00001vd7VzQAI/spot-com-using-urdf (2023). Accessed 13 Nov 2023

Gao, Y., Gong, Y., Paredes, V., Hereid, A., Gu, Y.: Time-varying ALIP model and robust foot-placement control for underactuated bipedal robotic walking on a swaying rigid surface. In: Proceedings of American Control Conference, pp. 3282–3287 (2023)

Gao, Y.: Time-varying control and state estimation for bipedal locomotion. Ph.D. thesis, University of Massachusetts Lowell (2023)

Gao, Y., Yuan, C., Gu, Y.: Invariant filtering for legged humanoid locomotion on a dynamic rigid surface. IEEE/ASME Trans. Mechatron. **27**(4), 1900–1909 (2022)

Gibson, G., Dosunmu-Ogunbi, O., Gong, Y., Grizzle, J.: Terrain-adaptive, alip-based bipedal locomotion controller via model predictive control and virtual constraints. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 6724–6731 (2022)

Gu, Y., Yao, B., Lee, C.S.G.: Feasible center of mass dynamic manipulability of humanoid robots. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 5082–5087 (2015)

He, Z., Teng, S., Lin, T.-Y., Ghaffari, M., Gu, Y.: Legged robot state estimation within non-inertial environments. In: Proceedings of Control and Decision Conference (2024, to appear)

Inc., M.A.: Motion Analysis: cortex software. https://www.motionanalysis.com/software/cortex-software/ (2024). Accessed 14 July 2024

Inc., M.A.: Motion Analysis: three types of camera, with a range of specs. https://www.motionanalysis.com/camera/ (2024). Accessed 14 July 2024

Iqbal, A., Gu, Y.: Extended capture point and optimization-based control for quadrupedal robot walking on dynamic rigid surfaces. In: Proceedings of IFAC Modeling, Estimation, and Control Conference, vol. 54, pp. 72–77 (2021)

Iqbal, A., Veer, S., Gu, Y.: Asymptotic stabilization of aperiodic trajectories of a hybrid-linear inverted pendulum walking on a vertically moving surface. In: Proceedings of American Control Conference, pp. 3030–3035 (2023)

Iqbal, A.: Modeling, planning, and control of quadrupedal locomotion on dynamic rigid surfaces with unknown vertical motions. Ph.D. thesis, University of Massachusetts Lowell (2023)

Iqbal, A., Gao, Y., Gu, Y.: Provably stabilizing controllers for quadrupedal robot locomotion on dynamic rigid platforms. IEEE/ASME Trans. Mech. **25**(4), 2035–2044 (2020)

Iqbal, A., Veer, S., Gu, Y.: Analytical solution to a time-varying LIP model for quadrupedal walking on a vertically oscillating surface. Mechatronics **96**, 103073 (2023)

Norton, A., Ahmadzadeh, S. R., Jerath, K., Robinette, P., Weitzen, J., Wickramarathne, T., Yanco, H., Choi, M., Donald, R., Donoghue, B., et al.: Decisive test methods handbook: Test methods for evaluating sUAS in subterranean and constrained indoor environments, pp. 1–107. arXiv:2211.01801 (2022)

Norton, A., Ahmadzadeh, S. R., Jerath, K., Robinette, P., Weitzen, J., Wickramarathne, T., Yanco, H., Choi, M., Donald, R., Donoghueet, B., Dumas, C., Gavriel, P., Giedraitis, A., Hertel, B., Houle, J., Letteri, N., Meriaux, E., Khavas, Z. R., Singh, R., Willcox, G., Yoni, N.: DECISIVE Benchmarking Data Report: sUAS Performance Results from Phase I. arXiv:2301.07853 (2023)

Radosavovic, I., Xiao, T., Zhang, B., Darrell, T., Malik, J., Sreenath, K.: Learning humanoid locomotion with transformers (2023). arXiv preprint arXiv:2303.03381

Shi, F., Zhang, C., Miki, T., Lee, J., Hutter, M., Coros, S.: Rethinking robustness assessment: adversarial attacks on learning-based quadrupedal locomotion controllers. In: Proceedings of Robotics: Science and Systems (2024)

Torricelli, D., Pons, J.L.: Eurobench: preparing robots for the real world. In: Wearable Robotics: Challenges and Trends: Proceedings of the International Symposium on Wearable Robotics, pp. 375–378. Springer (2019)

Weng, B., Castillo, G.A., Kang, Y.-S., Hereid, A.: Towards standardized disturbance rejection testing of legged robot locomotion with linear impactor: a preliminary study, observations, and implications. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 9946–9952 (2024)

Weng, B., Castillo, G.A., Zhang, W., Hereid, A.: On safety testing, validation, and characterization with scenario-sampling: a case study of legged robots. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5179–5186 (2022)

Weng, B., Castillo, G.A., Zhang, W., Hereid, A.: On the comparability and optimal aggressiveness of the adversarial scenario-based safety testing of robots. IEEE Trans. Robot. **39**, 3299–3318 (2023)

Westervelt, E.R., Chevallereau, C., Choi, J.H., Morris, B., Grizzle, J.W.: Feedback Control of Dynamic Bipedal Robot Locomotion. CRC Press, Boca Raton (2007)

Xiong, X., Ames, A.: 3-d underactuated bipedal walking via h-lip based gait synthesis and stepping stabilization. IEEE Trans. Robot. **38**, 2405–2425 (2022)

Zhao, Y., Gu, Y.: A non-periodic planning and control framework of dynamic legged locomotion. Int. J. Intell. Robot. Appl. **4**, 95–108 (2020)

Stephen Misenti received his M.S. degree in Mechanical Engineering at the University of Massachusetts Lowell (UML) with a Vibrations/Dynamics/Controls concentration in 2024. Before this he received a B.S. degree in Mechanical Engineering with a Minor in Robotics from UML. During his time studying, he served in the Massachusetts Air National Guard at Barnes ANGB in Westfield, MA as an F-15 crew chief. Stephen started his professional career at SIG Sauer as a Research & Development Test Engineer developing test plans and designing data acquisition systems.

Brendan Hertel received a B.S. degree in Mechanical Engineering (2022) and a M.S. in Computer Science (2023) at the University of Massachusetts Lowell (UML). He is currently pursuing his PhD in Computer Science at UML, where he performs research in the Persistent Autonomy and Robot Learning (PeARL) lab. His main research topics are learning from demonstration and robot autonomy. Interests in these areas include learning from failed demonstrations and metric creation and evaluation.

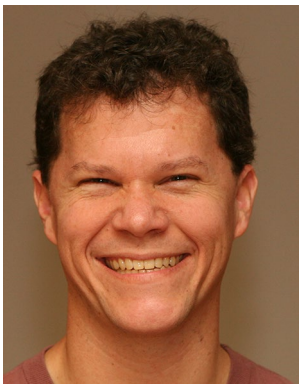
Dr. Bowen Weng is an Assistant Professor with Department of Computer Science at Iowa State University (ISU). His research focuses on the development, evaluation, and validation of artificial intelligence within embodied and physical domains interacting with humans and other robots, which are exemplified through applications involving wheeled and legged robotic systems. Before joining ISU, Dr. Weng was a Research Scientist at Transportation Research Center Inc. on assignment to National Highway Traffic Safety Administration (NHTSA), U.S. Department of Transportation, where he led the research and technical development of projects related to safety testing and performance evaluation of Automated Driving Systems. Mr. Weng received his Ph.D. in Electrical and Computer Engineering in 2023 from The Ohio State University, Columbus, USA.

Ryan Donals received his M.Sc. in 2024 from the Electrical and Computer Engineering Department at the University of Massachusetts Lowell (UML). His thesis focuses on the development of methods for measuring and increasing robot autonomy in dynamic environments. In 2023, Ryan received his B.Sc. degree in Electrical and Computer Engineering with a minor in Robotics from UML.

Advait (AJ) Jawaji earned his B.S. degree in Mechanical Engineering from Purdue University in 2024. During his junior year, AJ joined TRACE Labs, where he developed a data visualizer for the humanoid robot Digit. Following this, he gained extensive experience in the control and programming of industrial autonomous construction machines. Most recently, AJ has been focused on implementing controllers for Unitree's Go1 robot, enabling it to walk on a swaying treadmill with unknown motion. Additionally, AJ has been contributed in developing and enhancing an EPICS course that equips students with foundational skills in Python programming. Currently, he is investigating reinforcement learning based control strategies for legged robots.

Magnus-Tryggvi Kosoko-Thoroddsen received his B.S. degree in Mechanical Engineering from MIT in 2023. His education experience focused on dynamics and controls, as well as internship experience with Myomo Inc. and an undergraduate research position with the MIT-NASA MOXIE Group, has pushed him to pursue an M.S. at Purdue to further investigate non-linear control and control design of humanoid

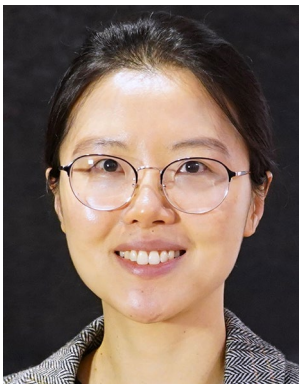
robotics. Magnus is currently working on evaluation and improvement of an existing quadrupedal locomotion controller.



Greg Trafton is the section head for the Human and Machine Intelligence Section in the Artificial Intelligence Center at the Naval Research Laboratory. Greg received his BS in computer science from Trinity University. He received a Ph.D from Princeton University in 1994. Greg works on cognitive robotics, human robot interaction, and artificial intelligence. He joined NRL in 1995. He mixes methods from cognitive modeling, AI, and cognitive science to improve (our understanding of) interaction.

Adam Norton is the Associate Director of the New England Robotics Validation and Experimentation (NERVE) Center at the University of Massachusetts Lowell. His research interests include the test and evaluation of robot, human, and human-robot performance, and the development of metrics, test methods, and benchmarking tools for robots. Adam has developed metrics and test methods and led evaluations for autonomous industrial vehicles, robot manipulators, exoskeletons, response robots, unmanned aerial systems, and human-robot interaction. He develops standards for robot systems on several ASTM International committees including E54.09, F45, and F48.

Reza Azadeh is an Associate Professor with the Miner School of Computer and Information Sciences at the University of Massachusetts Lowell (UML), where he directs the Persistent Autonomy and Robot Learning (PeARL) lab. His research interests encompass robot learning and autonomy. He is the recipient of an NSF CAREER Award and a senior member of IEEE. Before joining UML, he was a Postdoctoral Fellow with the School of Interactive Computing at Georgia Institute of Technology. He holds a Ph.D. in Robotics, Cognition, andInteractive Technologies from the University of Genoa, in collaboration with the Italian Institute of Technology.



Yan Gu received the B.S. degree in Mechanical Engineering from Zhejiang University in 2011 and the Ph.D. degree in Mechanical Engineering from Purdue University in 2017. She joined the faculty of the School of Mechanical Engineering at Purdue University in 2022. Her research interests include nonlinear control, hybrid systems, legged locomotion, and wearable robots. She was the recipient of the NSF CAREER Award in 2021 and ONR YIP Award in 2023.