Expressive Whole-Body Control for Humanoid Robotic Motions Using Model Predictive Control and Reinforcement Learning

SurveyForge

Abstract— Expressive whole-body control for humanoid robotic motions leverages advanced control and learning technologies to achieve fluid, stable, and adaptive behaviors in dynamic tasks. This survey reviews state-of-the-art methods combining Model Predictive Control (MPC) and Reinforcement Learning (RL), emphasizing their complementary strengths in constraint-based trajectory generation and policy generalization. MPC ensures real-time feasibility, stability, and physical compliance, while RL enables robust adaptation across unstructured, high-dimensional environments. Hybrid MPC-RL approaches demonstrate breakthroughs in combining high-level adaptability with safety-critical decision-making, supporting applications such as multi-contact locomotion, gesture-based interaction, and task-switching. Specialized datasets for human-like motion imitation, advanced solver architectures, and hierarchical frameworks further enhance whole-body expressivity by reducing computational bottlenecks and bridging the simulation-to-reality gap. Challenges remain in resolving energy efficiency constraints, scalability for high-degree-of-freedom (DoF) systems, and standardized benchmarks for cross-framework evaluations. Promising future directions include the integration of biomechanics-inspired models, multimodal sensor fusion, and probabilistic planning to unify stability, naturalness, and adaptability. This review underscores the transformative potential of expressive humanoid robots in domains such as healthcare, entertainment, and human-robot collaboration, while identifying critical gaps and research opportunities to drive innovation in adaptive whole-body control.

Index Terms—Hybrid control frameworks, Expressive humanoid robotics, Reinforcement learning synergy

1 Introduction and Foundations

The field of humanoid robotics stands at the nexus of control systems, artificial intelligence, and biomechanics, aspiring to create machines that mirror the versatility, expressiveness, and adaptability of human motion. Central to this pursuit is the concept of expressive whole-body control, a paradigm that focuses not only on the dynamics of locomotion or manipulation but also on achieving fluid, natural, and context-aware movements. This section delves into the foundational principles of expressive whole-body control, emphasizing the challenges posed by the high dimensionality and real-time constraints inherent to humanoid systems, while also introducing the potential of leveraging Model Predictive Control (MPC) and Reinforcement Learning (RL) as synergistic frameworks to tackle these challenges.

Expressive whole-body motion encompasses a spectrum of behaviors where humanoid robots must navigate their environment while coordinating multiple degrees of freedom (DoFs) with precision and adaptability. Achieving such motions is pivotal across numerous applications—ranging from healthcare and eldercare assistance, where intuitive and empathetic interactions are crucial [1], to entertainment and creative industries, where dynamic emotive motions are highly valued [2]. Compared to traditional control paradigms, expressive control places equal emphasis on balancing physical feasibility and biomechanical accuracy, ensuring robot-generated movements resonate with natural human characteristics [3]. However, the complexity of humanoid systems introduces unique challenges, such as computational overhead, high-dimensional state-action spaces, physical constraints, and adaptability in unstructured environments.

MPC stands out as a robust, optimization-based approach for real-time trajectory generation and control under complex dynamic constraints. Its ability to predict future system behavior over a finite time horizon while considering the robot's limitations, such as torque bounds and joint constraints, has positioned it as a cornerstone for stability and adaptability in humanoid control. For instance, Nonlinear MPC frameworks have been employed successfully to accommodate multi-contact dynamics, allowing robots to navigate complex terrains while dynamically managing their center of mass (CoM) and supporting forces [4]. Similarly, mobility in constrained environments, such as dynamic stair climbing or locomotion across uneven surfaces, has been effectively demonstrated using MPC in conjunction with explicit constraint handling mechanisms [5], [6]. However, MPC is inherently limited by the computational expense of solving optimization problems in high-dimensional systems within real-time constraints, necessitating advancements in solvers and hierarchical formulations [7].

In complementary fashion, RL offers a data-driven, reward-driven framework to achieve policy generalization and behavioral learning. RL approaches have demonstrated significant advancements in synthesizing robust and diverse humanoid skills, ranging from dynamic locomotion to human-like emotional expressiveness [8], [9]. Unlike MPC, which relies on accurate system models, RL derives policies through interaction with the environment, enabling adaptability in unpredictable or poorly modeled settings.

Notable examples include methods that utilize human motion capture data to train humanoids for complex tasks, overcoming the sim-to-real gap using innovations like domain randomization and robust reward shaping [10], [11]. However, RL's reliance on extensive exploration and the high cost of sample inefficiency in both computation and hardware wear remain critical limitations, particularly for humanoid systems with large action spaces and fine-grained control requirements [12].

Recent advancements have shown that combining MPC and RL can significantly mitigate the limitations of these individual frameworks while exploiting their respective strengths. MPC's constraint satisfaction mechanisms can be integrated into RL-based systems to support safe exploration and policy optimization under physical constraints [13], [14]. Conversely, pre-trained RL policies can augment MPC by refining high-level task planning and providing adaptability to dynamic environments [15]. This synergy unlocks new capabilities for expressive whole-body control, extending its applicability to dynamic, real-world scenarios.

Looking forward, the interplay between model-driven and learning-based approaches holds significant promise for transformative progress in humanoid robotics. By addressing foundational challenges like computational scalability, sim-to-real transfer, and the integration of multimodal sensory inputs, the field is poised to develop systems capable of mastering unpredictable human environments with fluid and natural expressiveness. The remainder of this survey extends this discussion by exploring overarching principles, emerging hybrid frameworks, and specialized applications that highlight the potential of MPC and RL in driving the next generation of expressive humanoid robots.

2 Fundamental Principles of Control and Learning in Humanoid Robotics

2.1 Foundations of Model Predictive Control (MPC) in Humanoid Robotics

Model Predictive Control (MPC) has emerged as a cornerstone methodology in humanoid robotics, offering a principled, optimization-based framework to address the intricacies of dynamic and high-dimensional control tasks. Designed to solve finite-horizon optimization problems at every control step, MPC enables humanoid robots to plan trajectories, predict future system states, and react adaptively to environmental changes—all while respecting critical physical and operational constraints. Its application to humanoid systems has reshaped the landscape of robot locomotion, manipulation, and whole-body balancing, rendering it indispensable in achieving dynamic behaviors under real-time constraints.

At its core, MPC operates by solving an optimal control problem (OCP) over a finite prediction horizon, seeking to minimize a cost function subject to the multivariate dynamics of the robot and its operational constraints. Given the high degrees of freedom (DoFs) in humanoids, the system dynamics are typically modeled using simplified representations like the Linear Inverted Pendulum Model (LIPM) [16] or centroidal dynamics [17]. These abstractions allow efficient computation while incorporating stability constraints, such as ensuring Zero Moment Point (ZMP) remains within the support polygon—a key requirement for maintaining balance [18]. However, as the environmental or task complexity increases, such simplified models may fail to capture the nonlinearities inherent in full-body interactions, motivating the adoption of whole-body MPC frameworks using nonlinear dynamics [19].

One of the most profound strengths of MPC lies in its ability to handle constraints systematically. For humanoid systems, this includes kinematic limits, torque bounds, contact force constraints, and friction cone conditions [4]. By explicitly encoding these requirements within the optimization problem, MPC enables humanoid robots to perform coordinated whole-body motions such as climbing stairs [1], executing agile maneuvers [20], or adjusting to dynamic terrain. Furthermore, MPC frameworks have been extended to model contact transitions explicitly, solving for optimal timing, location, and sequences of contacts without prespecification [21]. These capabilities are pivotal in enabling humanoid robots to navigate unstructured environments and execute multi-contact loco-manipulation tasks [22].

Nevertheless, the practical deployment of MPC in real-time applications presents significant computational challenges due to its simultaneous consideration of high-dimensional dynamics, long prediction horizons, and intricate constraints. Advanced methodologies, such as hierarchical MPC and frequency-shaped cost functions, have been developed to mitigate these challenges [23], [24]. Hierarchical approaches decouple planning into layers—often dividing high-level trajectory optimization from low-level whole-body control—to balance computational complexity and precision [25]. Methods employing frequency-domain regularizations further enhance robustness by accommodating actuator and contact bandwidth limitations while preserving computational tractability [23]. Additionally, hardware advancements such as GPU-accelerated solvers have enabled MPC to operate at rates exceeding 100 Hz, even when solving nonlinear whole-body problems—making real-time implementations increasingly feasible [7].

Despite these advances, there are notable limitations. Model inaccuracies, such as those arising from unmodeled actuator dynamics or environmental uncertainties, can degrade MPC performance, leading to infeasible solutions in practice [26]. Techniques leveraging probabilistic models and stochastic optimization have been introduced to tackle these uncertainties by incorporating probabilistic guarantees for constraint satisfaction [26]. Additionally, combining MPC with learning-based methods, such as reinforcement learning (RL), holds promise for addressing long-term adaptability while maintaining physical safety through constraint enforcement [13].

Emerging research is keenly focused on synergizing MPC with whole-body trajectory optimization for humanoid robots. The seamless integration of contact dynamics, adaptability in high-dimensional environments, and robust learning frameworks heralds the next phase of human-centric robotic applications. As computational resources advance, the emphasis on scalable formulations using sparse optimizations [21] and adaptive frameworks [14] will remain critical in bridging the gap between theoretical robustness and practical performance. Ultimately, MPC continues to establish itself as a versatile and foundational tool in advancing the expressiveness and functionality of humanoid robotics for diverse real-world tasks.

2.2 Reinforcement Learning (RL) as a Framework for Humanoid Skill Learning

Reinforcement Learning (RL) has emerged as a compelling paradigm to enable humanoid robots to autonomously acquire complex, task-relevant behaviors through interaction with their environment. In contrast to optimization-driven methods like Model Predictive Control (MPC), which depend heavily on accurate dynamic models and pre-specified constraints, RL leverages experience-driven exploration to generate adaptive policies for intricate tasks. This learning-driven approach proves particularly advantageous for humanoid robotics, given the challenges posed by high degrees of freedom (DoFs), non-linear dynamics, and unstructured, evolving environments that often outstrip the capabilities of traditional control mechanisms.

Recent advancements in RL have showcased its potential to synthesize diverse and expressive humanoid motions, from walking and jumping to fine manipulation and socially interactive behaviors. At its foundation, RL aims to optimize a policy of mapping actions to environmental states that maximizes a cumulative reward. This reward-driven framework enables RL to evolve task-specific and adaptable behaviors, demonstrated in applications such as robust bipedal locomotion and dynamic manipulation tasks, bolstered by progress in actor-critic and policy-gradient algorithms [27].

A critical aspect of RL design is reward shaping, which significantly influences its ability to efficiently converge on desirable behaviors while avoiding spurious or unsafe solutions. Poorly designed or sparse rewards can result in sample inefficiency or counterproductive emergent behaviors, such as energy-inefficient movements or excessive conservatism in navigating dynamic environments. To mitigate these challenges, recent works propose hierarchical and task-decomposed reward structures tailored to the unique demands of humanoids [28]. For example, in tasks involving dynamic locomotion over uneven terrain, rewards that concurrently prioritize energy efficiency, stability, and locomotion fluidity have successfully guided RL algorithms toward optimal policy learning [29].

Despite its versatility, RL faces notable limitations, particularly in the realms of sample efficiency and robustness for deployment in real-world environments. Training RL algorithms to handle high-dimensional humanoid systems commonly necessitates millions of simulation episodes, a process complicated by the inherent discrepancies between simulated and physical platforms. Domain randomization—a technique involving variations in simulation parameters such as friction, mass, and sensory noise—seeks to close this sim-to-real transfer gap by exposing policies to a wide range of uncertainties during training [30]. However, transferring these policies to physical robots remains a significant challenge, often requiring additional fine-tuning to accommodate unmodeled dynamics or unexpected variability in physical systems. Emerging approaches that integrate fine-tuned dynamics models with online adaptive learning hold promise in incrementally improving the performances of RL policies post-deployment [14].

One promising strategy to enhance RL's training efficiency is the use of motor primitives and compact action representations, which constrain policy exploration to physically feasible subspaces, dramatically reducing the search space [31]. Additionally, curriculum learning, an incremental approach where simpler tasks (e.g., standing) gradually transition to more complex challenges (e.g., object manipulation during walking), has been shown to accelerate the development of robust and versatile humanoid behaviors [27]. These hierarchical frameworks further extend to task-level decomposition, where higher-level RL policies plan adaptive sequences of actions while lower-level controllers ensure precise motor execution [30].

As humanoid robots increasingly interact in human-centric and unstructured environments, multi-modal learning has become a focal area. By integrating visual, proprioceptive, and tactile feedback, RL-trained agents demonstrate improved adaptability and responsiveness, especially in collaborative human-robot tasks [28]. Multi-modal frameworks not only enhance physical task performance but also foster human-aligned and socially intuitive behaviors. In this domain, innovations such as attention-based reward models and probabilistic reasoning mechanisms have further enriched the expressiveness and decision-making capacities of RL-driven robots [32].

Looking ahead, the integration of RL with complementary frameworks like MPC is poised to address RL's inherent limitations while capitalizing on its adaptability. Hybrid RL-MPC frameworks offer a balanced combination of safety-constrained learning and adaptive policy control, bridging the gap between the performance boundaries of optimization-driven and learning-driven paradigms [29]. Through robust sim-to-real techniques, scalable architectures for high-DoF systems, and improved adaptation to real-world variability, RL is set to play an integral role in advancing humanoid robotics toward robust, dynamic, and socially expressive behavior in complex environments.

2.3 Synergies Between MPC and RL in Humanoid Control

The integration of Model Predictive Control (MPC) and Reinforcement Learning (RL) has emerged as a promising strategy for humanoid robotics, leveraging the complementary strengths of these paradigms to achieve robust, adaptive, and expressive control. MPC excels in short-term optimization and precise constraint handling, while RL offers long-term adaptability and the ability to explore high-dimensional decision spaces. Together, their synergy enables hybrid approaches that can operate effectively in dynamic and uncertain environments, bridging the gap between reactive, optimization-based control and policy-driven adaptability.

A key advantage of coupling MPC with RL lies in addressing RL's inherent limitations during learning, particularly with safety and constraint compliance. RL often explores unsafe states during training due to its exploratory nature, which can lead to poor performance or hardware risks when applied to humanoid systems. By integrating MPC's predictive capabilities, RL frameworks can leverage constraint-based safeguards. For instance, MPC can dynamically enforce stability, balance, or torque constraints during RL training, guiding the learning agent to remain within physically plausible boundaries. Techniques such as constrained RL frameworks, where the exploration space is bounded by MPC-generated feasible sets, have been shown to improve both safety and sample efficiency in real-world tasks [33]. These methods result in policies that balance performance and feasibility, mitigating catastrophic exploration failures.

Conversely, RL augments MPC by addressing its computational limitations and enabling broader adaptability in high-dimensional and nonlinear tasks. Traditional formulations of MPC struggle with solving optimization problems in real-time for complex humanoid models, particularly in multi-contact scenarios or when precise dynamics are unavailable. RL, with its pre-trained policies, can serve as an auxiliary layer to predict task-relevant objectives or refine trajectory planning under computational constraints, thus accelerating MPC solvers [13], [34]. For example, RL-trained policies can provide action priors or precondition target trajectories for MPC, reducing the problem complexity during critical time frames. This hybrid dynamic has been successfully employed in tasks such as bipedal locomotion and reactive balance recovery, where RL enhances the adaptability of MPC to external disturbances and varying terrains [35], [36].

Moreover, hybrid MPC-RL frameworks demonstrate superior scalability in hierarchical control systems, where low-level controllers are governed by MPC for precise actuation, while high-level planning and decision-making leverage RL's policy flexibility. In such configurations, MPC handles short-horizon control problems, including balance and contact stabilization, while RL oversees high-level tasks such as gait modulation or environmental interaction. This architectural decomposition not only ensures real-time responsiveness but also emphasizes task specialization, as evidenced by studies employing hierarchical approaches to control humanoids during complex parkour-like maneuvers or dexterous loco-manipulation tasks [9], [37].

Despite these advances, several challenges persist. The success of hybrid MPC-RL approaches largely depends on the seamless integration of their respective modules, particularly in ensuring synchronization and compatibility between deterministic optimization and policy-driven exploration. Computational trade-offs remain an active area of research, as deploying RL-adaptive MPC in real-time scenarios often demands significant resources, including advanced solvers and hardware acceleration [13], [35]. Furthermore, bridging the sim-to-real gap presents a crucial hurdle. While MPC offers robustness against model inaccuracies, RL-generated policies trained in simulation often require domain randomization or fine-tuning for hardware deployment [38], [39].

Future research directions include the exploration of probabilistic MPC frameworks combined with policy-driven uncertainty estimation from RL to better handle dynamic environments and stochastic disturbances. Additionally, leveraging advances in large-scale datasets and motion libraries could enhance knowledge transfer between MPC and RL, fostering human-like expressiveness and interaction capabilities in humanoid robots [3], [40]. As these hybrid systems evolve, their potential to address increasingly intricate and human-centric tasks offers a significant leap toward truly adaptive and versatile humanoid robotics.

2.4 Challenges and Emerging Directions in Control Paradigms

The integration of Model Predictive Control (MPC) and Reinforcement Learning (RL) in humanoid robotics unlocks intriguing prospects for achieving both robust control and adaptive performance. Nonetheless, significant challenges persist in scaling these paradigms to meet the high-dimensional, real-time requirements of humanoid robots while effectively navigating uncertain, dynamic environments. This section delves into these pressing challenges and explores emerging research directions to overcome them.

A key challenge lies in the scalability of control for systems with high Degrees of Freedom (DoFs). Humanoid robots, characterized by their intricate kinematics and complex dynamic structures, necessitate computationally efficient solvers to address large-scale optimization problems in real-time. While existing MPC approaches have demonstrated effectiveness in handling stability and multi-contact constraints [41], extending these methods to achieve fine-grained control of full-body dynamics often encounters computational bottlenecks and struggles with nonlinear modeling fidelity [31]. Simultaneously, RL techniques for high-DoF systems face inefficiencies in exploring vast action spaces. Although strategies such as hierarchical policy learning [30] and action-space dimensionality reduction through motor primitives [42] have shown promise, practical implementation remains hampered by sample inefficiency and the difficulty of ensuring robustness across diverse tasks.

Another significant limitation is handling environmental uncertainty and maintaining system robustness. MPC's dependency on high-fidelity system models makes it vulnerable to external disturbances and deviations from assumed dynamics [14], particularly in dynamic scenarios such as uneven terrain or unexpected forces during walking and manipulation. Although RL introduces adaptability through learned policies, its robustness under physical inaccuracies remains a work in progress. Recent advancements, such as combining RL with uncertainty-aware modeling approaches like Bayesian multi-task learning [43] and probabilistic ensembles [44], offer potential for enhanced robustness. However, integrating uncertainty quantification into control architectures that can operate in real-time remains an unresolved challenge requiring further investigation.

The issues of energy efficiency and resource optimization further constrain the potential of humanoid robotics. Expressive and dynamic motions in humanoid systems inevitably demand considerable power, limiting operational timeframes and hardware longevity. The computational demands of MPC exacerbate these challenges, despite efforts to streamline formulations through frameworks such as cascaded-fidelity MPC [6] or hardware acceleration solutions [45]. On the RL front, energy-efficient methodologies that focus on minimizing power consumption during extended task executions are still relatively underexplored, providing an interesting avenue for future research.

The integration of multimodal sensory feedback into control paradigms adds another layer of complexity. To operate effectively in dynamic and human-centric environments, humanoid robots must be capable of synthesizing inputs from visual, proprioceptive, and tactile sensors in real time. While RL-based multi-modal learning frameworks show promise [46], particularly in collaborative or dexterous manipulation tasks, MPC-based methods involving multidimensional sensory inputs face increased computational burdens [47]. Balancing the trade-offs between robust sensory integration and computational feasibility remains a significant hurdle for both frameworks.

While the synergies between MPC and RL present compelling opportunities, achieving an effective balance between their strengths poses a noteworthy challenge. MPC excels in precision and constraint enforcement, while RL provides adaptability and policy refinement in unstructured environments. Achieving seamless integration requires addressing computational trade-offs and mitigating latency in real-time systems [48]. Promising techniques, such as leveraging RL-enhanced state predictions to initialize MPC optimization [13], highlight potential pathways for reducing computational complexity while retaining the complementary benefits of both paradigms.

Looking ahead, addressing these challenges calls for advancements in adaptive hierarchical control mechanisms, scalable state representations for high-DoF systems, and unified frameworks that integrate physics-based and learning-based approaches. Incorporating advanced computational tools, such as GPU-parallelized solvers and stochastic optimization techniques, into hybrid systems [49] may also accelerate progress. Furthermore, exploring biophysically inspired approaches—such as leveraging human biomechanics to synthesize natural and expressive motions [10]—represents a promising frontier that could bridge robustness and adaptability.

In conclusion, while MPC and RL individually offer transformative potential, their integration into humanoid robotics remains an ongoing challenge. Overcoming the key obstacles and leveraging emerging approaches will drive the field closer to achieving robust, adaptive, and expressive whole-body control for real-world humanoid robots.

3 Dynamics and Representation of Expressive Whole-Body Motions

3.1 Modeling Humanoid Dynamic and Kinematic Systems

The modeling of humanoid dynamic and kinematic systems lies at the heart of synthesizing responsive, expressive, and robust whole-body motions. Such modeling serves as the foundational framework for understanding the interactions of forces, torques, and motion trajectories within the humanoid robot's structure. The overarching aim is to achieve accurate representation of both global dynamics, such as center of mass (CoM) behavior, and local kinematics, such as limb coordination and joint compliance. Critical to this process is the seamless integration of dynamic and kinematic descriptions to capture the full complexity of high-degree-of-freedom systems.

Multi-limb coordination demands sophisticated representations of force and motion dynamics across the robot's limbs to maintain balance, stability, and motion fluidity. A significant body of research leverages hierarchical optimization to describe and synchronize whole-body movements. Hierarchical inverse dynamics approaches have shown promise in translating task requirements into coherent motion plans by resolving force and acceleration interactions between the limbs [50]. For instance, hierarchical controllers manage multi-task scenarios, such as balancing and manipulation, by defining prioritized objectives that ensure physical feasibility and stability under disturbances [51]. However, while these controllers simplify real-time implementation, their computational complexity often proves limiting in highly dynamic scenarios. Trade-offs arise between efficiency, solved via reduced-order formulations like centroidal momentum dynamics [52], and accuracy achieved through full-body models.

Accurate representation of CoM trajectories is especially critical as it directly influences stability during locomotion, manipulation, or dynamic tasks. To model CoM effectively, simplified dynamic frameworks such as the Linear Inverted Pendulum Model (LIPM) have been widely adopted for their computational efficiency and ability to convert CoM motion planning into tractable optimization problems [16], [21]. However, such simplifications can fail to capture intricate nonlinearities associated with rapid movements, leading to conservative predictions. To address this, centroidal dynamics models, which incorporate angular momentum and variable inertia, provide a superior alternative for maneuvering complex terrains or executing dynamic behaviors [26], [53]. By integrating CoM modeling with moment-based constraints, these models achieve robust stability in scenarios involving uncertain interactions, such as foot slippage or external impacts.

Achieving consistency between dynamic and kinematic representations poses another fundamental challenge. Dynamic representations focus on forces and torques within system constraints, while kinematic frameworks describe motion trajectories through joint-space configurations. Effective synthesis is necessary to ensure that high-level trajectories derived from dynamic models can be physically expressed through the robot's actuators. Whole-body optimization frameworks, such as model-predictive approaches, excel in bridging these domains by blending task-specific dynamics with kinematics constraints [19]. Hence, these approaches allow for seamless transitions between motion tasks, such as switching from walking to grasping, while preserving biomechanical coherence. However, such integration greatly increases computational costs, warranting the development of hierarchical techniques [25] and sparse formulations [7].

Efficient state representation for high-dimensional systems further amplifies the operational viability of humanoid motion modeling. Dimensionality reduction techniques, such as embedding where motor primitives capture joint-space features [54], enable manageable representations for real-time applications without sacrificing expressiveness. Similarly, latent-variable models have shown their potential in encoding diverse motion styles, enabling humanoids to learn, reuse, and generalize motion behaviors [55]. Despite these advances, challenges surrounding real-time performance, robustness under sensor noise, and integration with feedback mechanisms remain pressing.

Emerging trends and innovations highlight the need for structured and adaptive modeling frameworks. Methods such as nonlinear centroidal dynamics modeling [31] and hybrid representations combining dynamics and probabilistic control models [56] have shown promise in expanding the domains of feasible motion. Future research should prioritize unifying dynamic and kinematic models with multimodal sensory feedback and adaptive control schemes to achieve finer motion expressivity and reliability across broader operational conditions.

3.2 Motion Primitives and Feature Representations

Motion primitives and feature representations serve as pivotal tools in the modeling, synthesis, and control of expressive humanoid motions, bridging the gap between complex motor control and adaptable behavior design. By decomposing humanoid tasks into modular and reusable components, these frameworks provide the foundation for synthesizing adaptive, fluid, and context-aware movements. This subsection elaborates on the approaches and trends in motion primitives and feature representations, emphasizing their relevance in achieving dynamic and expressive whole-body motions.

Dynamic Movement Primitives (DMPs), inspired by biological systems, have demonstrated their utility in encoding flexible motion trajectories through nonlinear differential equations. By parameterizing motion behaviors with adaptable weights and goal states, DMPs enable humanoid robots to execute dynamic tasks, such as multi-contact locomotion and manipulation, with compact yet robust representations [19], [57]. Their capacity to preserve essential dynamics ensures smooth and reliable motion synthesis. However, DMPs' fixed structure often limits their applicability in scenarios involving sudden contact changes or hybrid events, necessitating more versatile formulations for such tasks.

Complementary to DMPs are optimization-based, trajectory-centric approaches, which derive motion primitives that encapsulate contact dynamics while maintaining task adaptability. These methods generate reference trajectories grounded in kinematic and dynamic models, enabling seamless transitions across interaction phases [4], [45]. For example, contact-implicit frameworks optimize trajectory and contact sequences simultaneously, overcoming complexities associated with task-specific transitions. Despite being computationally demanding, advancements in solver efficiency and convex relaxations have significantly improved the real-time applicability of such frameworks for humanoid platforms [58], [59].

Feature representations play an equally critical role in generating expressive humanoid motions by parameterizing characteristics such as amplitude, frequency, and phase. These representations allow humanoid robots to tailor behaviors like adaptive walking speeds, dynamic gestures, or stable locomotion patterns for varying contexts. For instance, integrating centroidal momentum dynamics into feature-based models supports momentum-conserving motions that enhance stability and energy efficiency while maintaining dynamic feasibility [60]. Such adaptive features enable humanoids to address challenging tasks, such as navigating uneven terrains or performing cooperative actions, by harmonizing expressiveness with physical and biomechanical constraints [4], [5].

Recent advancements highlight the integration of context-aware primitives that proactively incorporate environmental feedback and task-specific constraints into motion representations. This approach blends sensory feedback with trajectory optimization techniques, facilitating robotic adaptation to unstructured or dynamic conditions. For example, spring-contact models coupled with predictive control frameworks have enabled humanoid robots to reactively adjust their motion primitives in response to disturbances in real-time [61], [62]. Additionally, the advent of learning-based methods, such as neural network-derived feature representations, has further expanded adaptive capabilities. By capturing latent task-relevant variability, these models enable humanoid robots to generalize behavior across tasks with improved training efficiency and deployment flexibility [27], [32].

The modularity and reusability of motion primitives greatly enhance the diversity and adaptability of humanoid behaviors. Modular primitives facilitate seamless transitions between tasks, such as alternating between walking, gesturing, and object manipulation. However, the complexity of coordinating hybrid systems underscores the need for hierarchical frameworks. Hierarchical architectures, by decoupling high-level planning from low-level motor execution, support smooth behavioral transitions and complementary task achievement [30], [63].

While the development of motion primitives and feature representations has advanced considerably, challenges persist in ensuring effective generalization across diverse tasks, achieving scalable computational performance, and reducing the simulation-to-reality discrepancy. Future exploration should center on hybrid models that fuse data-driven learning with trajectory optimization, develop real-time task-adaptive frameworks, and integrate multimodal feedback for richer interactions. Solving these challenges will unlock broader potential for human-like, expressive, and versatile humanoid robot capabilities across complex real-world scenarios.

3.3 Physical Feasibility and Biomechanical Constraints

Ensuring the physical feasibility and biomechanical consistency of humanoid robot motions is a cornerstone of expressive whole-body control. This subsection explores the principles and approaches for enforcing fundamental physical constraints such as stability, contact force distribution, and joint torque limits, while addressing the challenges of energy efficiency and biomechanical modeling for naturalistic performance.

Stability and compliance are primary concerns in generating physically feasible motions, particularly in dynamic and interactive environments. Stable motion demands accurate modeling of the Zero Moment Point (ZMP) and Center of Pressure (CoP), which are widely used to ensure balance during locomotion and manipulation tasks. Techniques such as Model Predictive Control (MPC) have been successfully applied to predict ZMP trajectories and dynamically adjust foot placement under changing conditions, ensuring robust recovery from disturbances like external forces [35], [36]. Additionally, compliance strategies leverage hierarchical feedback controllers to achieve adaptable behavior in response to external forces, maintaining system integrity and avoiding rigid, unnatural motions. Reinforcement learning (RL) frameworks, when integrated with physically consistent constraints, have also demonstrated human-like recovery behaviors such as ankle and hip push-off strategies [64].

Contact dynamics and force distribution play a critical role in enabling humanoids to interact with the environment safely and effectively. The management of reaction forces and moments at contact points, such as foot-ground interactions during locomotion, ensures stability and prevents overloading joints or inducing slip. Whole-Body Locomotion Controllers (WBLCs), for instance, integrate dynamic task prioritization with inequality constraints to compute reaction forces efficiently [65]. Furthermore, frameworks such as Dynamical Complementarity Constraints explicitly model robot-environment interaction to optimize force distributions and adapt to complex hybrid dynamics, including switching contacts during walking [19]. These approaches highlight the importance of bridging quasi-static assumptions with the dynamic realities of physical robotics.

Incorporating joint torque and kinematic constraints is essential to mechanically viable and biologically-plausible motions. State-of-the-art methods integrate these constraints into optimization-based frameworks, such as Differential Dynamic Programming (DDP) and MPC, which inherently account for joint angle limits, actuator torque thresholds, and kinematic reachability. By reflecting structural constraints during motion generation, these models prevent damage to the robot's hardware and foster smoother, human-like motion [66]. Reinforcement learning approaches are particularly enhanced when augmented with biomechanics-inspired reward functions, allowing policies to discover torque-efficient strategies autonomously [40].

Energy efficiency further juxtaposes physical feasibility with the need for sustainable operation. Generating expressive behaviors often necessitates real-time computation of high-dimensional control parameters, which can lead to excessive energy consumption and overheating of actuators. Optimization techniques such as Bayesian optimization have been employed to tune motion generation costs, focusing on minimizing energy expenditure while ensuring robustness under uncertainty [34]. Additionally, hybrid frameworks that combine physics-informed control models with data-driven strategies provide an avenue for optimizing energy efficiency without sacrificing adaptability [13].

Emerging trends point toward achieving greater biomechanical fidelity in humanoid motions. Neural probabilistic motor primitives offer a modular structure to encode diverse motion patterns while maintaining compliance with torque and joint constraints [54]. Similarly, motion datasets harvested through imitation learning or physics-based modeling are used to establish priors for generating naturalistic motions without mechanical violations [3], [8]. Advances in sim-to-real transfer pipelines further ensure that learned motions retain biomechanical feasibility when deployed on physical robots, often leveraging domain randomization techniques to account for inaccuracies in physical models [11].

In conclusion, the integration of biomechanical constraints and models significantly enhances the feasibility and naturalness of humanoid robot motions. Future directions include advancing hybrid control-learning paradigms to holistically balance physical constraints, computational efficiency, and task expressiveness. Moreover, increasing adoption of multimodal sensory inputs—such as tactile and visual feedback—can enable robots to adaptively refine their actions in real time, ensuring robustness and compliance in unstructured environments. The realization of scalable, energy-efficient, and biomechanically consistent motions represents a crucial step toward achieving humanoid robots with lifelike agility and versatility across real-world applications.

3.4 Natural Motion Synthesis through Optimization and Learning

The synthesis of natural motion for humanoid robots embodies a fusion of physics-based optimization and learning-based techniques, emphasizing key attributes such as smoothness, responsiveness, and human-like dynamism. Positioned at the intersection of control theory and artificial intelligence, this multidisciplinary approach seeks to balance biomechanical realism with computational efficiency in high-dimensional humanoid systems. The complementary strengths of these paradigms have enabled significant strides in crafting motions that are not only physically feasible but also contextually expressive and adaptable to varied scenarios.

Optimization-based methods lie at the heart of deterministic motion generation, formulating constrained optimization problems rooted in the physical dynamics of humanoids. These methods, exemplified by Model Predictive Control (MPC), excel in enforcing stability and precise constraint-handling, ensuring real-time adaptability for tasks such as locomotion and object manipulation [67]. Advanced frameworks augment these capabilities by incorporating centroidal dynamics, thus enabling coordinated multi-limb motions and balance regulation through center-of-mass (CoM) control [68]. Hierarchical MPC frameworks, in particular, optimize both motion smoothness and dynamic consistency by introducing higher-fidelity models incrementally across prediction horizons [6]. However, despite their precision, these methods often face challenges related to high computational demands and limited adaptability in unstructured and unpredictable environments. As a result, achieving real-time performance in complex scenarios frequently necessitates compromises such as solver approximations.

Reinforcement learning (RL), in contrast, introduces a data-driven paradigm to motion synthesis. By interacting with simulated or physical environments, RL paradigms enable humanoid robots to autonomously discover control policies for complex, expressive behaviors including dancing, navigating rugged terrain, or performing gestural communication [10]. Effective frameworks such as masked multi-modal controllers have demonstrated robust capabilities in handling diverse real-world tasks like adaptive standing, walking, and partial-body motion imitation [46]. However, RL training often relies heavily on simulation, producing policies that require careful transfer techniques such as domain randomization to bridge the simulation-to-reality gap [69]. While RL excels in adaptability and fine-tuning for task-specific applications, its inherent challenges, including sampling inefficiency, intricate reward design, and the lack of clear interpretability in learned policies, underline the trade-offs between adaptability and reliability.

Hybrid approaches that combine the strengths of optimization and reinforcement learning present a promising solution to overcome the limitations of using either method exclusively. By aligning MPC's structured trajectory optimization with RL's flexible policy discovery, these frameworks achieve a balance between task adaptability and motion precision. For example, RL-augmented MPC frameworks leverage learned corrective policies to refine motion parameters such as foot placement or multi-contact interaction strategies in dynamic settings [29]. Conversely, the integration of MPC within RL training environments enhances safety and physical consistency by embedding constraint-aware trajectory sampling into the learning pipeline, thereby reducing trial-and-error exploration during policy generation [48]. Hybridization facilitates bidirectional advantages: optimization frameworks benefit from RL's generalization capabilities, while RL gains structure and efficiency from robust optimization constraints.

An overarching challenge in synthesizing natural motion lies in achieving computational tractability while preserving motion quality in systems with high degrees of freedom (DoFs). Recent efforts have sought to address this through innovations such as nullspace resolution techniques for inverse dynamics MPC [70] and hierarchical learning paradigms that reduce reliance on high-frequency policy updates [30]. These advancements streamline computational requirements by prioritizing critical dynamic constraints, minimizing redundant joint-space computations, and employing neural approximations for accelerated motion prediction [27].

Future research must continue to deepen the integration of biomechanics-inspired intuition with data-driven motion synthesis frameworks. Enhanced biomechanical modeling can improve the transferability and robustness of learned controllers by embedding characteristics such as natural torque patterns and energy-efficient gaits. Furthermore, probabilistic modeling fused with hybrid learning techniques holds potential for enabling robust exploration and safe adaptation in volatile environments [71]. Recognizing the growing importance of perception-driven control, innovations such as hierarchical world models for visual motion synthesis show promise in advancing interactive humanoid control [28]. Incorporating visual feedback into these frameworks not only bridges perception and action but also significantly enriches context-awareness for complex tasks.

In conclusion, the synergy between optimization-based control and reinforcement learning continues to propel advancements in the field of humanoid robot motion synthesis. Progress in computational efficiency, biomechanical fidelity, and adaptable learning paradigms paves the way for humanoid systems to realize the fluidity, expressiveness, and robust adaptability that define human movement, bringing humanoid robots closer to seamless interaction within dynamic, real-world environments.

4 Advancements in Model Predictive Control for Expressive Humanoid Motion

4.1 Advances in Real-Time Computational Strategies for MPC

The surge in the application of Model Predictive Control (MPC) for humanoid robotics, particularly for expressive and dynamic motions, has necessitated the development of computational strategies that achieve real-time performance. Humanoid robots impose unique challenges, given their high degrees of freedom (DoFs), nonlinearity, and dynamic complexity. This section focuses on the latest advancements in computational techniques that enable MPC to meet these demands, emphasizing algorithmic innovations, hardware acceleration, and the efficient exploitation of problem structures.

A foundational challenge in implementing real-time MPC for humanoid robots is the computational intractability resulting from the high-dimensional state and control spaces, especially when accounting for nonlinear dynamics, multi-contact scenarios, and strict constraints on stability and torque limits. Recent efforts toward addressing this issue exploit problem sparsity and structure, as demonstrated in hierarchical and sparse formulations. Approaches that leverage sparsity, such as those presented in [21], allow solvers to focus computational resources on critical regions of the control problem, greatly reducing solution time while maintaining physical fidelity. These sparse formulations are crucial for humanoid tasks requiring seamless transitions between dynamic motions, such as contact-rich manipulation or locomotion over uneven terrain.

Another significant line of progress involves preconditioning and warm-start strategies to expedite iterative solvers by leveraging prior solutions. Caching and reusing solutions from previous optimization cycles not only improves convergence rates but also ensures responsiveness in dynamically evolving humanoid tasks, as highlighted in [51]. By initializing MPC with solutions that reflect subtle changes in task dynamics, such strategies allow humanoid robots to maintain balance and stability under unforeseen disturbances, such as external pushes or terrain irregularities.

The advent of hardware acceleration, particularly through Graphics Processing Units (GPUs) and custom accelerators, has further revolutionized the ability to deploy MPC in real-world scenarios. GPU-accelerated optimization is particularly effective for high-frequency updates, enabling MPC to run at rates exceeding 90 Hz on high-DoF humanoid platforms [7]. The parallelization capabilities of GPUs have been demonstrated to handle the computationally intensive tasks of solving quadratic programs (QP) with constraints imposed by contact forces, stability, and joint kinematics.

Advancements in solver design also play a key role in improving both the speed and robustness of MPC implementations. Methods that rely on fast iterative solvers, such as those based on the Alternating Direction Method of Multipliers (ADMM), prioritize efficiency by iteratively decomposing large optimization problems into smaller, more manageable subproblems [4]. Additionally, frequency-shaped cost functions, as introduced in [23], extend robustness by incorporating actuator bandwidth limitations directly into the optimization, ensuring smooth and realistic trajectories suitable for the physical capabilities of humanoid hardware.

Emerging formulations, such as cascaded fidelity designs, offer promising avenues for handling the trade-off between model accuracy and computational feasibility. For instance, [6] proposes a hierarchical relaxation of constraints along the optimization horizon, allowing more efficient computation at coarse time scales while maintaining fine-grained accuracy in immediate predictions. Such systems demonstrate superior performance in synthesizing complex motions like dynamic barrel rolls and acrobatic maneuvers.

Despite these advances, significant challenges persist in closing the gap between offline optimization and real-time control on hardware. Many MPC frameworks still struggle with scaling gracefully to humanoids with higher DoFs or under extreme uncertainties, such as those encountered on slippery terrains or during interactions with deformable objects [26]. Techniques that couple MPC with machine learning-based predictive models offer a compelling direction. Learning-augmented MPC frameworks are increasingly capable of capturing nonlinearities and uncertainties from offline data while running near real-time during deployment [72].

In summary, the convergence of algorithmic optimizations, solver enhancements, and hardware acceleration continues to transform MPC from a conceptually robust framework into a computationally practical tool for humanoid robots. Future progress will likely lie in integrating these advancements into unified frameworks that seamlessly manage the complexities of multi-task scenarios and unstructured environments. Additionally, the hybridization of MPC with machine learning may unlock greater predictive accuracy and adaptability, enabling broader applications of MPC in expressive humanoid robotics.

4.2 Handling Complex and Multi-Contact Scenarios via Predictive Control

Model Predictive Control (MPC) has emerged as a versatile framework for addressing the inherent complexities of multi-contact humanoid robot control. Multi-contact scenarios, wherein humanoids engage with their environment through varied and potentially simultaneous contact points, introduce significant challenges due to the hybrid and non-smooth dynamics arising from contact initiation, maintenance, or transition phases. Successfully managing these challenges demands precise modeling of contact forces, stability constraints, and task prioritization to generate dynamic and stable humanoid behaviors. This subsection examines recent advancements adapting MPC methodologies to handle multi-contact interactions effectively, bridging the innovations discussed in the previous subsection on MPC efficiency and the following subsection on handling nonlinearities and uncertainties.

One prominent method for addressing multi-contact scenarios involves integrating explicit contact dynamics into the predictive control framework. By incorporating rigid body dynamics with contact force constraints, MPC can handle multi-phase hybrid systems where transitions between contact states are explicitly modeled. For example, nonlinear MPC formulations embedding contact dynamics have been shown to optimize contact sequences, locations, and timings, enabling real-time trajectory generation for tasks such as walking, pushing, or object manipulation [4], [22]. However, these approaches often suffer from high computational complexity due to the nonlinearity introduced by contact modeling, highlighting the necessity for advancements in solver efficiency similarly emphasized in the previous discussion on real-time MPC.

To mitigate computational challenges, sparse problem formulations and contact-aware approximations have been developed. Methods leveraging convex relaxations, such as Quadratically Constrained Quadratic Programs (QCQPs), provide an efficient means of handling contact forces while maintaining the physical fidelity of the humanoid's motion [73]. These strategies reduce the non-convexity inherent in multi-contact dynamics and have demonstrated their scalability to real-time rates. However, as highlighted in subsequent sections focusing on robustness, these convex relaxations may introduce inaccuracies in highly nonlinear or friction-sensitive interactions, presenting a trade-off between computational efficiency and precise terrain compliance.

The hybrid nature of multi-contact dynamics, involving periodic contact-breaking or impact dynamics, further complicates MPC implementation. Techniques like Hybrid iLQR MPC dynamically adjust contact sequences in response to disturbances, facilitating cohesive planning across a humanoid robot's whole-body dynamics. These methods have proven effective in enhancing MPC's adaptability to large perturbations, extending its versatility for robots operating in unpredictable multimodal environments [59]. Additionally, algorithms embedding time-optimization layers into MPC can dynamically adapt contact switching schedules, improving responsiveness for tasks involving fast-paced transitions [74]. Such approaches demonstrate how advancements discussed in this subsection are aligned with the broader exploration of real-time adaptability and robustness.

Maintaining stability during high-impact contacts presents another critical focus area. Impact-aware MPC frameworks, which anticipate and stabilize systems around potential impact states, have demonstrated robust performance in humanoid applications where contact timing or force predictability is uncertain. These frameworks enhance durability and safety by explicitly incorporating constraints such as peak force bounds and velocity continuity conditions, ensuring resilience to abrupt post-impact dynamics [75]. This aligns closely with the following subsection's emphasis on constraint management as a linchpin in addressing uncertainties.

Task-specific stability constraints embedded within multi-contact MPC further ensure robustness during locomotion and manipulation. For example, criteria based on Center of Mass (CoM) dynamics or momentum constraints enable robust trajectory stabilization, even in scenarios with variable terrain and contact conditions. Multi-contact MPC approaches incorporating defined cones of feasible CoM accelerations have improved planning robustness under stability constraints, enhancing operational reliability during challenging maneuvers [76]. However, these conservative formulations may limit dynamic exploration, presenting a trade-off explored later in the discussion on adaptability to nonlinear conditions.

Future advancements in multi-contact handling within MPC are likely to revolve around its integration with data-driven methodologies. Augmenting physics-based MPC with learned models trained on simulated or empirical contact data, for instance, promises to improve prediction accuracy under uncertain and nonlinear conditions [77]. Similarly, reinforcement learning-augmented frameworks for predicting feasible contact sequences or optimizing interaction forces during tasks like manipulation show potential for enhancing flexibility and generalizability [29]. These hybrid approaches underscore themes of adaptability and computational feasibility seen throughout this survey.

In conclusion, MPC's ability to adaptively manage hybrid and multi-contact challenges makes it an indispensable tool for expressive humanoid control. Recent advancements are pushing the boundaries of dynamic stability, computational efficiency, and environmental adaptability, complementing the trends discussed in prior sections on real-time MPC and foreshadowing future developments in uncertainty management. Sustained innovation in solver technologies, risk-aware formulations, and synergetic integration with learning-based enhancements will be critical to meeting the demands of increasingly intricate and human-centric robotic tasks.

4.3 Enhancing Robustness in Nonlinear and Uncertain Dynamics

Model Predictive Control (MPC) has emerged as a cornerstone for stable and dynamic control of humanoid robots; however, its application to nonlinear and uncertain dynamics remains limited by challenges in model fidelity, disturbance rejection, and real-world adaptability. Enhancing the robustness of MPC under such conditions demands advancements in both system modeling and computational strategies. This subsection delves into adaptive modeling techniques, robust constraint handling, and machine learning integrations that are transforming MPC's capability to address nonlinear and uncertain dynamics in humanoid robotics.

Central to improving MPC's robustness is the development of accurate dynamic models capable of capturing the system's nonlinear behavior and external disturbances. Traditional centroidal and reduced-order models, while computationally efficient, are often limited in representing complex phenomena such as variable inertia, flexible contact forces, or non-smooth dynamics observed during multi-contact motion [19]. Approaches that integrate variable-inertia representations, derived from detailed centroidal dynamics, offer significant improvements in capturing whole-body agility during high acceleration maneuvers and precise interaction with uncertain environments. For instance, the augmented Single Rigid Body Model (aSRBM) used in orientation-aware MPC [35] demonstrates the effectiveness of reasoning about orientation dynamics to stabilize humanoids under external torques, outperforming point-mass simplifications in dynamic tasks.

Handling external disturbances is critical for achieving consistent performance in uncertain scenarios. Explicit disturbance rejection mechanisms have been incorporated into MPC formulations to manage the unpredictable dynamics resulting from environmental variability or external forces. The use of receding horizon strategies to systematically predict and counteract disturbances, as demonstrated in the push-recovery strategies for the iCub platform [36], highlights the efficacy of predictive models in maintaining stability during aggressive perturbations. However, these methods often incur a trade-off between computational efficiency and robustness performance, particularly when disturbances involve stochastic, high-frequency noise.

Incorporating machine learning technologies to enhance predictive accuracy has increasingly been explored as a strategy to address model inaccuracies and operational uncertainties. Data-driven methods enable MPC to generalize across a broader spectrum of dynamics beyond conventional model-based parameterizations. Neural approximations of system dynamics, such as the use of probabilistic models or Gaussian processes, provide a mechanism for uncertainty-aware modeling during motion planning. For example, learning-based enhancements demonstrated in humanoid locomotion frameworks combine offline-trained models with online adaptation to refine contact interactions and dynamic stability in the presence of hybrid dynamics [56]. While effective, these methods present challenges related to integration latency, requiring efficient solver designs to enable real-time decision-making.

Additionally, constraint violation recovery remains a pivotal consideration in ensuring safe operation under nonlinear conditions. Techniques such as viability kernels and dynamic feasibility sets are employed to recover from states near infeasibility without jeopardizing task objectives [65]. By integrating dynamic constraint relaxations within MPC, recent research successfully demonstrates a balance between maintaining stability and advancing locomotion under contact-intensive tasks such as jumping and climbing [78], [79].

Emerging trends suggest that hybrid frameworks combining MPC with reinforcement learning (RL) represent a powerful approach for tackling complexities inherent in nonlinear systems. RL offers adaptability to learn optimal configurations during high-dimensional interactions, while MPC enforces safety-critical constraints during deployment. Strategies involving dual-layer optimization leverage RL to adapt task-specific parameters and MPC to execute real-time trajectory tracking, as seen in frameworks integrating model-based RL with MPC for adaptability in varied terrains [13]. However, hybrid solutions must overcome computational bottlenecks to maintain real-time viability in dynamic scenarios.

In summary, enhancing the robustness of MPC for nonlinear and uncertain dynamics relies on advancements that bridge model precision, real-time computational feasibility, and adaptability to environmental changes. The integration of learning-based methodologies, adaptive constraint handling, and hierarchical frameworks provides promising solutions but necessitates a careful balance of trade-offs. Future research should focus on harmonizing these developments to unlock higher locomotion capabilities, dynamic dexterity, and adaptive safety mechanisms for humanoid robots in real-world, ever-changing environments.

4.4 Task-Specific and Hierarchical Control Schemes

Hierarchical and task-specific control schemes in humanoid robotics effectively address the dual challenges of computational feasibility and motion strategy prioritization, particularly when leveraging Model Predictive Control (MPC). By employing layered control structures, these approaches decouple high-level task planning from low-level trajectory optimization, enabling efficient allocation of computational resources while maintaining adaptability across complex and dynamic scenarios.

Hierarchical frameworks simplify the intricacies of whole-body MPC by breaking control problems into distinct levels, each tailored to specific objectives. The high-level modules handle overarching task goals—such as locomotion, manipulation, or balance—while the low-level controllers refine trajectories to meet these objectives, ensuring adherence to system dynamics and constraints. For example, whole-body trajectory generation leveraging centroidal and kinematic dynamics has proven effective in real-time MPC formulations, where computational demand is managed by isolating and simplifying dynamic complexities [19]. Similarly, cascaded-fidelity frameworks like Cafe-Mpc balance computational speed and accuracy by progressively relaxing constraints and reducing model fidelity along prediction horizons, enabling dynamic maneuverability without sacrificing precision [6].

Another critical aspect of these frameworks is task prioritization, as humanoid robots often operate under multi-objective scenarios that require balancing locomotion, manipulation, balance maintenance, and interaction with the environment. Prioritized control strategies, such as hierarchical inverse-dynamics methods, assign relative weights to conflicting task objectives, ensuring that essential functions like balance or obstacle avoidance are preserved during high-demand conditions [50]. Such methods commonly employ quadratic programming (QP) formulations to resolve priority conflicts, effectively managing scenarios like dynamic walking paired with manipulation tasks [52]. These formulations also integrate contact-specific constraints, enabling adaptive performance during environmental disturbances or platform instability.

Recent trends in hierarchical MPC involve dynamically adjusting task parameters to accommodate a wide range of environmental conditions and behaviors. For instance, variable horizon control adjusts footstep timing and placement to support tasks involving uneven terrains or high-speed locomotion, improving stability and responsiveness [5]. Moreover, hierarchical designs have been extended to support the seamless transition between locomotion and manipulation tasks in multi-contact scenarios by integrating online optimization of timing and task-switching parameters [74]. Machine learning-driven strategies further enhance runtime adaptability by calibrating task-specific priorities dynamically, aligning high-level planning with real-time state adjustments.

Task-aware control representations have also advanced hierarchical MPC designs, leveraging dynamic control-affine formulations that prioritize sub-task decoupling for resource-efficient execution. These methods introduce robust null-space resolution techniques, enabling humanoid robots to spatially separate redundant degrees of freedom (DoFs) from primary tasks, streamlining complex optimization problems [70]. By adopting such structured approaches, hierarchical MPC frameworks efficiently mediate between computational intensity and real-time adaptability, ensuring robust performance under diverse conditions.

Integrating task-specific objective functions defined through domain expertise or learned via data-driven methods has further bolstered hierarchical MPC. For example, pre-trained policies can augment low-level controllers by refining their objectives during runtime, reducing sub-optimality and increasing robustness in whole-body controllers [30]. The alignment of hierarchical control frameworks with learned models ensures that adaptive, scalable systems remain responsive to changing environments while benefiting from both structured task prioritization and learned dynamics.

Nonetheless, significant challenges persist. Balancing model fidelity with real-time computational viability remains a key concern, especially in high-degree-of-freedom robots. Addressing multi-contact scenarios with non-smooth dynamics introduces further computational complexity, necessitating the development of more efficient solvers capable of resolving hybrid interaction constraints [59]. Meanwhile, incorporating adaptive mechanisms into hierarchical schemes—such as integrated task planning, variable parameter tuning, and task-specific learning—will be essential for sustaining performance in increasingly diverse and challenging applications.

In conclusion, hierarchical and task-specific MPC frameworks provide a powerful means of achieving expressive humanoid motions, combining computational efficiency with robust adaptability through decomposed and prioritized control structures. By integrating learning-based policy refinement and task-awareness, these approaches hold considerable promise for enabling scalable, flexible control systems in complex humanoid robotics. Future research should focus on advancing solver acceleration, optimizing task parameters, and seamlessly incorporating multi-contact dynamics, paving the way for versatile and efficient implementations capable of meeting the demands of real-world environments.

4.5 Advances in Safety, Viability, and Planning Horizons

Ensuring safety, feasibility, and long-term planning is a critical focus in Model Predictive Control (MPC) for expressive humanoid robots, as such systems must operate effectively in dynamic environments without compromising stability or violating constraints. Recent advancements in MPC, particularly developments in safe optimization methods, adaptive horizon adjustment, and planning under physical and computational constraints, have significantly enhanced the safety and viability of humanoid robot control.

To ensure safety during motion planning, many MPC frameworks have incorporated strict safety constraints, often formulated through control barrier functions (CBFs) or viability kernels that explicitly encode safe operating boundaries. For example, the integration of CBFs with MPC guarantees that trajectories respect collision avoidance constraints and self-collision bounds, even in highly dynamic motions [80]. These constraints ensure not only kinematic feasibility but also the robustness of derived plans under disturbances and uncertainties. Similarly, viability kernels have been employed to guarantee forward invariance, ensuring that the robot remains in a recoverable state throughout execution [26]. Such approaches systematically enhance the robot's ability to maintain stability and task continuity, even under unexpected disturbances.

Predictive horizon optimization has emerged as a key capability in improving both safety and computational feasibility. Traditional MPC applications often assume fixed planning horizons, which can limit adaptability to task-specific needs or environmental changes. Novel methods such as variable horizon MPC dynamically adjust the control horizon based on environmental feedback and system dynamics [5]. This flexibility balances computational efficiency with motion predictability, allowing the robot to account for complex scenarios, such as negotiating uneven terrain or sudden changes in task objectives. Additionally, recent research has explored tailored trade-offs between short and long horizons by using hierarchical formulations that combine coarse long-term planning with fine-grained, high-frequency corrections during execution [24].

Feasibility guarantees during motion execution remain one of the most challenging aspects of humanoid robot control. The inclusion of physically-based terminal constraints, such as terminal cost functions that encode desired stability conditions or reachable sets, ensures that trajectory endpoints support sustained dynamic balance [16], [18]. Moreover, data-driven approaches leveraging reinforcement learning (RL) to refine viability constraints provide another avenue for augmenting safety. For instance, RL-trained policies have been utilized to enhance the accuracy of robustness metrics, demonstrating promise in enabling MPC to handle highly nonlinear model inaccuracies more effectively [29].

To further address the compromises between safety and computational efficiency, advanced optimization-based solvers have been introduced for MPC frameworks. These solvers, such as ADMM-based methods or sparse problem formulations, optimize multi-step planning while adhering to the real-time constraints required by high-degree-of-freedom humanoid systems [7], [45]. By leveraging sparsity in the problem structure, these methods significantly reduce computational overhead, enabling safety-compliant MPC frameworks to operate at frequencies exceeding 100 Hz, even during multi-contact and hybrid dynamics scenarios.

Despite these achievements, challenges persist in ensuring safety during transitions between dynamic tasks. Hybrid control methodologies, such as those combining trajectory viability with probabilistic guarantees, are increasingly being explored to address uncertainties in contact forces and environmental interactions [19]. Emerging directions also include the use of multimodal sensory feedback, integrating vision with proprioception and force sensing, to refine motion planning and adaptability in unstructured or unknown environments [81].

In summary, advances in MPC for safety, feasibility, and planning horizons underscore a shift toward more adaptive, data-driven, and computationally robust frameworks. As these approaches evolve, the integration of learning-based methods with traditional optimization will likely play a key role in further enhancing the safety and scalability of humanoid robots. These efforts not only promise improved performance in cluttered or unpredictable environments but also lay the groundwork for bridging the gap between simulated policies and real-world deployments. Continued study into hybrid learning-control frameworks and methods for uncertainty quantification, particularly in long-horizon settings, remains an exciting frontier for further exploration.

5 Advances in Reinforcement Learning for Human-Centric Humanoid Control

5.1 Reinforcement Learning for Expressive and Dynamic Humanoid Behavior Adaptation

In recent years, reinforcement learning (RL) has emerged as a transformative approach for enabling expressive and adaptive behaviors in humanoid robots. By leveraging autonomous exploration and reward-driven policy optimization, RL facilitates the synthesis of fluid, human-like whole-body motions under complex, real-world constraints. This subsection explores key methodologies for designing task-specific reward frameworks, fostering behavioral realism, and promoting motion diversity within humanoid robotics, with a focus on their technical underpinnings, challenges, and recent advancements.

Central to RL-driven expressive humanoid behavior is the design of reward functions tailored to encourage both task success and biomechanical plausibility. For instance, task-specific rewards can encode priorities such as achieving energy-efficient walking, maintaining balance, or synthesizing human-like gestures, while ensuring natural motion trajectories. Prominent frameworks incorporate potential-based reward shaping (PBRS) to guide learning, as it preserves policy optimality while accelerating convergence [82]. However, the complexity of humanoid motion requires balancing competing objectives, such as task accuracy and physical realism, necessitating multi-objective reward formulations. Studies utilizing adversarial imitation reward functions have demonstrated success in learning policies that mimic human motion patterns for locomotion or manipulation tasks, while mitigating the discrepancy between simulated and real-world dynamics [55], [56].

A critical challenge in learning expressive motions lies in balancing goal-driven optimization and kinematic naturalness. One innovative approach is residual force control (RFC), where supplemental corrective forces are integrated into the humanoid's policy space to address dynamics mismatches between human motions and robotic constraints. This allows humanoids to synthesize agile and complex expressive motions, such as ballet spins or dramatic gesture transitions, with unprecedented fidelity [8]. Similarly, frameworks such as hierarchical reinforcement learning (HRL) introduce structured policies that decompose high-level objectives, such as dynamic movement sequencing, from low-level joint control [56].

Another essential area of focus in RL-based adaptation for expressive behaviors is achieving motion diversity and robustness while navigating high-dimensional action spaces. Techniques like motor primitive learning compress hundreds or thousands of expert behaviors into compact latent representations, which can further serve as reusable building blocks for generating diverse humanoid motions. Such probabilistic embedding spaces have successfully enabled humanoids to perform unseen tasks, such as blending locomotion with manipulation or switching between physically dissimilar motions [54]. Curriculum learning techniques complement these pipelines by progressively introducing increasingly complex motions as robots master foundational skills, avoiding local optima in training [10].

The inherent gap between simulated environments and real-world conditions—a pervasive issue in RL workflows—poses additional challenges. Advanced sim-to-real transfer methods, including domain randomization and fine-tuned adaptation, minimize discrepancies in factors such as contact friction, actuator delays, and kinematic imprecision [11]. Furthermore, hybrid learning frameworks augment RL policies with model-based control paradigms, enforcing real-time physical constraints during training to ensure that simulated behaviors remain viable on physical humanoids [13].

Emerging trends in RL research further emphasize multimodal frameworks for human-like expressivity and interaction. Multi-channel perceptions—encompassing proprioception, tactile feedback, and dynamic environment cues—are integrated into policy training to improve task adaptability and social responsiveness. For instance, integrating emotional state priors or social affordances into reward functions has enabled humanoids to exhibit expressive cues aligned with human social behaviors [3], [83]. Such techniques hold promise for advancing humanoid applications in interactive domains like entertainment, healthcare, and collaborative workspaces.

In conclusion, reinforcement learning provides an adaptive and scalable foundation for achieving dynamic, expressive humanoid behaviors. It strengthens the bridge between task-centric optimization and naturalistic motion synthesis, albeit at the cost of increased computational complexity and sensitivity to reward design. Future directions include the use of unified reward architectures, biomechanically accurate simulation-to-reality pipelines, and hybrid frameworks incorporating predictive task modeling. These avenues promise not only to narrow the sim-to-real gap but to unlock new layers of expressivity and versatility in humanoid robotics.

5.2 Advanced Techniques for High-Dimensional Action Spaces in Humanoid Robots

The high-dimensional action spaces characteristic of humanoid robots pose unique challenges in control and reinforcement learning (RL). With upwards of 30 degrees of freedom (DoFs), these robots exhibit complex interdependencies and redundancies across joints and limbs, complicating motion synthesis and policy optimization. Addressing these challenges demands advanced learning algorithms tailored for dimensionality reduction, hierarchical structuring, and computational scalability. This subsection explores cutting-edge approaches to managing these complexities, emphasizing hierarchical policy design, motor primitive utilization, and optimization-driven RL methods.

Hierarchical reinforcement learning (HRL) has emerged as a powerful strategy for managing the intricacies of high-dimensional action spaces by decomposing control into structured subproblems. HRL employs multi-layered policies in which higher-level controllers define abstract objectives, such as trajectories or goals, while lower-level controllers translate them into precise motor commands. This structure facilitates efficient coordination across limbs while localizing computational demands to each policy level. For instance, HRL paradigms have demonstrated effectiveness in tasks like locomotion, where high-level policies dictate stepping direction and low-level policies execute limb synchronization to maintain balance and prevent collisions [50]. However, designing interface mechanisms to seamlessly transfer information between policy levels remains a challenge. Recent advancements include adaptive hierarchical partitions that dynamically adjust policy roles in response to task states, thereby enhancing flexibility and minimizing information bottlenecks [29].

Complementing HRL, motor primitives offer a structured approach to represent high-dimensional action spaces. These primitives encode reusable motion templates, such as stepping or arm-swinging patterns, which compress redundant actions into more tractable abstractions. By operating within this latent action space, RL agents can bypass exhaustive exploration of full DoFs while maintaining biomechanical veracity. Techniques like Deep Dynamical Movement Primitives (DMPs) and learned latent-space embeddings have demonstrated significant acceleration in policy convergence while supporting expressive motion generation [5], [21]. For example, motor primitives enable humanoids to seamlessly integrate multi-contact locomotion tasks by constraining exploration to feasible regions of the action space [22]. However, further research is necessary to develop motion primitives that are context-aware and generalizable across diverse tasks and operating conditions, particularly in unstructured environments.

Curriculum learning represents another critical avenue for tackling the challenges posed by high-dimensional control. By structuring training to start with simple tasks and gradually increase complexity, curriculum learning enables RL agents to build foundational skills that pave the way for mastering advanced behaviors. For humanoids, this often involves learning static balancing before progressing to dynamic locomotion and manipulation [27]. Additionally, curriculum learning can synergize with domain adaptation techniques to extend simulated policy training to real-world conditions, bridging the simulation-to-reality gap [43]. This combined approach holds promise for refining RL agents' robustness in dynamic and unstructured scenarios.

Optimization-based RL offers another complementary pathway for improving the efficiency and robustness of high-dimensional humanoid control. These approaches integrate physical constraints, such as self-collision boundaries and torque limits, directly into the learning process, ensuring biomechanically viable motion synthesis at every policy iteration. Techniques such as inverse dynamics trajectory optimization and constrained optimization solvers have shown strong potential for real-time application while maintaining high fidelity [58]. Furthermore, innovations like warm-start initialization and parallelized solver architectures enhance scalability, enabling humanoid robots to generate responsive whole-body motions in temporally constrained scenarios [32].

In summary, approaches for managing high-dimensional action spaces in humanoid robots increasingly emphasize hierarchical policy architectures, motion primitives, curriculum-driven learning, and optimization-based RL frameworks. Together, these methods provide the computational efficiency and representational clarity necessary to handle complex motion redundancies while preserving the adaptability required for generalizing across diverse tasks. Nonetheless, critical gaps remain, such as improving the integration of model-based and model-free techniques, refining task transfer mechanisms, and developing unified frameworks that balance precision with high-level adaptability. Future research will benefit from exploring hybrid RL architectures enriched by biomechanical insights and automated structure selection, advancing humanoid robots toward more expressive and functional real-world deployment.

5.3 Simulation-to-Reality Transfer for Humanoid Control

The simulation-to-reality (sim-to-real) gap represents one of the most enduring challenges in deploying reinforcement learning (RL) policies for humanoid robot control. While simulation environments provide efficient, scalable platforms for training policies under controlled conditions, transitioning these learned behaviors to real-world humanoid robots poses significant obstacles. These challenges arise primarily from discrepancies between simulated and real-world dynamics, environmental uncertainties, and sensor noise. This subsection delves into the strategies employed to bridge the sim-to-real gap, analyzing techniques such as domain randomization, domain adaptation, and real-time fine-tuning, and highlighting their advancements, limitations, and future directions.

Domain randomization, a cornerstone methodology for addressing the sim-to-real gap, prepares policies to generalize across variations in real-world conditions by introducing stochastic perturbations to simulation parameters during training. Parameters such as mass, friction, joint stiffness, and external disturbances are randomized to ensure that the policy does not overfit to a specific simulated model, thereby improving robustness during deployment. For example, reinforcement learning frameworks addressing bipedal locomotion [84], [85] demonstrated that domain randomization enables robust walking and running on real robots, even in the presence of environmental inconsistencies like rough terrain or hardware variations. A significant strength of this approach lies in its scalability and low computational overhead during simulation. However, it often requires careful tuning of randomization ranges; excessively wide ranges may confuse the policy with unrealistic scenarios, while overly narrow ranges may fail to account for real-world variability.

Complementing domain randomization, domain adaptation techniques focus on minimizing discrepancies between simulated and real environments by refining learned policies after deployment. Adaptive fine-tuning methods leverage real-time sensory feedback to optimize RL policies during execution in a physical environment. For instance, simulation-to-reality learning paradigms employing current feedback and targeted dynamics randomization [86] allowed robust deployment of RL controllers on a humanoid robot with life-sized actuators by addressing inaccuracies in torque tracking. Similarly, approaches such as residual reinforcement learning combine an initial policy trained in simulation with an adaptive correction term during real-world deployment to rectify deviations stemming from unmodeled dynamics [8]. These methods demonstrate the potential of incremental adjustments to enhance policy robustness without extensive retraining.

Physics engine fidelity plays an instrumental role in reducing the initial gap between simulation and deployment. High-accuracy physics simulators have been particularly successful in humanoid control, as seen in frameworks such as Humanoid-Gym [39]. By integrating soft constraints into simulated interactions and validating RL policies using alternative simulators, these systems guarantee a higher likelihood of real-world transfer without additional fine-tuning. Nevertheless, reliance on high-fidelity models significantly increases computational costs and limits scalability to complex, high-dimensional scenarios.

Innovative hybrid methods combining RL with model-based control also show potential for sim-to-real transfer. In RL + Model Predictive Control architectures [78], on-demand trajectory optimization minimizes errors associated with model simplifications while enabling reliable policy generalization. Similarly, approaches coupling reduced-order models with RL enhance transferability by encoding essential dynamical constraints within the control framework [87]. These hybrid approaches effectively combine physics-based accuracy with the adaptability of data-driven learning, overcoming the limitations of standalone RL.

Recent advancements also explore leveraging large-scale simulation datasets for building versatile policies. For instance, universal motion representations conditioned on proprioceptive data [55] were found to enable long-term task generalization without retraining, while hierarchical policy architectures enhanced learning efficiency by decoupling exploratory behaviors across high- and low-level controllers [79]. These frameworks emphasize the importance of multi-stage policy training for sim-to-real success.

Despite these advancements, achieving robust sim-to-real transfer across diverse humanoid platforms remains an open problem. One key area is the development of improved simulation parameter estimators to match real-world conditions under highly dynamic and stochastic scenarios, such as push recoveries or agile transitions [88]. Another challenge lies in refining data-driven adaptations that are context-aware, where policies can utilize real-time variability in environmental feedback to modify their behaviors dynamically [89].

In conclusion, strategies for bridging the sim-to-real gap in humanoid robotics are rapidly evolving and transforming RL into a viable solution across physical platforms. However, critical challenges such as scalability, efficient feedback-driven refinement, and uncertainty quantification persist. Future research must focus on advancing hybrid control frameworks and leveraging structured hierarchies for efficient and robust policy deployment, thereby unlocking the full potential of learning-based humanoid control in real-world environments.

5.4 Multi-Modal Learning for Human-Robot Interaction

The integration of multi-modal inputs represents a crucial frontier for advancing expressive, interactive, and context-aware capabilities in humanoid robotics. Multi-modal learning, driven by Reinforcement Learning (RL), provides the computational foundation for synthesizing diverse sensory feedback—such as visual, tactile, proprioceptive, and auditory signals—into coherent and adaptive control policies. By leveraging these diverse inputs, humanoid robots can dynamically adjust their motions, gestures, and reactions in ways that enhance interaction quality, task performance, and operational safety, particularly in collaborative human-robot interaction (HRI) scenarios.

At the core of multi-modal RL frameworks is the ability to process and fuse heterogeneous sensory modalities into unified state representations that inform policy learning. For instance, visual inputs enable robots to perceive and interpret human activities, while tactile and proprioceptive signals offer feedback on contact forces and joint configurations to ensure physical compliance [46]. The integration of vision with proprioceptive feedback enhances control precision in tasks requiring environment interaction, such as door opening or object manipulation [47]. Similarly, the fusion of tactile and proprioceptive information enables robots to respond effectively to external disturbances, ensuring stable and adaptive behaviors during close-proximity interactions [71].

Hierarchical architectures have become a prominent approach in multi-modal RL systems, effectively decomposing complex control challenges into high- and low-level policies. High-level modules extract global features from visual or auditory inputs, while low-level controllers manage fine-grained proprioceptive and tactile feedback to command precise motor actions. For example, hierarchical world models align visual goals with proprioceptive state dynamics, enabling humanoids to synthesize coordinated behaviors while accounting for real-world uncertainties in interactive environments [28]. These hierarchical structures have demonstrated significant success in simultaneously enhancing task generalization and control robustness across diverse interaction contexts.

Nevertheless, multi-modal RL frameworks encounter challenges related to the asynchronous nature of sensory modalities and partial observability in dynamic scenarios. Delays in visual processing, for instance, may disrupt the coordination between visual and proprioceptive feedback, leading to suboptimal task execution or stability issues. To overcome such challenges, belief-based filtering mechanisms and recurrent structures, such as Long Short-Term Memory (LSTM) networks, have been deployed to reconstruct missing or delayed sensory signals, ensuring temporally consistent policy predictions [90]. Uncertainty-aware models, such as Gaussian processes, further address variability and noise within sensory streams, improving overall system resilience [71].

On the expressive front, multi-modal learning enables humanoid robots to achieve social and emotional expressivity critical for engaging in meaningful HRI. Reinforcement-guided policies trained on multi-sensory observations can adapt non-verbal cues, such as gestures and postures, to align with human social dynamics [10]. Furthermore, the integration of auditory feedback enriches these interactions, facilitating contextual verbal and non-verbal exchanges. For instance, behavioral imitation using visual motion capture data, combined with policy optimization, has allowed robots to replicate human-like movements with naturalistic accuracy [42].

Despite these advances, scalability and computational complexity remain significant barriers. Multi-modal RL often demands extensive training data and computational resources, especially given the high-dimensional state and action spaces characteristic of humanoid robots. Approaches such as curriculum-based training and domain adaptation have shown promise in addressing these challenges. By incrementally introducing complex tasks during training, curriculum learning ensures the modular integration of sensory modalities while preserving convergence efficiency [46]. Domain adaptation further aids in bridging simulation-to-reality gaps, enabling effective transfer of trained policies to real-world settings.

Future directions in this domain must prioritize tighter integration and coordination across sensory modalities to enhance adaptability and real-time performance. Emerging techniques, such as generative models using Variational Inference MPC [44] and transformer-based architectures, offer promising avenues for improving multi-modal representations and computational efficiency. Additionally, deploying multi-modal humanoids in sensitive applications, including caregiving and therapy, requires ethical frameworks that emphasize transparency, reliability, and user comfort [46].

In summary, multi-modal RL stands as a pivotal enabler for the next generation of interactive and context-aware humanoid robots. By unifying diverse sensory inputs within adaptive reinforcement frameworks, these systems hold transformative potential for creating socially expressive, safe, and versatile humanoid platforms capable of thriving in real-world environments.

5.5 Task Performance, Generalization, and Robustness

Reinforcement Learning (RL) has emerged as an essential paradigm for achieving versatility and adaptability in humanoid robots. One critical research challenge within RL-based control frameworks is ensuring that policies generalize effectively across diverse tasks, environments, and robots, while remaining robust under uncertainty and dynamic conditions. Generalization and robustness are fundamental to deploying RL strategies in complex, unstructured human-centric scenarios, as they address both the diversity of required applications and the harsh realities of unexpected disturbances.

Task generalization in humanoid robotics often necessitates a careful balance between the specificity of task rewards and the underlying policy structure. Approaches such as curriculum learning and hierarchical policies have been particularly effective in leveraging task similarities while accommodating task-specific dynamics. Staged training procedures, for example, incrementally expose policies to progressively complex tasks, thereby mitigating catastrophic forgetting and enhancing transferability across task domains [10], [46]. In addition, the use of hierarchical RL frameworks allows decoupling of high-level strategies (e.g., goal selection) from low-level controllers (e.g., joint space coordination), which has proven instrumental in managing high-dimensional humanoid action spaces [91]. However, a trade-off exists between the computational overhead of training hierarchical structures and the robustness gains they deliver, which remains an open question in resource-constrained scenarios.

One of the major hurdles in robust RL for humanoid robots lies in optimizing reward structures to ensure both task efficiency and behavioral stability. Sparse or deceptive rewards, common in real-world tasks, often impair learning convergence and robustness. Potential-based reward shaping techniques have been employed to address sparse rewards, ensuring that exploration remains focused on meaningful regions of the action space [12]. For instance, integrating viability metrics into RL reward functions encourages safe and stable exploration, particularly during dynamic humanoid locomotion [26]. Yet, overly biased reward shaping may limit the diversity of learned behaviors, underscoring the need for composable and explainable reward architectures [10].

Expanding on robustness, robust policy designs address various types of disturbances, such as external forces or sensor degradation, which humanoid robots are prone to encounter. Techniques such as adversarial training—where policies are exposed to simulated perturbations—improve the system's ability to recover and maintain stable operation [26], [92]. Notably, Bayesian optimization has also been used to systematically tune reward weights and cost functions post hoc in order to bolster stability and robustness in uncertain environments [34]. This iterative tuning overcomes discrepancies between training environments (often simulated) and real-world deployment, a critical concern for RL-driven designs.

Generalization and robustness further benefit from leveraging diverse sensory modalities, as demonstrated in multi-modal RL frameworks. Combining visual, tactile, and proprioceptive feedback enables humanoid robots to adapt to a wider range of environmental conditions by enriching their state representations [81]. Multi-modal learning has enhanced social and task-oriented behaviors alike by integrating these streams of feedback into coherent policy designs, albeit at the cost of additional computational complexity [46].

Simulation-to-reality (sim-to-real) transfer constitutes another pivotal aspect of generalization, as strategies learned in simulation must perform reliably under real-world dynamics. Domain randomization, for instance, introduces variability into simulation parameters to improve policy robustness against discrepancies between virtual and physical environments [69]. Real-time fine-tuning using adaptive RL during deployment further ensures robust sim-to-real performance, bridging gaps that even state-of-the-art simulation methods might leave unaddressed [93].

Despite these advancements, challenges remain. Current approaches often struggle with scaling to humanoids with higher degrees of freedom (DoFs) or transferring models across different humanoid platforms. The scalability issue is tied to the computational intensity of RL methods, while cross-platform transferability is hindered by variations in hardware design and dynamics. Emerging work in motor primitive-based action space reduction has shown promise for tackling these issues by encapsulating task-relevant behaviors into compact representations that simplify policy adaptation [46].

Future directions call for integrating probabilistic methods, such as Bayesian RL, to model and adaptively manage uncertainties during task execution. Combining RL with Model Predictive Control (MPC) frameworks also offers considerable potential to enhance robustness, as evidenced by RL-augmented MPC designs that refine foot placement and whole-body dynamic behaviors [29]. Moreover, exploring meta-learning paradigms could enable policies to rapidly adapt to new tasks with minimal retraining, a crucial ability for practical humanoid deployment.

In summary, advancements in RL for task generalization and robustness in humanoid robotics have laid a strong foundation for achieving versatile and reliable human-centric control. However, addressing scalability, computational efficiency, and cross-platform adaptability will require innovative designs that synergize RL with complementary methodologies. Such interdisciplinary approaches could unlock the next stage of humanoid robotic autonomy and expressiveness in dynamically complex environments.

5.6 Ethical, Social, and Computational Considerations

The deployment of reinforcement learning (RL) in humanoid robots for expressive and human-centric behavior bridges advanced computational methodologies and critical societal considerations. This subsection examines these dimensions, focusing on the trade-offs inherent in designing RL-powered systems for real-time adaptability while addressing the broader impact of humanoid robots exhibiting human-like behaviors.

The computational challenges of RL-based controllers for humanoid robots are rooted in the high-dimensional action spaces and the simultaneous requirements for adaptability and stability. Training such policies to handle multi-task objectives within real-world time constraints remains a significant hurdle. Hierarchical reinforcement learning has proven effective, decomposing tasks into high-level objectives and low-level motor primitives to mitigate computational overhead while preserving an expressive range [46], [55]. Similarly, dimensionality reduction techniques, such as structured latent spaces and motor primitive embeddings, enable efficient policy representation while maintaining motion fidelity [46], [55]. However, achieving real-time control demands high-performance computational hardware, such as GPUs, and highlights the trade-offs between task generalization across diverse scenarios and the complexity of scaling expressive control to real-world deployments [7], [94].

Ethical considerations arise significantly when designing humanoid robots capable of expressive, human-like behavior. Anthropomorphism can improve trust and engagement in applications such as caregiving, education, and social robotics. However, it also carries risks of user manipulation, emotional dependency, and the formation of unrealistic expectations. Design transparency is pivotal in mitigating these risks. Incorporating explainable decision-making mechanisms into RL-driven control frameworks can reduce ambiguity in robot actions and foster trustworthiness [28], [90]. Additionally, interpretable feedback loops and reward structures that align with pre-defined ethical constraints ensure that robots adhere to safe and socially acceptable behavior patterns during real-world deployments [28].

A particularly pressing concern is the ethical boundary for emotional expression and the extent to which humanoid behaviors should mirror human social cues. The mimicry of human-like gestures and interactions can unintentionally exploit user vulnerabilities, creating "deceptive" perceptions about autonomy and empathy. These challenges necessitate integrating ethical considerations directly into RL training and reward design processes. Reward mechanisms should prioritize functional expressiveness (e.g., improving task clarity) over emulating emotional engagement, balancing usability with responsible design principles [55].

Scalability across different humanoid designs adds further complexity, both technically and ethically. RL systems must adapt expressive policies to accommodate diverse physical morphologies, which may involve varying energy footprints and structural constraints. Scalable frameworks, such as those leveraging lightweight architectures for skill retargeting [95], highlight the trade-offs between enabling expressive behavior and minimizing computational and environmental costs. Incorporating energy-efficient reward paradigms that balance expressiveness with energy conservation, stability, and other task demands offers a pathway to sustainable RL-driven humanoid systems [96], [97].

Simulation-to-reality transfer compounds these challenges, as real-world deployments often introduce discrepancies due to incomplete modeling of physical dynamics and human-robot interactions. Techniques such as domain randomization, adaptive real-time policy adjustments, and fine-tuned physics models address these discrepancies and enhance robustness during deployment [7], [98]. However, iterative fine-tuning also raises ethical concerns, particularly regarding the autonomy of adaptation, underscoring the importance of maintaining human oversight to ensure compliance with behavioral norms.

In conclusion, RL-powered humanoid robots hold transformative potential for advancing human-centric tasks, such as caregiving, creative collaboration, and precision automation. However, achieving this vision demands continuous innovation in policy efficiency, scalability, and ethical governance. Addressing the computational demands requires hybrid approaches that combine optimization techniques with learning-based strategies, while societal challenges necessitate embedding ethical constraints into training objectives. Balancing technological advancements with societal responsibility is imperative to the sustainable evolution of expressive and human-centric humanoid robotics [52], [99].

6 Hybridization of Model Predictive Control and Reinforcement Learning

6.1 Foundations of the Hybrid Framework

The hybridization of Model Predictive Control (MPC) and Reinforcement Learning (RL) represents a pivotal innovation in the realm of humanoid robotics, particularly in achieving expressive whole-body control. By unifying MPC's strength in handling multi-variable constrained optimization with RL's capacity for adaptive and exploratory policy learning, hybrid frameworks are uniquely positioned to address both the predictability required for physical system constraints and the flexibility needed to operate in dynamic, unstructured environments. This subsection establishes the theoretical and conceptual foundation of hybrid MPC-RL systems, evaluates their complementary properties, and emphasizes their transformative potential in humanoid robot control.

Fundamentally, MPC excels in planning optimal motion trajectories over finite horizons by solving constrained optimization problems. It provides actionable solutions that adhere to physical feasibility and safety constraints, such as balance maintenance, joint torque limits, and inter-limb coordination [4], [51]. MPC's capacity to enforce spatial and temporal constraints in real time makes it a reliable mechanism for ensuring the stability and robustness critical to humanoid motion tasks. However, despite these strengths, standalone MPC faces significant limitations in high-dimensional systems. Its reliance on accurate dynamic models and the computational complexity of solving constrained optimization at high frequencies restrict its applicability to environments with large uncertainties or substantial real-time variability [4], [7].

Conversely, RL offers a complementary paradigm that leverages experience-driven learning to discover policies that can generalize across task variations. RL-based methods, powered by trial-and-error interactions, excel in tasks requiring adaptability, such as navigating unstructured terrains or executing dynamic transitions between behaviors [56], [100]. However, a major limitation of RL is the absence of guarantee over safety or physical feasibility during exploration. This is particularly challenging in the context of humanoid robots, where failures during learning can lead to irreparable damage to hardware or dangerous states in real-world interactions [8].

The hybridization of MPC and RL leverages their strengths to overcome these limitations in tandem. Perhaps the most immediate advantage of hybrid systems lies in MPC's ability to constrain RL's exploration space, introducing safety envelopes that prevent the system from entering unsafe states during training or execution [13], [80]. For instance, MPC can enforce constraints such as capture point limits for balance recovery [18] or friction cone constraints in multi-contact scenarios [19]. This synergy further enhances sample efficiency, a critical bottleneck in RL, by guiding exploration to areas of the state space where useful, feasible policies are likely to emerge.

Another promising variant of hybridization involves RL-augmented MPC schemes, where RL-trained policies complement MPC by providing high-level decisions or adaptive modules for real-time refinement [8]. For example, in locomotion tasks, RL can be employed to predict optimal stepping strategies or dynamic foot placement, while the finer trajectory optimization is delegated to MPC, which ensures real-time convergence [35]. This paradigm enables humanoids to tackle tasks that require rapid decision-making and adaptability, such as traversing uneven terrains under external disturbances [26].

Hybrid frameworks can also be classified based on their architectural integration. In model-based RL approaches, MPC-generated trajectories serve as structured priors in RL training, thereby reducing policy divergence caused by inaccuracies in simulated models [72]. An alternative approach, MPC-enhanced RL, embeds an MPC module within an RL feedback loop, dynamically updating intermediate reference states to ensure feasibility at each time step [17]. Parallel hybridization, meanwhile, allows modular systems where MPC ensures low-level safety constraints while RL governs high-level tasks, including long-term scene planning and social interactions [15].

Despite these strengths, several challenges persist in hybrid MPC-RL systems. Achieving computational efficiency while preserving real-time constraints remains a significant hurdle, particularly in high-dimensional systems like humanoid robots. Advances in solver acceleration, including GPU-based computations and approximate solutions [6], [7], are pivotal for bridging this gap. Moreover, robust sim-to-real transfer mechanisms are essential to minimize discrepancies between training environments and physical robot dynamics [69].

In conclusion, the foundation of hybrid MPC-RL frameworks is built on the recognition of their complementary properties—MPC's constrained optimization and RL's policy generalization. This synthesis not only enables humanoid robots to exhibit robust, expressive, and human-like behaviors but also paves the path for significant future advancements. Collaborative research in efficient hybrid architecture design, dynamic safety constraint integration, and scalable policy learning will be instrumental in addressing remaining challenges and unlocking the full potential of hybrid systems.

6.2 Learning-Safe Strategies Enabled by MPC

The hybridization of Model Predictive Control (MPC) and Reinforcement Learning (RL) offers a robust pathway to enabling expressive and adaptable humanoid robot control by embedding learning-safe strategies. MPC, with its ability to define and impose formal safety constraints, ensures that RL agents operate within stable and feasible boundaries during both training and deployment. This subsection delves into methodologies that leverage MPC's constraint-handling capabilities to enforce critical safety guarantees, which, when combined with RL's adaptability, synthesize operational safety with efficient exploration and learning.

At its foundation, MPC provides a framework for enforcing constraints drawn from physical laws and environmental dynamics, addressing requirements such as balance stability, joint torque limits, and self-collision avoidance through optimal control problem-solving at each decision step. These constraints, indispensable for high-dimensional and dynamic humanoid systems, can be seamlessly integrated into RL frameworks via "safety envelopes" or viability kernels during training and policy execution. For instance, control barrier functions embedded within MPC, as demonstrated in [101], ensure collision-free and invariant trajectories while optimizing system performance. This safety-aware constraint formulation can also adopt probabilistic definitions, allowing the incorporation of model uncertainties to maintain adherence to pre-defined safety thresholds.

A key advantage of learning-safe strategies lies in safety-aware exploration. MPC restricts RL's exploration domain to operationally viable regions, mitigating risks of catastrophic failures—such as hardware damage or infeasible policy convergence—during training or execution. For example, predictive stability constraints showcased in [16] enable real-time locomotion tasks to preserve balance, even as RL policies probe novel movement patterns. Furthermore, robust dynamics modeling enhances this exploration by continuously updating control adjustments based on deviations between predicted and observed states. Probabilistic uncertainty frameworks, like the approach detailed in [43], quantify model inaccuracies to guide RL to explore novel strategies while maintaining task feasibility.

A significant challenge in integrating safety with learning emerges from the nonlinear and hybrid dynamics unique to humanoid robots. Contact-rich scenarios and highly dynamic maneuvers introduce discontinuities that conventional RL exploration struggles to reconcile. Here, MPC-augmented frameworks have demonstrated remarkable efficacy. For example, [77] extends MPC to reason about hybrid contact dynamics, enabling trajectory generation without relying on pre-specified contact sequences. By embedding contact constraints into the policy-learning process, this approach ensures RL outcomes are grounded in physically realizable trajectories. A similar synergy is observed in [59], which utilizes MPC to stabilize dynamic transitions under diverse contact mode variability, enabling RL to operate seamlessly across a broader range of scenarios.

However, the pursuit of learning-safe strategies involves trade-offs between computational overhead and exploratory conservatism. MPC's frequent resolution of constrained optimization problems can pose substantial runtime challenges, particularly when combined with RL's iterative updates. Advances in solver acceleration, such as GPU-based optimizations described in [102], address these bottlenecks by maintaining robust constraint enforcement at reduced computational cost. Conversely, overly conservative constraint imposition may cause RL policies to underexplore high-reward regions, hindering long-term adaptability. Dynamic relaxation of inactive constraints, as implemented in Transformer-enhanced MPC [32], provides computational efficiency while enabling more extensive RL exploration.

Future progress in the domain aims to further expand the expressivity and applicability of learning-safe MPC in humanoid robotics. Integrating learned models into MPC formulations, particularly through residual learning frameworks, offers a promising avenue for improved real-time deployment [30]. These approaches allow MPC to incorporate data-driven updates to predictive dynamics, enabling systems to adapt to unforeseen complexities in the environment while guiding RL policies toward safe and efficient solutions. Continued advancements in co-optimization of MPC and RL computation pipelines will be integral to resolving trade-offs in adaptability, safety, and runtime performance, unlocking scalable and generalizable strategies for next-generation humanoid robots.

6.3 Real-Time Control Using RL-Augmented MPC

The hybridization of reinforcement learning (RL) and model predictive control (MPC) presents a promising paradigm for the real-time control of expressive humanoid motions. This subsection delves into the integration of RL into MPC frameworks to address critical limitations of MPC such as computational overhead, adaptability in dynamic environments, and horizon-limited optimization. By leveraging RL's capacity to encode complex, high-dimensional dynamics and enable rapid adaptation, RL-augmented MPC systems demonstrate significant potential for enhancing the robustness and expressivity of humanoid control.

Traditional MPC implementations are constrained by their reliance on simplified system models and finite-horizon receding optimizations. While effective for ensuring constraint satisfaction and optimality under predictable conditions, they lack the capacity to generalize across rapidly changing environments or compensate for unmodeled dynamics. Reinforcement learning compensates for these shortcomings by training policies that leverage learned dynamics representations, often derived from high-dimensional data or prior trajectories [34], [103]. RL can also introduce feedback refinements, enabling adaptation to unforeseen perturbations in real time. For instance, pre-trained RL policies can augment the initialization of MPC optimization trajectories, reducing computational latency and addressing inadequacies in model fidelity, as illustrated in [13].

One of the primary approaches in RL-augmented MPC is the pre-computation of high-level strategic behaviors via RL, followed by the use of MPC to enforce low-level physical constraints and refine task execution. For example, using RL-trained policies that encode high-dimensional motion strategies, MPC can adjust for necessary constraints such as self-collision, balance, and force limits in real-time control loops. This synergy is particularly effective for humanoid robots performing dynamic tasks such as parkour [9] or multi-step walking on irregular terrains [19]. Here, RL provides a probabilistic model of task complexity, while MPC determines control outputs that satisfy the real-time feasibility requirements.

An important innovation within RL-augmented MPC frameworks is the use of learned dynamics approximations or residual policies. For instance, residual force policies trained via RL complement MPC to improve the system's ability to handle task transitions or achieve skill expressivity not captured by coarse-grained models [8]. Furthermore, uncertainty-aware RL integration with MPC allows systems to quantify and bound exploration risks during dynamic tasks, mitigating potential safety violations [33]. Probabilistic MPC approaches, enhanced with uncertainty estimations derived from Gaussian processes or neural network value approximations, facilitate safe exploration while retaining control robustness.

Despite these advances, challenges remain in balancing the computational loads imposed by MPC's optimization routines and the iterative evaluation demands of RL policies. Hierarchical control frameworks, where RL operates as a high-level task planner and MPC as a low-level trajectory stabilizer, are a pragmatic solution. Such hierarchies decompose the learning and control problem, enabling computational co-optimization. For example, task space RL frameworks have been shown to prioritize coarse locomotion objectives, allowing MPC to execute fine-grained motor control in humanoid systems [78], [104]. Additionally, GPU-based solvers and warm-start initialization techniques have emerged as practical solutions to bridge real-time implementation gaps, further enabling the feasibility of RL-augmented MPC frameworks [35].

The empirical success of RL-augmented MPC is reflected in experimental and simulated benchmarks. For instance, integrating RL-driven footstep adaptation into orientation-aware MPC has yielded significant advancements in robustness to external disturbances [35]. Similarly, combining reference trajectory generation with RL policies has accelerated sample efficiency for training multi-skill policies in complex locomotion tasks [87]. Notably, these systems also achieve zero-shot transfer to real-world environments, underscoring the practical applicability of hybrid frameworks [39].

Looking forward, ensuring seamless integration between learning-driven policies and optimal control mechanisms will require advances in both compute infrastructure and modeling paradigms. Future research may explore conditions under which RL-augmented MPC can autonomously infer task priorities or co-evolve policy gains alongside changing task objectives, thereby enabling long-term adaptability. Moreover, leveraging multi-agent interactions, as exemplified by frameworks like WoCoCo [79], could extend RL-augmented MPC to collaborative and interactive tasks, advancing the frontier of expressive humanoid control. This hybrid approach, while computationally demanding, offers an unmatched combination of adaptability, safety, and precision, providing a clear roadmap for the development of next-generation humanoid robotic systems.

6.4 Design and Implementation of Computational Co-Optimization

The co-optimization of Model Predictive Control (MPC) and Reinforcement Learning (RL) represents a pivotal step in harnessing the computational efficiency and functional synergy of these complementary paradigms for expressive humanoid control. Building upon the hybrid framework discussions, this subsection delves into methodologies and technical strategies that merge MPC's deterministic optimization with RL's adaptive learning, addressing challenges of computational scalability, model precision, and dynamic performance.

Hierarchical architectures are widely employed in co-optimization strategies, leveraging the strengths of MPC at the lower level for constraints and stability while assigning high-level decision-making and task adaptability to RL. A key challenge in this paradigm is achieving a balance between the computational demands of real-time MPC solving and the iterative application of RL policies. To overcome this, recent work [13], [48] has explored embedding RL-trained policies as initial guesses or guidance within the MPC optimization horizon. This integration reduces solver iterations and enhances responsiveness. Conversely, MPC can structure RL training by enforcing safety envelopes, guiding exploration away from unsafe or dynamically infeasible states, further strengthening the system's reliability.

Layered co-optimization structures are particularly prominent for managing high-dimensional, long-horizon control problems. By decomposing these challenges into sub-modules, task planning and constraint handling can be separately addressed to improve computational efficiency. For instance, hierarchical frameworks described in [30], [105] assign RL to handle nonconvex task-level objectives while reserving MPC to enforce stability and physical feasibility by simplifying subsets of constraints. This division reduces computational load per cycle and matches update frequencies to the distinct dynamics of each layer, enabling scalable real-time application.

Hardware acceleration techniques, such as GPU-based solvers and parallelization, have further improved the scalability of hybrid frameworks. Structured algorithms like the Alternating Direction Method of Multipliers (ADMM), as showcased in [106], enable rapid computations within the constraints of onboard controllers. When combined with RL's sparse policy gradients, these techniques create robust systems capable of balancing the fine-grained trajectory refinement of MPC with the adaptability of RL, even in high degrees of freedom (DoFs) humanoid platforms. This synergy ensures that time-critical tasks can be performed without compromising computational feasibility.

To maximize effectiveness, co-optimization frameworks must also address trade-offs between generalization and precision in both predictive modeling and policy design. While RL is exceptionally flexible, deployment outside its training distribution can lead to errors. Pairing RL's adaptability with predictive capabilities of MPC, as demonstrated in [13], [44], produces systems that maintain robustness even under uncertain conditions. For example, a Bayesian perspective in RL enables uncertainty quantification that MPC can exploit, adaptively adjusting constraints or planning horizons to maintain performance amidst unpredictable variables.

However, computational bottlenecks remain prominent, particularly for tasks requiring long-horizon predictions and frequent updates in highly dynamic environments. Cascaded fidelity frameworks, such as those proposed in [6], provide an efficient solution by employing low-fidelity models for distant planning horizons and reserving high-order models for immediate decision-making. This method reduces MPC's computational demands without degrading task performance. The inclusion of RL, which supplies probabilistic priors for less critical state-space regions, further enhances the system's computational efficiency and applicability across dynamic tasks.

Despite these advancements, significant challenges persist in ensuring the robustness of co-optimized policies under high-dimensional and dynamic environments laden with model inaccuracies and variability. Sim-to-real transfer methods [46], [71] and sampling-based optimization strategies, such as trajectory stitching or stochastic trajectory ensembles [107], are being actively developed to address these challenges. These approaches enhance robustness by promoting diverse training experiences or dynamically blending weighted sub-policies that adapt to varying scenarios, paving the way for improved reliability in practical settings.

Looking to the future, hybrid MPC-RL frameworks must increasingly incorporate uncertainty-aware algorithms and adaptive resolution techniques to enable dynamic resource efficiency. Developing modularized architectures, as discussed in [28], where specific components can be independently optimized or replaced, will also promote broader applicability across diverse humanoid platforms. As advances in computational hardware enable tighter integration of MPC's optimization-based planning and RL's stochastic adaptability, these frameworks will continue to evolve, offering novel pathways to enhance humanoid expressivity and autonomy in complex environments.

6.5 Case Studies and Benchmarks of Hybrid Approaches

Hybrid approaches combining Model Predictive Control (MPC) and Reinforcement Learning (RL) have demonstrated significant promise in tackling the complexities of humanoid robot control, particularly for achieving dynamic, expressive whole-body motions. This subsection delves into practical implementations and benchmarking efforts, addressing the interplay between MPC's optimization-based stability mechanisms and RL's adaptability in dynamic tasks. Through a comparative evaluation of case studies, the effectiveness, limitations, and potential applications of these integrated frameworks are critically analyzed.

The ability of hybrid MPC-RL approaches to handle tasks requiring both precision and adaptability has been vividly showcased in applications such as bipedal locomotion, balance recovery, and dexterous manipulation. For instance, hybrid frameworks have been employed to stabilize locomotion and navigate uneven terrains by complementing MPC's constrained optimality with RL's policy generalization for adaptive behaviors. In one notable example, RL-augmented MPC was leveraged for bipedal footstep locomotion, where RL improved sub-optimal step planning by compensating for full-body dynamics overlooked in traditional simplified MPC models [29]. This hybrid integration resulted in measurable improvements in tracking diverse walking speeds, negotiating complex terrains, and enhancing the overall robustness against disturbances. Similarly, in humanoid push recovery tasks, MPC was utilized to enforce stability through constrained optimization, while RL facilitated the adaptation to unexpected perturbations, such as external pushes or irregular ground surfaces [18].

A prominent benchmark feature of these studies is their exploration of trade-offs. Hybrid systems balance computational efficiency and real-time feasibility by distributing responsibility: MPC focuses on lower-level control enforcing physical constraints, while RL governs higher-level decision-making under uncertain or dynamic conditions [59]. For example, Bayesian approaches, such as Variational Inference MPC, have been employed to model uncertainties and optimize dynamic foot placements, demonstrating superior asymptotic performance compared to standalone

MPC [44]. However, this computational stratification introduces challenges in synchronizing the disparate time horizons of RL and MPC, particularly in tasks demanding rapid adaptation.

Applications in dexterous manipulation provide another rich benchmark for hybrid frameworks. Locomotion tasks interwoven with manipulation objectives, such as carrying payloads or pushing objects, require precise coordination of multi-contact dynamics—a task well-suited for hybrid MPC-RL systems. For instance, multi-contact MPC frameworks integrated with RL were demonstrated for humanoid loco-manipulation, enabling the seamless execution of tasks such as object transportation over uneven terrain while maintaining balance and stability [22]. In these instances, RL enhanced adaptability to unmodeled interaction forces, while MPC ensured force constraints and feasibility.

Comparative benchmarking emphasizes that hybrid frameworks often achieve better results than their standalone counterparts in terms of stability, robustness, and task adaptability. For instance, controllers employing RL-facilitated step planning consistently outperformed pure MPC methods in tasks requiring agile yet stable locomotion, especially under dynamic environmental constraints [26]. Similarly, hybrid methods incorporating RL for policy refinement showed increased robustness and energy efficiency in locomotion tasks compared to MPC-only strategies [27].

Despite these successes, several areas require further investigation. The computational overhead introduced by combining MPC's optimization loops with RL's iterative policy learning remains a critical bottleneck for real-time applications, particularly in high-degree-of-freedom humanoid systems. For instance, frameworks such as Adaptive Horizon MPC have demonstrated improved scalability, yet their integration with RL across dynamic and multi-contact scenarios still presents challenges in terms of real-time convergence rates and hardware constraints [5]. Additionally, the sim-to-real transfer problem persists as a fundamental hurdle, with hybrid systems often relying heavily on domain randomization to ensure policy robustness across simulated and real-world deployments [10].

Emerging trends suggest promising avenues for further exploration. Advances in hardware acceleration, including GPU-based parallelization and solver optimization, have begun to mitigate some of the computational burdens, facilitating deployment on physical robots [106]. Furthermore, probabilistic hybrid frameworks that incorporate uncertainty modeling in both the MPC planner and RL policies hold significant potential for improving robustness in unpredictable environments [44].

In conclusion, case studies and benchmarks underscore the remarkable potential of hybrid MPC-RL frameworks in humanoid robotics, achieving unprecedented adaptability and stability across dynamic and multi-dimensional tasks. However, future research must address scalability, real-time synchronization, and transfer learning to enable these systems to operate efficiently in complex, real-world environments. Integrating insights from biomechanics and exploring new hybrid paradigms, such as hierarchical RL-MPC architectures, will further refine the intersection of optimization and learning for expressive humanoid control.

6.6 Challenges and Design Guidelines for Hybrid Systems

The hybridization of Model Predictive Control (MPC) and Reinforcement Learning (RL) demonstrates substantial promise for expressive humanoid motion control by blending physics-based optimization with adaptive learning to deliver both real-time robustness and human-like expressivity. However, seamlessly integrating these paradigms while maintaining computational efficiency, stability, and adaptability under diverse operational constraints remains challenging. This subsection explores the key difficulties inherent to hybrid MPC-RL systems and proposes design principles to advance their development for humanoid robots.

One critical challenge in hybrid systems is addressing the simulation-to-reality (sim-to-real) transfer problem, particularly in adapting control policies and dynamics models. Both MPC and RL face inherent limitations when moving from simulation to real-world applications due to inaccuracies in robot dynamics modeling and uncertainties during environmental interactions. While MPC depends on high-fidelity models, constructing such models for high-degree-of-freedom (DoF) humanoids with contact-rich dynamics is often computationally prohibitive [50], [108]. Similarly, RL policies trained in simulation frequently falter in real-world conditions due to insufficient generalization and unmodeled disturbances. Hybrid approaches compound these challenges as MPC constraints and RL policies must co-adapt during sim-to-real transitions. Existing solutions like domain randomization and adaptive policy fine-tuning show promise but add significant computational overhead, limiting scalability [109]. The integration process thus requires co-designing simulation frameworks that strike a balance between approximate yet robust interaction modeling and scalable data collection to support RL learning.

Real-time computational demands present another formidable obstacle. MPC typically requires solving constrained optimization problems at high frequencies to maintain control stability, while RL-based components require computationally intensive forward propagation of high-dimensional policies. Embedding MPC within RL systems to dynamically manage physical constraints often introduces computational bottlenecks, especially in multi-contact scenarios requiring fine-grained contact force optimization [94], [108]. Conversely, incorporating RL into MPC frameworks risks undermining real-time feasibility. Advances in hardware accelerators, GPU-based solvers, and sparse problem representations have begun alleviating these bottlenecks, although further progress is needed to scale effectively for high-dimensional humanoid systems [7], [108].

Another significant issue is robustness to dynamic environmental disturbances and non-stationary conditions. While MPC excels in stabilizing predefined motion strategies via real-time trajectory corrections, it often falls short in highly stochastic scenarios where task adaptability is essential [18], [60]. RL, on the other hand, provides adaptability to variable conditions but lacks stability guarantees critical for safety-sensitive tasks. Hybrid frameworks must resolve this trade-off by using MPC's constraint enforcement capabilities to establish safe exploration boundaries for RL.

Within this bounded space, RL can safely adapt high-level task behaviors [18], [97].

Designing effective hybrid MPC-RL systems requires modularity, hierarchical architectures, and computational scalability. MPC should serve as a lower-level controller focused on fine-grained stabilization and constraint satisfaction, while RL should govern higher-level decision-making such as trajectory adaptation and approximate state representation. This hierarchical structure ensures that each paradigm exploits its strengths, with MPC acting as a safety framework for RL training [16], [25]. Additionally, probabilistic frameworks should be incorporated into hybrid systems to quantify uncertainties in real-time, enabling probabilistic MPC or RL to manage disturbances more effectively [80].

Hybrid systems must also integrate adaptive inertial modeling and energy efficiency to optimize performance in challenging dynamic tasks. For instance, incorporating adaptive centroidal dynamics modeling into hybrid systems enables real-time adjustments to variable inertial profiles, significantly improving performance in tasks such as jumping, complex manipulations, or highly uneven terrain locomotion [19], [97]. Moreover, prioritizing energy-aware optimization within hybrid MPC-RL frameworks enhances operational longevity and hardware reliability [110].

Despite these challenges, hybrid MPC-RL architectures represent a highly promising avenue for advancing bipedal stability, dynamic task adaptability, and expressive humanoid motion synthesis. Future research must focus on developing standardized task evaluation metrics and benchmarking tools tailored to hybrid systems, enabling consistent and fair performance comparisons across studies [111]. Addressing these challenges through thoughtful co-design and guided implementation could unlock the full potential of hybrid control systems, pushing the boundaries of humanoid robot expressivity and functionality across real-world applications.

7 Applications of Expressive Humanoid Robotics

7.1 Human-Robot Collaboration and Interaction

Expressive humanoid robotic motions bring unprecedented potential for enhancing human-robot collaboration and interaction, advancing functionality in assistive, service, and communicative roles. By replicating the nuanced dynamics of human expressions, these robots can transcend mere functionality, fostering more intuitive, responsive, and adaptive behaviors critical for tasks requiring close human-robot synergy. This subsection evaluates methodologies, applications, and challenges in developing expressive humanoid motions to improve collaboration and interaction.

In physical assistance scenarios, expressive humanoid motions play a pivotal role in healthcare and eldercare settings, assisting users with tasks that require balancing functionality with approachability. For instance, humanoids equipped with whole-body admittance control can integrate both center-of-mass (CoM) strategies and end-effector adjustments to provide physical support while maintaining stability in dynamic environments [1]. Advanced frameworks such as model predictive control (MPC) further ensure that the robot's physical interactions remain responsive to real-time uncertainties, optimizing leg mechanics for stability and coordination even in uneven terrain or dynamic environments [26]. Expressive robots offering subtle motion cues, such as compliant joint actuation while delivering medications or aiding in mobility, establish trust and comfort—a factor critical in eldercare and rehabilitation.

Service robotics places additional emphasis on seamless, contextual interaction within hospitality, retail, and public spaces. Expressive humanoids enhance task execution by combining precise, functional physical movements with naturalistic gestures and non-verbal cues. For example, hierarchical visuomotor controllers enable robots to integrate visual feedback with coordinated joint and motor responses, allowing for robust navigation and task-specific adaptations [91]. Similarly, approaches informed by reinforcement learning (RL) have been used to train robots to mimic human-like gestures that enhance engagement in retail or hospitality sectors, creating smoother user interactions [10]. However, high-dimensional action spaces inherent to humanoid systems remain a challenge when performing such tasks. Techniques such as hierarchical policies and motor primitive embeddings [54] are being explored to reduce the computational complexity of expressive motions, ensuring real-time adaptation without compromising naturalness.

Non-verbal communication is indispensable in human-robot collaboration, where expressive gestures enable robots to convey intent, acknowledge instructions, or synchronize tasks efficiently. For example, robots employing expressive motion can leverage synchronized hand and head gestures for disambiguating multistep tasks or signaling intent in ambiguous scenarios. Techniques like dynamic movement primitives (DMPs) [93] and neural probabilistic modeling [54] allow fine-grained synthesis of motion patterns, enabling humanoids to respond to human gestures in tactile or auditory interactions. Moreover, by coupling inverse kinematics frameworks with adaptive whole-body controllers, expressive robots can avoid self-collisions during high-dimensional gestures [80], enabling more effective collaboration in constrained environments.

Nonetheless, achieving robust, real-world deployment of expressive humanoid collaboration systems presents substantial challenges. Simulation-to-reality transfer remains a bottleneck, as discrepancies between simulated training environments and physical robots often hinder generalization and robustness [12]. Techniques such as domain randomization [11] and adaptive augmentation strategies show promise in mitigating these challenges, improving policy transfer for interaction-rich environments. Meanwhile, there is growing interest in expanding emotional expressivity through latent gesture spaces linked to user feedback, such as mapping trajectories to latent Valence-Arousal-Dominance (VAD) spaces for social robots [83].

Future research must operationalize innovations by bridging control methodologies—such as hybrid MPC-RL systems that balance precision with adaptability [13]—to achieve synergistic design. Priorities include broadening human-likeness in motion representation while ensuring energetically efficient behaviors necessary for sustained interaction. As expressive humanoid collaboration progresses, there is a critical need for standardized metrics evaluating motion fluency, responsiveness, and sociability under real-world conditions, promoting accountability in human-robot partnerships. Interdisciplinary advancements in biomechanics, control theory, and human-computer interaction remain essential to fully harness the transformative potential of expressive humanoid robots in collaborative domains.

7.2 Entertainment and Social Engagement

Entertainment and social engagement represent some of the most compelling arenas where expressive humanoid robotics showcase their potential, leveraging advanced motion generation and control techniques to interact meaningfully with human audiences. By simulating lifelike gestures, emotions, and behaviors, humanoid robots enhance immersive qualities in creative and social experiences, captivating audiences in theater, gaming, and art installations. The integration of predictive control frameworks and data-driven learning enables these robots to dynamically adapt to evolving environmental and participatory constraints, heightening their ability to perform expressive, synchronized, and responsive movements.

In theatrical performances and choreographed dance, expressive humanoid robots have garnered attention for embodying both predefined and improvisational roles. Model Predictive Control (MPC) frameworks enable precise synchronization and real-time adaptability, particularly in stage scenarios requiring seamless integration of dynamic movements. Frameworks such as IS-MPC [16] and hierarchical task-based architectures [30] support stable yet dynamic gait transitions and balance recovery, ensuring safety in fast-paced stage environments while achieving aesthetically fluent movements aligned with artistic intent. However, achieving the biomechanical nuance characteristic of human performers remains a challenge, as trajectory optimization methods often lack the flexibility to produce highly expressive motions unless redesigned with carefully tuned constraints [112]. Hybrid systems that merge deep reinforcement learning (RL) with MPC [29] show promise in overcoming these limitations, enabling robots to autonomously explore diverse and fluid expressive cues.

In gaming and virtual reality (VR), humanoid robots serve as interactive agents that physically embody in-game characters or guide participants in highly engaging ways. These dynamic interactions require holistic control of multi-contact scenarios and responsiveness to simultaneous environmental feedback. Multi-contact MPC frameworks [22] and adaptive control methods in stochastic contexts [27] allow robots to robustly navigate interactive gaming scenes, ensuring both participant safety and immersion. Furthermore, humanoid robots enhance VR experiences by mimicking human-like behaviors during player interactions, creating more realistic and engaging virtual environments. However, integrating humanoids within the physical constraints of gaming and VR environments requires innovations such as GPU-accelerated solvers [59] and contact stability optimization to meet the computational demands of predictive planning while maintaining seamless operation.

In the arts, expressive humanoid robots have emerged as collaborators or creators, making profound contributions to visual art, music, and kinetic sculptures. Advanced control frameworks, such as those incorporating motion primitives [107], enable these robots to follow personalized trajectories, mirroring rhythms, tempos, or gestures of human artists while fostering new modes of artistic expression. Robots equipped with orientation-aware MPC [35] further demonstrate capabilities such as traversing uneven terrains or dynamically moving within constrained spaces, enabling large-scale exhibitions and installations. Meanwhile, RL-driven creativity frameworks, where adaptive reward shaping prioritizes exploratory and innovative motion, allow robots to autonomously develop novel aesthetic behaviors [27]. Nevertheless, limitations in hardware actuators impose constraints on motion quality, restricting the range and richness of artistic expressiveness these robots can achieve.

Emerging trends position humanoid robots as both storytellers and performers, integrating expressive social cues into their repertoire. The incorporation of multi-modal sensing frameworks [43] enables robots to process and combine proprioceptive and auditory inputs, synthesizing gestures that resonate emotionally. For instance, auditory feedback systems are increasingly enabling robots to synchronize their movements in real-time with live music performances, leveraging hybrid RL-MPC systems [29]. As robots assume greater prominence in co-creative roles, ethical questions around authorship, bias in generated content, and copyright will need to be addressed, shaping the future of human-robot collaboration in the arts.

In summary, the deployment of expressive humanoid robots in entertainment and artistic domains underscores an intricate interplay between technological innovation and human creativity. Advancements in hybrid methodologies — blending physics-based MPC algorithms and data-driven RL frameworks — represent a vital pathway for robots to achieve the nuanced complexity required for creative industries. Balancing computational efficiency, biomechanical authenticity, and high-level expressiveness will remain pivotal in unlocking the full potential of humanoid robots as collaborators in entertainment and the arts.

7.3 Physical Tasks Requiring Dexterity

The ability of humanoid robots to execute physical tasks that require precise dexterity, adaptability, and robust whole-body coordination is critical for applications in industrial automation, healthcare, and service robotics. These tasks often entail seamless integration of fine motor control, dynamic balancing, and anticipatory adjustments, necessitating the synthesis of advanced whole-body control frameworks. Expressive humanoid robotics, leveraging techniques such as Model Predictive Control (MPC) and Reinforcement Learning (RL), has enabled significant strides in this domain by combining physical feasibility with task-specific adaptability and robustness.

Industrial and manufacturing applications represent a primary domain that showcases the necessity for dexterous humanoid robotics. These robots are not only tasked with performing repetitive functions, such as assembly or sorting, but also complex manipulation tasks requiring adaptive and safe interaction with their environment. For example, planning algorithms embedded in MPC frameworks have been shown to deliver robust real-world functionality by predicting and coordinating multi-body dynamics, as seen in applications like dynamic box manipulation, where robots adaptively lift, balance, and transport loads while maintaining stability [38]. Furthermore, state-of-the-art reinforcement learning approaches address challenges in handling complex action dynamics by training policies for reliable multi-step manipulation tasks in unstructured settings, including carrying non-rigid objects while managing external forces and perturbations [113].

Adaptive locomotion combined with dynamic manipulation forms another frontier in tasks requiring dexterity, where humanoid robots must operate in environments with uneven terrains or shifting constraints. Advanced control frameworks employing bio-inspired concepts, such as angular momentum-based Linear Inverted Pendulum Models (ALIP), coupled with RL-guided high-level planners, have enabled humanoids to transition seamlessly between tasks, such as walking on complex surfaces and manipulating heavy payloads [78]. Similarly, MPC formulations incorporating orientation-aware dynamics and step optimization have showcased significant improvements in providing humanoids with the agility to maintain balance while executing demanding physical activities [35].

Tasks requiring precision, such as surgical assistance or high-accuracy instrument handling in delicate industrial workflows, demand integration of fine motor skills with overarching whole-body stability. Neural motor primitives using latent-variable bottleneck models have emerged as a promising approach to encapsulate motion diversity while maintaining robustness. These allow robots to adapt instantaneously to changing task specifications by leveraging modular skill embeddings for manipulatory subtasks like screw fastening or fragile-object handling [54]. In parallel, data-efficient learning methods combining imitation learning with robust optimization have exhibited great promise in enabling robots to mimic human-level dexterity in micro-motion tasks while mitigating the unrealistic torque limits imposed by hardware constraints [56].

A notable characteristic of developments involving whole-body dexterous tasks is the increasing reliance on hybrid control systems combining MPC and RL. For instance, hybrid frameworks exploit MPC's inherent stability guarantees to enforce strict torque and joint constraints, while RL-driven policies enhance adaptability for tasks such as error recovery during dynamic human-robot interactions or sudden external perturbations [13]. Such systems have successfully generalized performance across a spectrum of dexterous activities, spanning from grasp-and-transport scenarios to maintaining robustness under push disturbances across diverse operational conditions [34].

Despite these advancements, challenges persist in achieving greater scalability of these control paradigms. High degrees of freedom (DoFs) in humanoid robotics impose significant computational overhead, particularly in high-speed, multi-contact tasks, limiting their real-time applicability. Additionally, energy efficiency and hardware wear-and-tear during extended periods of operation remain key implementation barriers. Emerging approaches, such as hierarchical policy architectures and task-space control frameworks, show potential for mitigating these limitations by enabling decomposed task learning in shared latent spaces, thereby reducing computational and sample complexity [104].

Future directions in this field must emphasize integrating multi-modal sensory feedback. Visual and tactile inputs can enrich proprioceptive guidance, allowing humanoid robots to better understand and adapt to subtle environmental changes. The development of physically consistent training simulators and standardization in dexterity benchmark datasets could further enhance the assessment and refinement of such control policies [39]. By addressing these ongoing challenges, expressive humanoid robotics is poised to become a pivotal force in automating complex dexterous tasks across diverse domains.

7.4 Enhancing Accessibility and Rehabilitation

The application of expressive humanoid robotics in accessibility and rehabilitation represents a transformative intersection of technology and human-centric care, building upon the advanced motor control and adaptive learning paradigms discussed earlier. These systems, equipped with expressive whole-body motions and robust control algorithms, effectively address physical, cognitive, and emotional barriers faced by individuals with disabilities. This subsection delves into the methodologies, trade-offs, and emerging trends that underpin the deployment of these robots in accessibility enhancement and therapeutic contexts, emphasizing their capacity to empower users and improve rehabilitation outcomes.

As assistive systems, expressive humanoid robots have demonstrated remarkable utility for individuals with mobility constraints or communication impairments. By leveraging optimized whole-body coordination algorithms, these robots facilitate activities such as walking assistance or object handling, employing dynamic Model Predictive Control (MPC) frameworks to ensure stability and real-time adaptability [18], [114]. For instance, dynamically optimizing contact forces and trajectories allows these robots to navigate uneven terrains, offering an alternative where traditional devices, like wheelchairs, may falter. Furthermore, reinforcement-learning-augmented systems excel in enabling humanoid robots to mimic user-specific gait patterns over time. Training frameworks that combine motion imitation with hierarchical learning [10] allow tailoring mobility support to individual users, aiding in the recovery or improvement of motion capabilities lost due to conditions such as stroke or other neurological impairments.

Rehabilitation therapy, another critical domain, is also benefiting from the involvement of expressive humanoid robots. Therapists often rely on repetitive exercises to restore motor functions, but these can be monotonous and challenging to monitor in traditional settings. Robots utilizing hierarchical whole-body MPC frameworks [30] inject stability and adaptability into these exercises by adjusting task difficulty in response to patient progress. This adaptability, achieved through personalized MPC-optimized trajectories or reinforcement learning paradigms applied alongside motion imitation learning [10], helps maintain user motivation and engagement. Real-time human-robot interaction is also enhanced by embedded motion tracking algorithms, providing clinicians with precise progress metrics and enabling highly targeted, personalized therapy regimens.

Beyond physical rehabilitation, the cognitive and social dimensions of therapy demonstrate the unique advantages of expressive humanoid robots. Through multimodal feedback mechanisms that integrate visual, tactile, and auditory channels [22], these robots recreate the dynamics of human-like social interactions. This capability is particularly valuable in addressing conditions such as autism spectrum disorder (ASD), where engagement with emotionally expressive robots can reduce social barriers and encourage the adoption of positive behaviors. Reinforcement learning algorithms combined with bio-inspired motion controllers [10] further refine the ability of these robots to connect at an interpersonal level, facilitating emotional bonding and reducing isolation, especially in pediatric therapy settings. Such emotionally resonant interactions make these robots highly effective in play-based interventions aimed at developing social and emotional skills in young users.

Nevertheless, challenges remain as these robots transition from controlled environments to broader real-world applications. The sim-to-real transfer problem continues to pose a significant hurdle, as the robustness of learned behaviors can degrade amidst the noise and unpredictability of unstructured settings [10], [48]. Promising solutions, such as domain randomization and online adaptation techniques, aim to enhance the reliability and deployment potential of these robots in dynamic environments [48]. Additionally, energy efficiency emerges as a concern, given the resource demands of sustaining expressive, high-fidelity motions. Innovations like lightweight hardware and low-resolution MPC solvers [106] are crucial to managing these constraints, extending operational longevity and supporting effective real-world use.

Emerging research directions offer further optimism for the field. The integration of bio-sensing technologies such as electromyography (EMG) and electroencephalography (EEG) with humanoid control systems promises biofeedback-driven rehabilitation [96]. By closing the feedback loop between user physiology and robot behavior, these advancements enable precisely customized therapies attuned to individual patient needs. Concurrently, ongoing development of multi-task and scenario-generalizable controllers [29] anticipates a future where humanoid robots can seamlessly transition between assistive functions and rehabilitation programs, enhancing their adaptability and expanding their roles in accessibility and care.

In conclusion, the deployment of expressive humanoid robotics in accessibility and rehabilitation showcases the profound synergy between advanced control technologies and human-focused applications. By effectively addressing challenges such as energy constraints, sim-to-real transfer limitations, and scalability in adaptive learning frameworks, these robots are poised to usher in equitable and impactful assistive technologies. As interdisciplinary research continues to drive technological and methodological advancements, these systems hold immense potential to foster independence, improve quality of life, and humanize care through sophisticated motion and interaction capabilities.

8 Evaluation Metrics, Datasets, and Benchmarking Frameworks

8.1 Metrics for Stability, Efficiency, and Expressiveness

Evaluating the performance of humanoid robots in terms of *stability*, *efficiency*, and *expressiveness* is critical for assessing their capability to execute dynamic and human-like whole-body motions in real-time. These metrics serve as a foundation for understanding how control strategies succeed or fail in replicating human behaviors while recognizing trade-offs in computational cost, adaptability, and robustness.

Stability metrics are inherently tied to the ability of humanoid robots to maintain balance under various dynamic conditions, and they often draw from classical control theories as well as biomechanical models of human motion. One standard metric for evaluating stability is the Zero Moment Point (ZMP), which ensures that the center of pressure lies within the robot's support polygon, a critical factor in preventing falls. Complementary metrics such as the Center of Mass (CoM) deviation or capture point estimation are also frequently utilized to measure disturbance rejection and predict potential recovery behavior during high-impact motions [18]. Advanced control systems, including hierarchical inverse dynamics models and nonlinear Model Predictive Control (MPC), incorporate real-time stabilization strategies while accounting for contact forces and hybrid motion dynamics [50] [4]. Balancing tasks often extend to dynamic environments where contact locations or external disturbances occur unpredictably; robust stability under such conditions is increasingly being analyzed through dynamic complementarity constraints and viability kernels [19] [26].

Efficiency, a vital metric, often focuses on evaluating energy consumption and resource allocation during complex humanoid motions. Metrics such as mechanical work, torque utilization, and actuator power expenditure provide insight into how optimally a humanoid achieves its tasks without over-utilizing mechanical components. Energy efficiency is of particular concern in aggressive maneuvering tasks such as jumping or high-speed locomotion, where maintaining energy balance significantly impacts hardware longevity. The rise of frequency-aware MPC demonstrates advancements in achieving smooth and energy-efficient motion trajectories without compromising task feasibility, even under bandwidth-limited actuation systems [23]. Emerging frameworks like Control Lyapunov Function (CLF)-MPC hybridizations specifically embed energy-optimal trajectories to adapt behaviors dynamically based on the robot's payload or terrain [14]. Furthermore, lightweight humanoid platforms such as "Adam" [115] highlight the role of optimized mechanical design in improving both energy efficiency and dynamic performance.

Expressiveness metrics, conversely, build on the qualitative emulation of human-like motions, focusing on fluidity, naturalness, and task-relevant adaptiveness. Quantitative measures include trajectory smoothness (e.g., jerk minimization), motion naturalness indices, and deviation from reference human motion trajectories. Expressiveness also overlaps with higher-level human-robot interaction goals, where robots must perform adaptive, socially intelligible movements. Residual Force Control (RFC)-based frameworks exemplify how humanoids can handle significant dynamics mismatches while synthesizing complex, graceful behaviors such as ballet or emotionally resonant gestures, by seamlessly augmenting their physical limitations with optimized control policies [8]. Reinforcement learning-based techniques have further enabled humanoids to learn nuanced expressive motions by combining imitation data with reinforcement objectives, balancing biomechanical feasibility with aesthetic accuracy [3] [11]. Additionally, systems like OmniH2O integrate whole-body kinematic poses with multimodal sensory control to boost expressiveness across tasks such as manipulation and dynamic walking [116].

Despite these developments, trade-offs remain across stability, efficiency, and expressiveness. High stability-centric strategies, such as strict ZMP controllers, frequently suppress natural movement tendencies, while expressiveness targets tend to push computational and energy demands higher. Similarly, achieving long-term balance and expressiveness simultaneously, especially in unstructured environments, remains an unsolved challenge. Benchmarking efforts are further complicated by the lack of universally applicable datasets and standardized protocols. Motion retargeting datasets and frameworks, such as those leveraging human motion capture data, have expanded test scenarios, but gaps remain in cross-platform generalization [95] [56].

Future work must focus on unifying these metrics into cohesive benchmarking protocols. For instance, hybrid control systems that integrate MPC and reinforcement learning demonstrate early promise in balancing precision and adaptability across stability, efficiency, and expressiveness domains [13]. Incorporating biomechanical priors, such as joint-level flexibility and human-effort models, may further refine comparative evaluations under a broader spectrum of motions and environmental contexts. As humanoid robotics progress, the interplay of these performance dimensions will guide design decisions toward developing robots that are robust yet relatable in human-centric applications.

8.2 Simulation and Testing Platforms

Simulation and testing platforms are indispensable for advancing the performance, stability, and expressiveness of humanoid robot motion control systems. Acting as controlled environments, these platforms enable the replication of physically complex scenarios, enforcement of model constraints, and iterative refinement of control algorithms before real-world deployment. In the context of expressive whole-body control via Model Predictive Control (MPC) and Reinforcement Learning (RL), simulation environments play a critical role in systematically analyzing and improving hybrid frameworks. By enabling evaluations under diverse and realistic conditions, they facilitate advancements in adaptive, high-dimensional control strategies.

Among the most widely adopted platforms in humanoid robotics simulation is MuJoCo, valued for its accurate physics engine and flexibility in modeling high-degree-of-freedom (DoF) robots. Its compatibility with Reinforcement Learning frameworks further enhances its applicability to tasks such as walking, jumping, and contact-rich interactions [117]. However, the computational demand of MuJoCo can limit its effectiveness for real-time applications, especially in highly dynamic scenarios requiring long-term predictive horizons. Alternatively, Gazebo is popular for its rich sensor simulation capabilities, making it highly suitable for human-robot interaction tasks. Gazebo's modularity supports diverse applications, though inconsistencies in performance metrics persist when dealing with complex whole-body control [118]. Another key platform, DART, excels in handling agile and dynamic humanoid motions, particularly in scenarios involving impact-aware control and stable contact dynamics [75].

A primary trade-off inherent in these platforms revolves around balancing simulation fidelity and computational efficiency. High-fidelity simulations, including Nonlinear MPC applications, are instrumental for understanding complex hybrid systems, such as centroidal momentum control or adaptive contact dynamics [60]. Nevertheless, even advanced solvers—such as Sparse Quadratic Programming—struggle to meet real-time control rates when simulating large humanoid robots exceeding 32-DoF [102]. These computational limitations often necessitate approximations, prioritizing real-time responsiveness at the expense of precision. To address resource constraints, emerging frameworks like TinyMPC leverage efficient numerical backends optimized for low-compute hardware [106]. While promising for cost-effective systems, these approaches are not yet well-suited for the high-dimensional motion evaluations required in expressive humanoid controllers.

Open-source solutions such as the Control Toolbox (CT) are increasingly being utilized to benchmark and prototype MPC and RL algorithms. By incorporating automatic differentiation techniques and derivative code generation, CT simplifies the design process, allowing developers to implement embedding constraints and optimize parameters with greater ease [118]. Its versatility makes it suitable for benchmarking various tasks, including locomotion, manipulation, and collaborative human-robot tasks, bridging gaps between simulation and real-world fidelity.

Despite their capabilities, current simulation platforms face significant challenges in meeting the demands of expressive whole-body control. For instance, simulating soft contacts, tactile interactions, and vision-driven biomechanical adaptations remains problematic [28], [76]. Moreover, accurately modeling motion qualities such as smoothness and naturalistic human expressiveness is a work in progress. Recent initiatives like HumanoidBench have begun standardizing expressive motion testing across various control architectures, although integrating sparse, customizable reward functions into these benchmarks remains an ongoing research challenge [117].

Looking toward the future, the integration of physics-informed machine learning has the potential to revolutionize simulation platforms for humanoid control, offering a means of closing the gap between simplified models and real-world performance. Techniques such as domain randomization and adversarial training are poised to enhance sim-to-real transfer, particularly in the context of RL-based optimization [14], [27]. Additionally, future platforms will need to accommodate advanced control demands, such as neuromorphic or bio-inspired behaviors, and expand their capabilities to support dynamic, perception-driven tasks. Developing scalable, multimodal, and modular simulation environments will be essential to enabling the iterative progression of expressive humanoid robotics and addressing the benchmarks outlined in subsequent sections.

8.3 Specialized Datasets for Humanoid Expressions

The use of specialized datasets has been pivotal in advancing expressive whole-body control for humanoid robots, especially as researchers strive to replicate human-like motion and behavior. These datasets serve as the foundation for training, evaluating, and benchmarking control frameworks, ensuring that the synthesized motions are not only functional but also expressive and naturalistic. This subsection discusses key datasets used in humanoid robotics, evaluates their suitability for expressive control, and highlights challenges and future directions.

Human motion capture (mocap) datasets remain the cornerstone for training humanoid robots to emulate expressive whole-body behaviors. Datasets such as the CMU Motion Capture Dataset and Human3.6M offer extensive annotated motion sequences, capturing diverse movements like walking, gesturing, dancing, and interactions. The usability of such datasets lies in their high temporal and spatial resolution. However, their direct application to humanoid robotics poses challenges due to discrepancies between the high degrees of freedom (DoFs) in human anatomy and the constrained DoFs of humanoid robots [3]. Techniques such as motion retargeting and reformatting into robot-compatible trajectory spaces have proven effective in mitigating these gaps. For instance, methods that embed retargeted motions into latent representations compress high-dimensional human motion data, enabling efficient imitation and interpolation of expressive behaviors [8], [54].

Beyond open human mocap datasets, task-specific datasets tailored to humanoid applications are gaining prominence. For example, datasets developed for push recovery scenarios capture dynamic responses to external perturbations, providing valuable benchmarks for robustness assessments [36]. Similarly, datasets for complex multi-contact or dynamic behaviors, as showcased in research on loco-manipulation tasks [37], provide trajectories that integrate balance maintenance with dexterous manipulation. These datasets often include specific robot-centric annotations, such as center of mass (CoM) trajectories or ground reaction forces, which directly support control system optimization.

A notable trend in dataset usage for humanoid expressions is the integration of domain adaptation methodologies, particularly for sim-to-real challenges. Domain randomization applied to mocap-derived datasets has shown success in bridging simulation and reality by perturbing environmental and physical parameters during training [39]. This approach enables datasets to generate robust motion policies that generalize across real-world variability, as seen in applications like uneven terrain navigation and dynamic manipulation [103].

While reliance on mocap datasets and task-specific datasets has driven significant progress, emerging techniques contribute to expanding dataset utility and diversity. For example, large-scale augmentation pipelines that combine synthetic environments and human motion textures can generate novel trajectories for advanced expressive tasks, such as parkour and collaborative interactions [9], [116]. Importantly, multimodal datasets encoding visual, tactile, and proprioceptive signals are starting to enable learning expressive behaviors that integrate perception-driven dynamics, crucial for social and emotional expression tasks [28].

Despite these advancements, critical gaps remain in the field. First, the lack of standardized, humanoid-specific datasets hinders direct benchmarking across different control frameworks. Existing datasets often focus on specific behaviors or robot configurations, making comparative evaluations challenging [56]. Additionally, the need for datasets that span longer time horizons and support hierarchical control frameworks is pressing, as many applications require sequence-level understanding of motion transitions [79]. Techniques to encode task-reward relationships into datasets, as demonstrated in research leveraging potential-based rewards [82], are promising avenues for future work.

In conclusion, specialized datasets have unlocked rich avenues for expressive whole-body control in humanoid robotics, though further standardization and augmentation are needed to achieve comprehensive, cross-domain applicability. Moving forward, the creation of open-access, task-diverse, and modality-rich datasets tightly aligned with benchmarking protocols could catalyze broader adoption and innovation in the field. Simultaneously, improvements in sim-to-real augmentation techniques and hierarchical motion representation will ensure that datasets remain scalable, interpretable, and relevant for emerging applications in dynamic and social environments.

8.4 Benchmarking Frameworks and Protocols

As humanoid robots increasingly assume roles in interactive and dynamic applications, the need for robust and standardized benchmarking frameworks for evaluating control paradigms becomes paramount. Such frameworks play a crucial role in ensuring the reproducibility of results, facilitating cross-method comparisons, and accelerating the development of expressive whole-body control systems. They address critical gaps in assessing diverse performance metrics such as stability, efficiency, adaptability, and expressiveness. Building on the role of specialized datasets outlined in the previous subsection and serving as a foundation for subsequent discussions on advanced control paradigms, this subsection examines existing approaches, highlights current challenges, and explores strategies for advancing benchmarking in humanoid robotics.

Current benchmarking frameworks for humanoid robotics control have primarily centered on task-specific performance evaluations. Many studies rely on scenario-based testing, where controllers are assessed on tasks such as walking, balancing, or manipulation under predefined disturbances or environmental variations. For example, the benchmarking of Divergent-Component-of-Motion (DCM) architectures for walking control has focused on analyzing position and velocity control across varied hardware setups [109]. While effective for specific tasks, these approaches often lack the flexibility to generalize across novel or unforeseen scenarios, limiting their utility in comprehensively evaluating control strategies for diverse operational settings.

A core challenge in benchmarking arises from the heterogeneous nature of humanoid robots, with variations in hardware components such as degrees of freedom (DoFs), actuator capabilities, and sensory configurations exerting substantial influence on performance outcomes. Comparisons of optimization-based controllers, such as Model Predictive Control (MPC) schemes, often depend heavily on the dynamic constraints and computational capabilities unique to each robotic platform [7], [97]. These dependencies underscore the importance of standardized, architecture-agnostic metrics in benchmarking. Foundational metrics such as Zero Moment Point (ZMP) convergence for stability, energy efficiency trade-offs, and expressiveness quantifiers (e.g., smoothness or trajectory deviation from human motion patterns) must form the baseline for future protocols [12], [18].

Efforts to enhance benchmarking reproducibility and scalability have increasingly leveraged simulation platforms integrated with benchmarking protocols. High-fidelity platforms like MuJoCo and Gazebo enable researchers to simulate complex scenarios and test whole-body behaviors under controlled conditions [117], [119]. However, these platforms face limitations, such as inconsistencies in task definitions and ground truth data across different environments. Domain randomization techniques have begun to address challenges related to the simulation-to-reality gap by perturbing environmental parameters during training, but they require standardized validation strategies on physical systems [46]. Developing open-source simulators with adjustable domain settings and globally recognized metrics can therefore enhance reproducibility and scalability.

Emerging multi-criteria benchmarking frameworks recognize the importance of evaluating hybrid control strategies, particularly those combining Model Predictive Control and Reinforcement Learning (MPC-RL), across a spectrum of tasks. For example, RL-augmented MPC methods validated through physical experiments have been shown to maintain system efficiency while recovering from external disturbances, making them well-suited for dynamic locomotion, robust manipulation, and interactive applications [29], [30]. To accurately assess the scalable adaptability of such hybrid systems, future benchmarking protocols must integrate evaluations across diverse and multi-task scenarios, ensuring cross-domain applicability and robustness.

Despite the progress made, significant challenges remain for benchmarking frameworks to scale to the full spectrum of high-dimensional and expressive behaviors expected in humanoid robots. For instance, designing evaluation criteria to assess performance in complex, multi-lateral tasks such as synchronized loco-manipulation is still an open issue [22]. Additionally, the standardization of evaluation datasets and public repositories remains limited. Existing human motion datasets, such as Human3.6M and the CMU Motion Capture database, provide valuable resources but require expansion to support humanoid-specific tasks fully. Data augmentation techniques, including probabilistic adjustments like Gaussian modeling, could complement these datasets and allow for a broader range of realistic and scalable evaluations [43].

Future progress in the field hinges on the development of task-agnostic benchmarking protocols that encompass critical metrics, including stability, energy efficiency, and expressiveness, across a wide range of operational contexts. Integrating hardware-neutral evaluation frameworks with ontology-based task definitions can further enhance transparency, reproducibility, and scalability. By fostering community-wide collaboration and embracing open benchmarking repositories that evolve alongside ongoing technological advancements, these frameworks will cement their role as pivotal enablers for fair, standardized, and effective evaluations in expressive humanoid robotics. This will not only strengthen comparative studies but also serve as an essential bridge to the advanced controllers and learning frameworks discussed in subsequent sections.

9 Challenges, Open Questions, and Future Directions

9.1 Bridging the Simulation-to-Reality Gap in Expressive Humanoid Motions

Humanoid robots hold great promise for effectively interacting in unstructured human environments, but the challenge of seamlessly transferring expressive whole-body control from simulation to reality remains a significant obstacle to their widespread deployment. This simulation-to-reality (sim-to-real) gap arises from discrepancies in modeling fidelity, unmodeled dynamics, sensor noise, and environmental variability, which lead to a divergence between simulated training outcomes and actual performance in the physical world. Addressing this gap for expressive humanoid motions involves systematically aligning simulated training environments with real-world dynamics and developing robust control and adaptation mechanisms.

A key strategy to bridge this gap is domain randomization, which involves exposing the control policy to varied dynamics and sensory conditions in simulation to build robust generalization. By randomly perturbing parameters such as friction, mass distribution, actuation delays, and sensor noise in the simulated environment, policies are better prepared for real-world uncertainties. For example, domain randomization has been shown to improve the robustness of humanoid locomotion and manipulation tasks by enabling controllers to adapt to diverse scenarios during training [11], [12]. However, while domain randomization can enhance robustness, overly wide parameter ranges may lead to unnecessarily conservative policies, reducing task-specific expressivity.

In addition to randomization, high-fidelity physics simulations are critical for narrowing the gap. Frameworks that leverage advanced physics engines, such as MuJoCo or Bullet, can capture detailed interactions at contact points and model actuator dynamics with greater precision, enabling accurate control transfer [119], [120]. Nevertheless, computational complexity and real-time constraints limit the incorporation of highly nonlinear dynamics and contact-specific parameters. Simplified models such as the Linear Inverted Pendulum Model (LIPM) remain widely used, but their limitations must be mitigated through novel hybrid paradigms integrating dynamics optimizations with learned corrections [5], [31].

Another promising approach is adversarial training, where policies are trained against simulated perturbations or adversarial environments to enhance robustness under extreme conditions. Residual Force Control (RFC), for example, augments learned policies with external corrective forces, enabling humanoid robots to stabilize dynamic, human-like motions such as ballet dancing and rapid transitions [8]. Similarly, adversarial perturbation modeling has shown success in push-recovery tasks, where robust gait stability is achieved by incorporating feedback-based adaptation to disturbances [18], [26].

Sensor fusion and real-time feedback mechanisms are also crucial in overcoming dynamic real-world uncertainties. Adaptive frameworks that integrate visual, tactile, and proprioceptive inputs enable humanoids to perceive and respond to unexpected changes in physical environments, such as uneven terrain or external forces. For instance, incorporating contact dynamics into predictive control frameworks allows for online adjustment of stepping trajectories, significantly enhancing stability and gait adaptability [5], [53]. However, achieving real-time feedback integration remains restricted by hardware processing speeds and communication latencies.

Evaluating sim-to-real transfer robustness presents its own challenges. Systematic benchmarking frameworks are required to assess robustness across diverse tasks, environments, and robot morphologies. Benchmarking studies have highlighted the critical role of simulation fidelity and task diversity in evaluating the performance of control policies [12], [82]. Moreover, metrics such as trajectory compliance, stability indices, and energy efficiency provide quantitative insights into sim-to-real performance gaps.

Future solutions necessitate a multi-pronged approach combining advancements in transfer learning, adaptive fine-tuning, and hybrid methodologies. Techniques such as zero-shot sim-to-real transfer that directly deploy policies from simulation onto hardware without additional tuning show promise. Notably, frameworks that incorporate large-scale human motion data, such as Human3.6M, with sim-to-real methods have generated policies that achieve unprecedented motion expressiveness in dynamic real-world tasks like dancing and object manipulation [8], [121]. Continued research must focus on integrating biomechanical insights, optimizing hardware-software co-design, and developing reliability safeguards to enable humanoids to seamlessly perform expressive, versatile motions in complex, unstructured environments.

9.2 Scalability and Adaptability in High-DOF Humanoids

The complexities of scaling expressive control frameworks to humanoids with high degrees of freedom (DoF) remain a critical challenge in advancing adaptive whole-body control. As humanoid robots increase in mechanical redundancy and physical intricacy, so too does the need for scalable and adaptable motion control paradigms capable of harmonizing expansive configuration spaces with dynamic constraints. This subsection explores state-of-the-art methodologies for addressing these challenges, highlights their strengths and limitations, and identifies future research directions for achieving robust scalability and task adaptability in diverse real-world applications.

High-DoF humanoids require control strategies that not only manage the extensive coupling between their limbs but also account for intricate dynamic interactions with their environment. Model Predictive Control (MPC) has emerged as a foundational approach due to its proven robustness in navigating constraints. Yet, its direct application to high-DoF systems presents formidable computational hurdles. To mitigate these issues, hierarchical MPC formulations have gained traction, decomposing control into high-level task optimization and low-level trajectory execution. This modularization improves tractability while ensuring constraint satisfaction, as evidenced in applications such as whole-body motion planning [28], [31]. However, this division often incurs delays and compromises global motion optimality, a trade-off that is particularly pertinent when striving for expressive and fluid robot behaviors.

Advances in computational efficiency for high-DoF MPC often center on exploiting structure. Sparse formulations, which prioritize dynamically relevant variables, significantly alleviate computational demands in complex scenarios, such as multi-contact motion generation [21]. Likewise, techniques like GPU-based parallelization and warm-start initialization have accelerated solvers, enhancing the feasibility of real-time control for systems with substantial DoF [102], [106]. Despite these breakthroughs, the challenge of achieving control rates beyond 100 Hz for highly complex, whole-body controllers persists, limiting smooth real-time execution in physically dynamic settings.

Reinforcement Learning (RL) offers an alternative approach to navigating the expansive action spaces inherent in high-DoF humanoids. Latent-space encodings and motor primitives serve as effective dimensionality reduction techniques, enhancing RL's ability to efficiently learn and adapt expressive skills across different robot platforms [28], [29]. Furthermore, hybrid frameworks combining RL with MPC have demonstrated considerable promise in bridging scalability gaps. For instance, MPC-enforced constraints during RL training prevent unsafe behaviors, while RL's adaptability expands the action repertoire for unstructured environments [29]. These integrations illustrate the potential for achieving both task-specific expressivity and operational robustness.

Task adaptability, another cornerstone for high-DoF control, becomes particularly important when scaling motion policies across heterogeneous robots or scenarios. Transfer learning and domain adaptation techniques, such as parameterized dynamics models and probabilistic priors, provide mechanisms to address this challenge, ensuring resilience against variations in morphology [43]. Additionally, inertia-aware control frameworks have furthered capabilities for dynamic tasks such as jumping or traversing rough terrain by incorporating variable centroidal dynamics into planning [35], [97]. These systems illustrate the progress made in enhancing adaptability, yet multi-task transitions and long-term autonomy across diverse domains remain challenging.

One of the most pressing limitations lies in mitigating overfitting to specific operational scenarios, which can hamper the generalizability of high-DoF controllers. Hybrid hierarchical frameworks offer a promising direction, leveraging high-level planners to facilitate adaptive task-switching while retaining contextual awareness [24], [31].

To further this progress, the development of standardized benchmarks, such as those utilized in humanoid control challenges [117], becomes essential for fostering cross-platform scalability and evaluating a system's capacity to generalize across tasks and morphologies.

Looking ahead, the intersection of physics-based models with learning-based approaches signifies a pivotal frontier in scaling high-DoF control. Future frameworks must prioritize biomechanically informed strategies, enabling humanoids to mirror human-like motion dynamics while respecting structural constraints. Simultaneously, advancing computational efficiency—through more sophisticated solvers and the integration of hardware accelerators—remains critical for closing the performance gap in high-dimensional, real-time motion planning. By addressing these interwoven challenges, the field moves closer to realizing scalable, adaptive, and expressive whole-body control for humanoid robots operating in dynamic and unstructured environments.

9.3 Ethical and Societal Implications of Expressive Humanoid Robots

The ethical and societal implications of deploying expressive humanoid robots in sensitive application domains, such as caregiving, social interactions, and education, are multifaceted and demand rigorous scrutiny. The intrinsic human-like appearance and expressive capabilities of such robots are intended to foster intuitive and emotionally resonant interactions. However, these very characteristics invoke significant ethical challenges, including issues of trustworthiness, transparency, fairness, anthropomorphism, cultural sensitivity, and user-centered design.

Expressive humanoid robots in caregiving roles, for instance, aim to support aging populations with physical assistance and companionship. These applications bring forth questions about trust and dependency. Users may anthropomorphize such robots, attributing human-like intentions or capabilities that exceed their programmed functionality, potentially leading to misplaced trust. A critical consideration is the transparency of decisions and behaviors exhibited by humanoids, particularly when high-stakes or emotionally charged tasks, such as eldercare, are involved. Frameworks that integrate interpretable models for motion generation, such as those developed for dynamic complementarity conditions [19], may provide a foundation for increasing decision transparency, but similar strides are required for expressive interactions.

Another challenge lies in mitigating algorithmic bias in social expressivity. Reinforcement learning-based controllers, extensively used for humanoid robots, generally depend on training datasets with human motion capture or behavioral data [8], [55]. If these datasets inadvertently encode cultural, gender, or age-based biases, the resulting expressiveness of humanoid robots might reflect such discrepancies, perpetuating stereotypes or alienating certain demographics. For example, biases in motion imitation frameworks [121] have the potential to affect trust and inclusivity, particularly in diverse social environments. Addressing this requires representation-aware datasets and algorithmic fairness metrics tailored to expressive behaviors.

Furthermore, the deployment of expressive robots in educational contexts or social interactions raises concerns regarding emotional manipulation and autonomy. Unlike explicit control tasks, expressive gestures and behaviors often serve to influence human emotions subtly, creating ethical tension between designing robots for effective interaction and avoiding undue psychological influence. Studies showcasing whole-body control for interaction-rich scenarios [79], [122] underscore the importance of user-centered design paradigms to avoid exploitation of innate human tendencies to infer social meaning from gestures. The inclusion of formal safety and ethical constraints in reinforcement learning processes, such as those in training frameworks that build motion policies under structured constraints [33], offers promising directions for ensuring responsible robot interactions.

Cultural sensitivity and inclusivity are equally critical in shaping expressive robot behavior. Movements, postures, and gestures that are expressive in one cultural context might be inappropriate or offensive in another. As robots are deployed globally, adapting to culturally specific norms and expectations becomes an essential aspect of their design. For example, advancements in motion retargeting demonstrated in [40] could be extended to encode regional or cultural variations in expressiveness, enhancing acceptance and usability across diverse settings.

Interpretable and explainable frameworks for generating expressive motions will also be pivotal. Multi-modal learning frameworks that integrate visual and proprioceptive feedback [46] could be extended to provide contextual explanations for decisions, thereby fostering trust. Similarly, structured benchmarking protocols for evaluating expressiveness in sensitive environments, akin to the methodologies outlined in [12], may help assess and standardize ethical performance metrics in expressive whole-body control.

Looking toward the future, socio-technical collaborations between roboticists, ethicists, and sociologists are imperative to navigate these challenges. Innovations in simulation-to-reality transfer [39] that maintain ethical alignment from the simulation stage to deployment and incremental learning frameworks [123] offer potential to develop adaptive, ethical, and context-specific behaviors. While the promise of expressive humanoid robots is immense, their societal acceptance hinges on prioritizing ethical considerations at every stage, from algorithm design to long-term integration into human-centric environments.

9.4 Energy Efficiency and Hardware Limitations in Humanoid Motions

Humanoid robotic systems are often constrained by their hardware, particularly in terms of energy efficiency and mechanical design, which pose significant challenges to achieving prolonged and expressive whole-body interactions. The dual objectives of creating human-like motions and maintaining operational longevity must contend with the energy-intensive nature of actuators, suboptimal weight distribution, and inefficiencies in control architectures. This subsection examines these limitations, evaluates solutions proposed in the literature, and highlights promising pathways for advancing energy-efficient and hardware-constrained expressive humanoid motions.

Energy consumption in humanoid robotics is largely driven by the reliance on high-capacity actuators and motors. These are further burdened by mechanical inefficiencies in joint assemblies, creating trade-offs between energy efficiency and motion expressivity. For instance, torque-controlled approaches enable precise, dynamically adaptive motions but require considerable energy, exacerbating challenges in maintaining operational longevity. Model Predictive Control (MPC)-based algorithms have demonstrated potential for achieving precise, dynamically stable behaviors through optimization-based planning. Yet, their application to hardware-constrained robots remains limited due to the computational demands of real-time trajectory optimization [49], [67]. Nonlinear and resource-intensive dynamics formulations, such as those incorporating variable centroidal inertia within MPC, offer improved motion adaptability but further increase computational and energy demands [97].

Weight distribution and mechanical design also impose hardware-related barriers to energy efficiency. Lightweight yet robust structural components are critical to reducing energy consumption while maintaining force closure during dynamic multi-contact scenarios. Robots like RH5 illustrate the importance of hybrid series-parallel architectures, which enhance dynamic performance and payload handling without sacrificing structural efficiency [66]. Furthermore, compliance mechanisms in joint designs can mitigate torque demands and overheating, especially during repetitive or high-torque motions [50]. However, these benefits come at the cost of increased manufacturing complexity and higher implementation expenses, limiting their scalability.

Real-time optimization techniques tailored for low-resource systems represent a promising direction for overcoming these limitations. Lightweight hardware accelerators and solvers, such as GPU-optimized quadratic programming, have significantly improved the computational feasibility of MPC frameworks, facilitating their deployment on resource-limited platforms [106]. By enabling faster updates to resource-intensive optimization frameworks, these methods allow for adaptive tuning of planning horizons and prioritization of low-energy configurations. Cascaded fidelity approaches, which simplify dynamic models over larger prediction horizons, have also achieved notable gains in both computational efficiency and energy savings while preserving control accuracy [6].

Nonetheless, the robustness of hardware to wear and tear over extended operational durations remains a critical challenge. Mechanical fatigue, particularly in actuators subjected to high-frequency vibrations or high-torque requirements, reduces system longevity and introduces additional energy inefficiencies over time. Dynamic parameter estimation and hybrid adaptive control frameworks, such as hybrid learning–MPC systems, propose promising solutions by introducing redundancy within task-specific control mechanisms [71], [96]. These approaches also facilitate predictive maintenance, improving long-term energy efficiency and minimizing operational disruptions.

Emerging technologies present additional opportunities to address energy efficiency constraints without compromising expressivity. Variable-stiffness actuators and passive mechanisms, for example, allow expressive actions while redistributing energy demands during stabilization tasks [46]. Similarly, bio-inspired control paradigms derived from human biomechanics have demonstrated reduced energetic footprints in humanoid locomotion while preserving natural and expressive movement patterns [10].

Looking ahead, advancing energy-efficient, hardware-conscious humanoid motions will require innovations in ultra-lightweight materials with high structural integrity, low-power high-torque actuators, and modular hardware designs optimized for task-specific functionality. Energy-harvesting mechanisms, such as regenerative braking for actuators, hold particular promise in transforming energy dynamics in robotics [48]. Furthermore, hybrid MPC–Reinforcement Learning frameworks, where RL-derived policies prioritize energy-efficient behaviors within the constraints of MPC, offer a synergistic pathway toward balancing energy conservation with expressive capabilities [29], [124]. Finally, fault-tolerant designs that ensure resilience to mechanical failures will play a pivotal role in creating energy-efficient, reliable robots capable of safe and expressive interactions across varied operational domains.

9.5 Developing Unified Evaluation Metrics and Benchmarking Protocols

Unified evaluation metrics and benchmarking protocols are pivotal to advancing expressive whole-body control in humanoid robotics. Despite significant developments in Model Predictive Control (MPC) and Reinforcement Learning (RL), the lack of comprehensive and standardized evaluation frameworks has impeded objective comparisons across methodologies, hindering cross-pollination of insights. This subsection discusses the need for establishing unified evaluation protocols, analyzing existing approaches, and proposing pathways toward addressing the gaps in benchmarking.

Expressiveness in humanoid robot motion involves a delicate interplay between stability, energy efficiency, and naturalness in dynamic, high-dimensional environments. While stability metrics like the Zero Moment Point (ZMP) and Center of Pressure (CoP) trajectories are well-established [18], they are often insufficient for capturing the nuanced dynamics of expressive behaviors. Metrics explicitly tailored for expressivity, such as trajectory deviation, smoothness indices, and biomechanical plausibility, remain inconsistently defined, limiting their applicability across applications. Studies such as [93] have incorporated smoothness and adaptability into motion frameworks but lack standardization for validation. Similarly, energy efficiency—a critical factor for real-time deployment—has been traditionally quantified through torque utilization and energy expenditure [34], yet adaptation to highly expressive motions remains a challenge due to dynamic task variability.

A major obstacle in benchmarking arises from the non-uniformity of task-specific evaluation protocols. Current approaches are often tailored to specialized use cases, such as multi-contact scenarios [73], locomotion on uneven terrain [5], or gesture-based interactions [46]. As tasks vary widely in complexity and constraints, existing protocols predominantly focus on single domains, making cross-method performance comparisons infeasible. Furthermore, while task-specific standardized datasets such as human motion capture datasets (e.g., CMU Motion Capture Dataset, Human3.6M) facilitate expressive motion imitation frameworks [10], simulation-to-reality transfer inconsistencies complicate their applicability in practical robotics systems.

Creating scenario-diverse and multi-criteria benchmarks holds immense potential for unifying evaluation. Multi-contact models like those reviewed in [21] highlight the importance of dynamic transitions across contacts, but more inclusive protocols are needed to evaluate adaptability across shifting constraints. Similarly, evaluations must address the interplay between precision and generalization, as hybrid control frameworks (e.g., MPC-RL approaches [29]) may excel in dynamic settings but falter under complex environmental interactions. A unified protocol could account for static metrics like minimum divergence from reference trajectories, while incorporating dynamic factors such as environment responsiveness.

To bridge the simulation-to-reality gap, reproducibility in hardware testing must also be emphasized. Advanced simulation platforms such as MuJoCo [119] and Gazebo offer controlled experimental settings but often fail to model nuanced properties such as soft impacts and friction inconsistencies. Domain randomization techniques address some stochastic discrepancies in policy evaluation [69], suggesting an urgent need for incorporating these into benchmarking frameworks to better simulate real-world conditions.

Another key consideration is the open sharing of datasets, tools, and protocols within the research community. Recent hardware demonstrations, such as HECTOR's dynamic loco-manipulation [37], serve as valuable exemplars but underscore the necessity of creating accessible datasets for validating control mechanisms under real-time dynamic scenarios. Encouraging collaboration by establishing open-access repositories could standardize diverse experimental conditions, fostering reproducibility.

Future efforts should also prioritize multi-objective benchmarking frameworks integrating expressivity, stability, safety, and efficiency. Techniques like Bayesian optimization for metric balancing [26] or multi-phase optimization to accommodate hierarchical task dependencies [24] could streamline evaluations of simultaneous locomotion and manipulation tasks. Furthermore, incorporating advanced sensors and multimodal feedback mechanisms [91] promises to refine the quality and breadth of scenario evaluations, catalyzing breakthrough developments in expressive humanoid control.

In conclusion, developing unified metrics and protocols will require collaborative standardization across methods and domains. Standardized evaluation criteria and task-specific benchmarks should reflect the field's diversity, accounting for whole-body dynamics, expressivity, safety, and environmental adaptability to drive tangible progress in humanoid robotics research.

9.6 Emerging Directions and Long-Term Research Opportunities

Building on recent advancements in expressive humanoid robotics, this subsection investigates transformative long-term research trajectories designed to address foundational limitations and unlock new applications, thereby advancing the broader goals of standardized and expressive whole-body control. A persistent challenge remains the synthesis of precise, reactive behaviors with robust adaptability, necessitating frameworks that deeply fuse physics-driven optimization and machine learning paradigms to meet the complex requirements identified in existing benchmarking protocols.

One particularly promising avenue lies in the development of hybrid systems that integrate the strengths of Model Predictive Control (MPC) and Reinforcement Learning (RL). While MPC excels in precision and enforces biomechanical constraints, its reliance on computationally intensive solvers limits real-time adaptability in high-degree-of-freedom (DoF) systems [16]. Conversely, RL offers the ability to learn adaptive, task-specific policies but often lacks the stability guarantees critical for safety-sensitive applications. Emerging hybrid frameworks, where MPC serves to constrain RL's exploration during training, have shown promise in generating more reliable policies for expressive humanoid motions [125]. Future research should expand on hierarchical architectures that pair RL-induced task planning with real-time constraint handling by MPC, leveraging advancements in parallel hardware acceleration for improved scalability and robustness in both training and deployment [94].

In parallel, closing the gap between robotics and biomechanics provides a crucial research direction. Many existing humanoid control paradigms rely on oversimplified dynamic models, limiting their capacity to capture the nuanced, non-linear coordination observed in human motion. Recent developments, such as variable inertia models that dynamically adapt upper-body and limb dynamics to improve stability and energy efficiency, demonstrate the potential of biomechanics-inspired techniques [97], [126]. Incorporating human-motion data into learning frameworks could enable humanoid robots to move beyond generic behavior and achieve highly expressive, naturalistic actions, such as soccer kicks and rapid arm movements [96], [127]. Future strides should focus on embedding musculoskeletal insights from simulations into whole-body controllers, fostering biomechanical compliance that enhances both motion realism and task adaptability within the diverse scenarios proposed by unified evaluation protocols.

Equally significant is the development of collaborative frameworks for multi-agent humanoid-robot interaction, a critical frontier highlighted in benchmarking efforts. Tasks requiring synchronized movements, whether involving multiple robots or human-robot interactions, challenge traditional design paradigms by demanding unparalleled adaptability and reactive capabilities. Existing approaches, such as multi-contact MPC frameworks for dynamic loco-manipulation, provide a strong foundation but are not yet scalable to complex, interactive environments [22]. Addressing these challenges will require exploration into probabilistic and graph-based planning schemes capable of predicting human intent and aligning robot trajectories in dynamic, unstructured environments [128]. These efforts promise new opportunities in human-humanoid collaboration and task orchestration.

Beyond physical adaptability, augmenting emotional and social intelligence through multi-modal learning represents a vital opportunity for enhancing expressivity in humanoid robots. Integrating sensory data from visual, auditory, tactile, and proprioceptive modalities enables robots to generate nuanced, context-aware gestures and expressions aligned with human-robot collaboration paradigms [46], [128]. Reinforcement learning approaches tailored to emotional and social behaviors can open new avenues for humanoids in caregiving, entertainment, and education by advancing their ability to interact meaningfully in human-centered environments [28]. Continued advancements will necessitate deeper integration of sensory data streams and adaptive learning pipelines, fostering high-quality, versatile interactions.

Energy efficiency remains another critical bottleneck in extending the operational feasibility of humanoid robots, a factor increasingly emphasized in multi-objective benchmarking criteria. While preliminary progress has been made through lightweight actuators and adaptive control algorithms [110], future paradigms must directly integrate energy dynamics into motion planning frameworks. Balancing expressivity, stability, and resource constraints stands as a fundamental trade-off, with dynamic time-optimal trajectory planning offering promising pathways for mitigating energy inefficiencies [60].

Lastly, the creation of open-access datasets and standardized benchmarking protocols, as emphasized in previous subsections, remains an overarching necessity. Current advancements in simulation training and domain-randomization techniques address some of the challenges associated with the simulation-to-reality gap, but their limited diversity and physical accuracy still hinder progress. Initiatives to develop task-diverse and physically consistent benchmarking frameworks hold great potential to address these shortcomings [50], [108]. Such platforms could enable global collaboration and reproducibility while accelerating innovation across all areas, from biomechanics-driven control schemes to human-centric interaction technologies.

In conclusion, the long-term trajectory of expressive humanoid robotics lies in interdisciplinary advancements that combine computational efficiency, biomechanical fidelity, adaptive learning, and human-centric design principles. By addressing these multifaceted challenges, researchers can unlock the full potential of humanoid robots to redefine both motion control and their role within broader human-robot interaction paradigms.

10 Conclusion

This survey has critically examined the landscape of expressive whole-body humanoid control by analyzing the integration of Model Predictive Control (MPC) and Reinforcement Learning (RL). Through an extensive review of state-of-the-art approaches, we synthesized key insights into the complementary strengths of these paradigms, the progress they enable, and the gaps that remain in achieving dynamic, adaptive, and expressive humanoid motions for real-world applications.

Fundamentally, MPC's deterministic optimization methods provide precise control by leveraging model-based predictive capabilities, ensuring stability under constraints such as balance, contact dynamics, and actuator limits. For instance, hierarchical inverse dynamics and momentum-based control strategies demonstrated robust performance in balancing and tracking tasks on torque-controlled humanoid robots [50], [51]. Similarly, nonlinear MPC frameworks have showcased effectiveness in whole-body motion planning by incorporating advanced dynamics, such as centroidal momentum and contact force cost functions, which rapidly adapt to dynamic environments [4], [120]. However, the computational complexity of MPC systems remains a bottleneck, particularly when addressing high degrees of freedom (DoFs) in humanoids or multi-contact scenarios, necessitating advancements in real-time solver efficiency [7].

Conversely, RL thrives in settings requiring adaptability and generalization by leveraging data-driven policies that learn directly from interactions. Recent successes in training expressive humanoid controllers through large-scale human motion datasets have enabled high-fidelity whole-body behaviors, such as dancing, gesturing, and rough-terrain locomotion [3], [11]. RL provides the flexibility to optimize behaviors for complex dynamics and unstructured tasks, as evidenced in frameworks that enable robust sim-to-real transfer during dynamic humanoid applications [100]. Nonetheless, RL faces persisting limitations regarding sample inefficiency and difficulty meeting safety or constraint requirements during training and deployment [8].

The convergence of MPC and RL, as outlined in hybrid approaches, has begun to address the individual shortcomings of these techniques. Hybrid control systems harness MPC for physical constraints and real-time safety guarantees while using RL for adaptability and task-level decision-making [13]. These systems have demonstrated efficacy in balancing long-term planning with short-horizon precision, as seen in applications such as adaptive footstep planning and responsive push recovery [18], [35]. However, achieving real-world deployment remains challenging due to real-time integration complexities, computational burdens, and transferring hybrid policies from simulation to physical systems reliably.

Looking ahead, critical advances are required to fully realize the promise of expressive humanoid robotics. First, future research must address the scalability of unified MPC-RL frameworks for higher-dimensional humanoids while mitigating their computational overhead. Continued development in real-time solvers, perhaps through GPU acceleration or tailored decomposition techniques, is essential to enable decision-making under tight constraints [14], [25]. Standardized metrics and benchmarking protocols will further be indispensable for evaluating hybrid systems' stability, expressiveness, and transfer capabilities across diverse tasks and hardware [12].

Finally, interdisciplinary collaboration between control theorists, machine learning researchers, and roboticists will play an increasingly pivotal role. Bridging biomechanics insights with control strategies may unlock breakthroughs in both natural human-like behaviors and efficient robotic designs [83], [121]. By fostering synergies across these disciplines, we can push the boundaries of adaptive, socially cognizant humanoid robotics, enabling applications in healthcare, entertainment, and beyond. This synthesis reflects a field standing at the confluence of remarkable technical capability and expansive opportunities yet to be explored.

REFERENCES

[1] S. Caron, A. Kheddar, and O. Tempier, "Stair climbing stabilization of the hrp-4 humanoid robot using whole-body admittance control," *2019 International Conference on Robotics and Automation (ICRA)*, pp. 277–283, 2018. 1, 2, 22

[2] Z. Li, C. Cummings, and K. Sreenath, "Animated cassie: A dynamic relatable robotic character," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3739–3746, 2020. 1

[3] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang, "Expressive whole-body control for humanoid robots," *ArXiv*, vol. abs/2402.16796, 2024. 1, 4, 7, 13, 26, 27, 33

[4] M. Neunert, M. Stäuble, M. Giftthaler, C. D. Bellicoso, J. Carius, C. Gehring, M. Hutter, and J. Buchli, "Whole-body nonlinear model predictive control through contacts for quadrupeds," *IEEE Robotics and Automation Letters*, vol. 3, pp. 1458–1465, 2017. 1, 2, 6, 9, 17, 25, 33

[5] E. Daneshmand, M. Khadiv, F. Grimminger, and L. Righetti, "Variable horizon mpc with swing foot dynamics for bipedal walking control," *IEEE Robotics and Automation Letters*, vol. 6, pp. 2349–2356, 2021. 1, 6, 11, 12, 13, 21, 28, 29, 31

[6] H. Li and P. M. Wensing, "Cafe-mpc: A cascaded-fidelity model predictive control framework with tuning-free whole-body control," *ArXiv*, vol. abs/2403.03995, 2024. 1, 4, 7, 9, 11, 18, 20, 31

[7] C. Khazoom, S. Hong, M. Chignoli, E. Stanger-Jones, and S. Kim, "Tailoring solution accuracy for fast whole-body model predictive control of legged robots," *IEEE Robotics and Automation Letters*, vol. 9, pp. 11 074–11 081, 2024. 1, 2, 5, 8, 12, 16, 17, 18, 21, 28, 33

[8] Y. Yuan and K. Kitani, "Residual force control for agile human behavior imitation and extended motion synthesis," *ArXiv*, vol. abs/2006.07364, 2020. 1, 7, 12, 14, 17, 19, 26, 27, 29, 30, 33

[9] Z. Zhuang, S. Yao, and H. Zhao, "Humanoid parkour learning," *ArXiv*, vol. abs/2406.10759, 2024. 1, 4, 19, 27

[10] C. Yang, K. Yuan, S. Heng, T. Komura, and Z. Li, "Learning natural locomotion behaviors for humanoid robots using human bias," *IEEE Robotics and Automation Letters*, vol. 5, pp. 2610–2617, 2020. 2, 5, 7, 13, 15, 16, 21, 22, 24, 25, 31, 32

[11] M. Taylor, S. Bashkirov, J. Rico, I. Toriyama, N. Miyada, H. Yanagisawa, and K. Ishizuka, "Learning bipedal robot locomotion from human movement," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2797–2803, 2021. 2, 7, 13, 22, 26, 28, 33

[12] B. J. van Marum, A. Shrestha, H. Duan, P. Dugar, J. Dao, and A. Fern, "Revisiting reward design and evaluation for robust humanoid standing and walking," *ArXiv*, vol. abs/2404.19173, 2024. 2, 16, 22, 28, 29, 30, 33

[13] D. oh Kang, J. Cheng, M. Zamora, F. Zargarbashi, and S. Coros, "Rl + model-based control: Using on-demand optimal control to learn versatile legged locomotion," *IEEE Robotics and Automation Letters*, vol. 8, pp. 6619–6626, 2023. 2, 4, 5, 7, 10, 13, 17, 19, 20, 22, 24, 26, 33

[14] M. Minniti, R. Grandia, F. Farshidian, and M. Hutter, "Adaptive clf-mpc with application to quadrupedal robots," *IEEE Robotics and Automation Letters*, vol. 7, pp. 565–572, 2021. 2, 3, 4, 25, 26, 33

[15] Y. Liu, C. Chen, and L. Yi, "Interactive humanoid: Online full-body motion reaction synthesis with social affordance canonicalization and forecasting," *ArXiv*, vol. abs/2312.08983, 2023. 2, 18

[16] N. Scianca, D. D. Simone, L. Lanari, and G. Oriolo, "Mpc for humanoid gait generation: Stability and feasibility," *IEEE Transactions on Robotics*, vol. 36, pp. 1171–1188, 2019. 2, 5, 12, 18, 22, 23, 32

[17] G. Romualdi, S. Dafarra, G. L'Erario, I. Sorrentino, S. Traversaro, and D. Pucci, "Online non-linear centroidal mpc for humanoid robot locomotion with step adjustment," *2022 International Conference on Robotics and Automation (ICRA)*, pp. 10 412–10 419, 2022. 2, 18

[18] M. Shafiee-Ashtiani, A. Yousefi-Koma, M. Panahi, and M. Khadiv, "Push recovery of a humanoid robot based on model predictive control and capture point," *2016 4th International Conference on Robotics and Mechatronics (ICROM)*, pp. 433–438, 2016. 2, 12, 17, 20, 21, 22, 24, 25, 28, 29, 31, 33

[19] S. Dafarra, G. Romualdi, and D. Pucci, "Dynamic complementarity conditions and whole-body trajectory optimization for humanoid robot locomotion," *IEEE Transactions on Robotics*, vol. 38, pp. 3414–3433, 2022. 2, 5, 6, 7, 10, 11, 12, 17, 19, 22, 25, 30

[20] C. Mastalli, W. Merkt, G. Xin, J.-E. Shim, M. Mistry, I. Havoutis, and S. Vijayakumar, "Agile maneuvers in legged robots: a predictive control approach," *ArXiv*, vol. abs/2203.07554, 2022. 2

[21] A. Herzog, S. Schaal, and L. Righetti, "Structured contact force optimization for kino-dynamic motion generation," *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2703–2710, 2016. 2, 5, 8, 13, 29, 32

[22] J. Li and Q. Nguyen, "Multi-contact mpc for dynamic loco-manipulation on humanoid robots," *2023 American Control Conference (ACC)*, pp. 1215–1220, 2022. 2, 9, 13, 21, 23, 25, 28, 32

[23] R. Grandia, F. Farshidian, A. Dosovitskiy, R. Ranftl, and M. Hutter, "Frequency-aware model predictive control," *IEEE Robotics and Automation Letters*, vol. 4, pp. 1517–1524, 2018. 2, 9, 25

[24] H. Li, R. J. Frei, and P. M. Wensing, "Model hierarchy predictive control of robotic systems," *IEEE Robotics and Automation Letters*, vol. 6, pp. 3373–3380, 2020. 2, 12, 29, 32

[25] M. Zafar, S. Hutchinson, and E. A. Theodorou, "Hierarchical optimization for whole-body control of wheeled inverted pendulum humanoids," *2019 International Conference on Robotics and Automation (ICRA)*, pp. 7535–7542, 2018. 2, 5, 22, 33

[26] M. H. Yeganegi, M. Khadiv, A. Prete, S. Moosavian, and L. Righetti, "Robust walking based on mpc with viability guarantees," *IEEE Transactions on Robotics*, vol. 38, pp. 2389–2404, 2020. 2, 5, 9, 12, 16, 18, 21, 22, 25, 29, 32

[27] N. Karnchanachari, M. I. Valls, D. Hoeller, and M. Hutter, "Practical reinforcement learning for mpc: Learning from sparse objectives in under an hour on a real robot," in *Conference on Learning for Dynamics & Control*, 2020, pp. 211–224. 3, 6, 8, 13, 21, 23, 26

[28] N. Hansen, V. JyothirS, V. Sobal, Y. LeCun, X. Wang, and H. Su, "Hierarchical world models as visual whole-body humanoid controllers," *ArXiv*, vol. abs/2405.18418, 2024. 3, 8, 15, 17, 20, 26, 27, 29, 33

[29] S. Bang, C. A. Jov'e, and L. Sentis, "Rl-augmented mpc framework for agile and robust bipedal footstep locomotion planning and control," *ArXiv*, vol. abs/2407.17683, 2024. 3, 8, 10, 12, 13, 16, 20, 23, 25, 28, 29, 31, 32

[30] K. Ishihara, H. Gomi, and J. Morimoto, "Hierarchical learning framework for whole-body model predictive control of a real humanoid robot," *ArXiv*, vol. abs/2409.08488, 2024. 3, 4, 6, 8, 11, 18, 20, 23, 24, 28

[31] A. Meduri, P. Shah, J. Viereck, M. Khadiv, I. Havoutis, and L. Righetti, "Biconmp: A nonlinear model predictive control framework for whole body motion planning," *IEEE Transactions on Robotics*, vol. 39, pp. 905–922, 2022. 3, 4, 6, 28, 29

[32] V. Zinage, A. Khalil, and E. Bakolas, "Transformermpc: Accelerating model predictive control via transformers," *ArXiv*, vol. abs/2409.09266, 2024. 3, 6, 14, 18

[33] Y. Kim, H. Oh, J. Lee, J. Choi, G. Ji, M. Jung, D. Youm, and J. Hwangbo, "Not only rewards but also constraints: Applications on legged robot locomotion," *IEEE Transactions on Robotics*, vol. 40, pp. 2984–3003, 2023. 4, 19, 30

[34] M. H. Yeganegi, M. Khadiv, S. Moosavian, J.-J. Zhu, A. Prete, and L. Righetti, "Robust humanoid locomotion using trajectory optimization and sample-efficient learning*," *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, pp. 170–177, 2019. 4, 7, 16, 19, 24, 31

[35] Y. Ding, C. Khazoom, M. Chignoli, and S. Kim, "Orientation-aware model predictive control with footstep adaptation for dynamic humanoid walking," *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, pp. 299–305, 2022. 4, 7, 10, 18, 19, 23, 24, 29, 33

[36] S. Dafarra, F. Romano, and F. Nori, "A receding horizon push recovery strategy for balancing the icub humanoid robot," in *International Conference on Robotics in Alpe-Adria-Danube Region*, 2017, pp. 297–305. 4, 7, 10, 27

[37] J. Li, J. Ma, O. Kolt, M. Shah, and Q. Nguyen, "Dynamic loco-manipulation on hector: Humanoid for enhanced control and open-source research," *ArXiv*, vol. abs/2312.11868, 2023. 4, 27, 32

[38] J. Dao, H. Duan, and A. Fern, "Sim-to-real learning for humanoid box loco-manipulation," *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 16 930–16 936, 2023. 4, 24

[39] X. Gu, Y.-J. Wang, and J. Chen, "Humanoid-gym: Reinforcement learning for humanoid robot with zero-shot sim2real transfer," *ArXiv*, vol. abs/2404.05695, 2024. 4, 14, 19, 24, 27, 30

[40] T. Haarnoja, B. Moran, G. Lever, S. H. Huang, D. Tirumala, M. Wulfmeier, J. Humplik, S. Tunyasuvunakool, N. Siegel, R. Hafner, M. Bloesch, K. Hartikainen, A. Byravan, L. Hasenclever, Y. Tassa, F. Sadeghi, N. Batchelor, F. Casarini, S. Saliceti, C. Game, N. Sreendra, K. Patel, M. Gwira, A. Huber, N. Hurley, F. Nori, R. Hadsell, and N. Heess, "Learning agile soccer skills for a bipedal robot with deep reinforcement learning," *Science Robotics*, vol. 9, 2023. 4, 7, 30

[41] S. Mason, N. Rotella, S. Schaal, and L. Righetti, "Balancing and walking using full dynamics lqr control with contact constraints," *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pp. 63–68, 2016. 4

[42] G. A. Castillo, B. Weng, A. Hereid, and W. Zhang, "Reinforcement learning meets hybrid zero dynamics: A case study for rabbit," *2019 International Conference on Robotics and Automation (ICRA)*, pp. 284–290, 2018. 4, 15

[43] E. Arcari, M. Minniti, A. Scampicchio, A. Carron, F. Farshidian, M. Hutter, and M. Zeilinger, "Bayesian multi-task learning mpc for robotic mobile manipulation," *IEEE Robotics and Automation Letters*, vol. 8, pp. 3222–3229, 2022. 4, 13, 18, 23, 28, 29

[44] M. Okada and T. Taniguchi, "Variational inference mpc for bayesian model-based reinforcement learning," in *Conference on Robot Learning*, 2019, pp. 258–272. 4, 15, 20, 21

[45] S. L. Cleac'h, T. A. Howell, S. Yang, C.-Y. Lee, J. Zhang, A. L. Bishop, M. Schwager, and Z. Manchester, "Fast contact-implicit model predictive control," *IEEE Transactions on Robotics*, vol. 40, pp. 1617–1629, 2021. 4, 6, 12

[46] P. Dugar, A. Shrestha, F. Yu, B. J. van Marum, and A. Fern, "Learning multi-modal whole-body control for real-world humanoid robots," *ArXiv*, vol. abs/2408.07295, 2024. 5, 7, 15, 16, 20, 28, 30, 31, 33

[47] M. Minniti, R. Grandia, K. Fäh, F. Farshidian, and M. Hutter, "Model predictive robot-environment interaction control for mobile manipulation tasks," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1651–1657, 2021. 5, 15

[48] M. Zanon and S. Gros, "Safe reinforcement learning using robust mpc," *IEEE Transactions on Automatic Control*, vol. 66, pp. 3638–3652, 2019. 5, 8, 20, 25, 31

[49] J. Nubert, J. Köhler, V. Berenz, F. Allgöwer, and S. Trimpe, "Safe and fast tracking control on a robot manipulator: Robust mpc and neural network control," *ArXiv*, vol. abs/1912.10360, 2019. 5, 31

[50] A. Herzog, N. Rotella, S. Mason, F. Grimminger, S. Schaal, and L. Righetti, "Momentum control with hierarchical inverse dynamics on a torque-controlled humanoid," *Autonomous Robots*, vol. 40, pp. 473 – 491, 2014. 5, 11, 13, 21, 25, 31, 33

[51] A. Herzog, L. Righetti, F. Grimminger, P. Pastor, and S. Schaal, "Balancing experiments on a torque-controlled humanoid with hierarchical inverse dynamics," *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 981–988, 2013. 5, 8, 17, 33

[52] D. Kim, J. Lee, O. Campbell, H. Hwang, and L. Sentis, "Computationally-robust and efficient prioritized whole-body controller with contact constraints," *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–8, 2018. 5, 11, 17

[53] D. Kim, S. J. Jorgensen, J. Lee, J. Ahn, J. Luo, and L. Sentis, "Dynamic locomotion for passive-ankle biped robots and humanoids using whole-body locomotion control," *The International Journal of Robotics Research*, vol. 39, pp. 936 – 956, 2019. 5, 29

[54] J. Merel, L. Hasenclever, A. Galashov, A. Ahuja, V. Pham, G. Wayne, Y. Teh, and N. Heess, "Neural probabilistic motor primitives for humanoid control," *ArXiv*, vol. abs/1811.11711, 2018. 5, 7, 13, 22, 24, 27

[55] Z. Luo, J. Cao, J. Merel, A. Winkler, J. Huang, K. Kitani, and W. Xu, "Universal humanoid motion representations for physics-based control," *ArXiv*, vol. abs/2310.04582, 2023. 5, 12, 14, 16, 17, 30

[56] C. Yang, K. Yuan, W. Merkt, T. Komura, S. Vijayakumar, and Z. Li, "Learning whole-body motor skills for humanoids," *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pp. 270–276, 2018. 6, 10, 12, 17, 24, 26, 27

[57] M. Menner, P. Worsnop, and M. Zeilinger, "Predictive modeling by infinite-horizon constrained inverse optimal control with application to a human manipulation task," *ArXiv*, vol. abs/1812.11600, 2018. 6

[58] V. Kurtz, A. Castro, A. Özgün Önol, and H. Lin, "Inverse dynamics trajectory optimization for contact-implicit model predictive control," *ArXiv*, vol. abs/2309.01813, 2023. 6, 14

[59] N. J. Kong, C. Li, G. Council, and A. M. Johnson, "Hybrid ilqr model predictive control for contact implicit stabilization on legged robots," *IEEE Transactions on Robotics*, vol. 39, pp. 4712–4727, 2022. 6, 9, 11, 18, 20, 23

[60] B. Pontón, A. Herzog, S. Schaal, and L. Righetti, "On time optimization of centroidal momentum dynamics," *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–7, 2017. 6, 21, 26, 33

[61] S. Han, M. Iskandar, J. Lee, and M. J. Kim, "Online multi-contact feedback model predictive control for interactive robotic tasks," *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11 556–11 562, 2024. 6

[62] F. Hogan, E. R. Grau, and A. Rodriguez, "Reactive planar manipulation with convex hybrid mpc," *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 247–253, 2017. 6

[63] F. M. Smaldone, N. Scianca, V. Modugno, L. Lanari, and G. Oriolo, "Gait generation using intrinsically stable mpc in the presence of persistent disturbances," *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, pp. 651–656, 2019. 6

[64] C. Yang, T. Komura, and Z. Li, "Emergence of human-comparable balancing behaviours by deep reinforcement learning," *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, pp. 372–377, 2017. 7

[65] D. Kim, J. Lee, and L. Sentis, "Robust dynamic locomotion via reinforcement learning and novel whole body controller," *ArXiv*, vol. abs/1708.02205, 2017. 7, 10

[66] J. Eßer, S. Kumar, H. Peters, V. Bargsten, J. Gea, C. Mastalli, O. Stasse, and F. Kirchner, "Design, analysis and control of the series-parallel hybrid rh5 humanoid robot," *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*, pp. 400–407, 2021. 7, 31

[67] S. Kuindersma, F. Permenter, and R. Tedrake, "An efficiently solvable quadratic program for stabilizing dynamic locomotion," *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2589–2594, 2013. 7, 31

[68] J.-P. Sleiman, F. Farshidian, M. Minniti, and M. Hutter, "A unified mpc framework for whole-body dynamic locomotion and manipulation," *IEEE Robotics and Automation Letters*, vol. 6, pp. 4688–4695, 2021. 7

[69] Q. Liao, B. Zhang, X. Huang, X. Huang, Z. Li, and K. Sreenath, "Berkeley humanoid: A research platform for learning-based control," *ArXiv*, vol. abs/2407.21781, 2024. 8, 16, 18, 32

[70] C. Mastalli, S. Chhatoi, T. Corbéres, S. Tonneau, and S. Vijayakumar, "Inverse-dynamics mpc via nullspace resolution," *IEEE Transactions on Robotics*, vol. 39, pp. 3222–3241, 2022. 8, 11

[71] J. Matschek, J. Bethge, and R. Findeisen, "Safe machine-learning-supported model predictive force and motion control in robotics," *IEEE Transactions on Control Systems Technology*, vol. 31, pp. 2380–2392, 2023. 8, 15, 20, 31

[72] C. Xie, S. Patil, T. Moldovan, S. Levine, and P. Abbeel, "Model-based reinforcement learning with parametrized physical models and optimism-driven exploration," *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 504–511, 2015. 9, 18

[73] B. Pontón, A. Herzog, S. Schaal, and L. Righetti, "A convex model of humanoid momentum dynamics for multi-contact motion generation," *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pp. 842–849, 2016. 9, 31

[74] S. Katayama and T. Ohtsuka, "Whole-body model predictive control with rigid contacts via online switching time optimization," *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8858–8865, 2022. 9, 11

[75] Y. Wang, N. Dehio, A. Tanguy, and A. Kheddar, "Impact-aware task-space quadratic-programming control," *The International Journal of Robotics Research*, vol. 42, pp. 1265 – 1282, 2020. 9, 26

[76] S. Caron and A. Kheddar, "Multi-contact walking pattern generation based on model preview control of 3d com accelerations," *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pp. 550–557, 2016. 10, 26

[77] G. Kim, D.-K. Kang, J. ha Kim, S. Hong, and H.-W. Park, "Contact-implicit model predictive control: Controlling diverse quadruped motions without pre-planned contact modes or trajectories," *The International Journal of Robotics Research*, 2023. 10, 18

[78] G. A. Castillo, B. Weng, S. Yang, W. Zhang, and A. Hereid, "Template model inspired task space learning for robust bipedal locomotion," *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8582–8589, 2023. 10, 14, 19, 24

[79] C. Zhang, W. Xiao, T. He, and G. Shi, "Wococo: Learning whole-body humanoid control with sequential contacts," *ArXiv*, vol. abs/2406.06005, 2024. 10, 14, 19, 27, 30

[80] C. Khazoom, D. G. Diaz, Y. Ding, and S. Kim, "Humanoid self-collision avoidance using whole-body control with control barrier functions," *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, pp. 558–565, 2022. 12, 17, 22

[81] M. Zambelli, A. Cully, and Y. Demiris, "Multimodal representation models for prediction and control from partial information," *Robotics Auton. Syst.*, vol. 123, 2019. 12, 16

[82] S. Jeon, S. Heim, C. Khazoom, and S. Kim, "Benchmarking potential based rewards for learning humanoid locomotion," *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9204–9210, 2023. 12, 27, 29

[83] A. Sripathy, A. Bobu, Z. Li, K. Sreenath, D. S. Brown, and A. Dragan, "Teaching robots to span the space of functional expressive motion," *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 13 406–13 413, 2022. 13, 22, 33

[84] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Reinforcement learning for versatile, dynamic, and robust bipedal locomotion control," *ArXiv*, vol. abs/2401.16889, 2024. 14

[85] R. P. Singh, M. Benallegue, M. Morisawa, R. Cisneros, and F. Kanehiro, "Learning bipedal walking on planned footsteps for humanoid robots," *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, pp. 686–693, 2022. 14

[86] R. P. Singh, Z. Xie, P. Gergondet, and F. Kanehiro, "Learning bipedal walking for humanoids with current feedback," *IEEE Access*, vol. 11, pp. 82 013–82 023, 2023. 14

[87] R. Batke, F.-Y. Yu, J. Dao, J. Hurst, R. Hatton, A. Fern, and K. R. Green, "Optimizing bipedal maneuvers of single rigid-body models for reinforcement learning," *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, pp. 714–721, 2022. 14, 19

[88] F. Yu, R. Batke, J. Dao, J. Hurst, K. R. Green, and A. Fern, "Dynamic bipedal maneuvers through sim-to-real reinforcement learning," *ArXiv*, vol. abs/2207.07835, 2022. 14

[89] L. M. Smith, Y. Cao, and S. Levine, "Grow your limits: Continuous improvement with real-world rl for robotic locomotion," *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 10 829–10 836, 2023. 14

[90] Z. Wang, W. Wei, R. Yu, J. Wu, and Q. Zhu, "Toward understanding key estimation in learning robust humanoid locomotion," *ArXiv*, vol. abs/2403.05868, 2024. 15, 17

[91] J. Merel, A. Ahuja, V. Pham, S. Tunyasuvunakool, S. Liu, D. Tirumala, N. Heess, and G. Wayne, "Hierarchical visuomotor control of humanoids," *ArXiv*, vol. abs/1811.09656, 2018. 16, 22, 32

[92] M. Djeha, P. Gergondet, and A. Kheddar, "Robust task-space quadratic programming for kinematic-controlled robots," *IEEE Transactions on Robotics*, vol. 39, pp. 3857–3874, 2023. 16

[93] A. Rai, G. Sutanto, S. Schaal, and F. Meier, "Learning feedback terms for reactive planning and control," *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2184–2191, 2016. 16, 22, 31

[94] B. Pontón, M. Khadiv, A. Meduri, and L. Righetti, "Efficient multicontact pattern generation with sequential convex approximations of the centroidal dynamics," *IEEE Transactions on Robotics*, vol. 37, pp. 1661–1679, 2020. 16, 21, 32

[95] K. Darvish, Y. Tirupachuri, G. Romualdi, L. Rapetti, D. Ferigo, F. J. A. Chavez, and D. Pucci, "Whole-body geometric retargeting for humanoid robots," *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, pp. 679–686, 2019. 17, 26

[96] J. Foster, S. McCrory, C. DeBuys, S. Bertrand, and R. J. Griffin, "Physically consistent online inertial adaptation for humanoid loco-manipulation," *ArXiv*, vol. abs/2405.07901, 2024. 17, 25, 31, 32

[97] S. Bang, J. Lee, C. Gonzalez, and L. Sentis, "Variable inertia model predictive control for fast bipedal maneuvers," *ArXiv*, vol. abs/2407.16811, 2024. 17, 22, 28, 29, 31, 32

[98] A. Sathya, H. Bruyninckx, W. Decré, and G. Pipeleers, "Efficient constrained dynamics algorithms based on an equivalent lqr formulation using gauss' principle of least constraint," *IEEE Transactions on Robotics*, vol. 40, pp. 729–749, 2023. 17

[99] Y. Gong and J. Grizzle, "Zero dynamics, pendulum models, and angular momentum in feedback control of bipedal locomotion," *ArXiv*, vol. abs/2105.08170, 2021. 17

[100] X. Gu, Y.-J. Wang, X. Zhu, C. Shi, Y. Guo, Y. Liu, and J. Chen, "Advancing humanoid locomotion: Mastering challenging terrains with denoising world model learning," *ArXiv*, vol. abs/2408.14472, 2024. 17, 33

[101] J. Zeng, B. Zhang, and K. Sreenath, "Safety-critical model predictive control with discrete-time control barrier function," *2021 American Control Conference (ACC)*, pp. 3882–3889, 2020. 18

[102] W. Jallet, E. Dantec, E. Arlaud, J. Carpentier, and N. Mansard, "Parallel and proximal constrained linear-quadratic methods for real-time nonlinear mpc," *ArXiv*, vol. abs/2405.09197, 2024. 18, 26, 29

[103] Z. Li, X. Cheng, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Reinforcement learning for robust parameterized locomotion control of bipedal robots," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2811–2817, 2021. 19, 27

[104] H. Duan, J. Dao, K. R. Green, T. Apgar, A. Fern, and J. Hurst, "Learning task space actions for bipedal locomotion," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1276–1282, 2020. 19, 24

[105] D. Bertsekas, "Model predictive control and reinforcement learning: A unified framework based on dynamic programming," *ArXiv*, vol. abs/2406.00592, 2024. 20

[106] A. Alavilli, K. Nguyen, S. Schoedel, B. Plancher, and Z. Manchester, "Tinympc: Model-predictive control on resource-constrained microcontrollers," *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–7, 2023. 20, 21, 25, 26, 29, 31

[107] J. Jankowski, L. Brudermüller, N. Hawes, and S. Calinon, "Vp-sto: Via-point-based stochastic trajectory optimization for reactive robot behavior," *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 10 125–10 131, 2022. 20, 23

[108] S. Singh, R. Russell, and P. M. Wensing, "Efficient analytical derivatives of rigid-body dynamics using spatial vector algebra," *IEEE Robotics and Automation Letters*, vol. 7, pp. 1776–1783, 2021. 21, 33

[109] G. Romualdi, S. Dafarra, Y. Hu, and D. Pucci, "A benchmarking of dcm based architectures for position and velocity controlled walking of humanoid robots," *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pp. 1–9, 2018. 21, 27

[110] J.-S. Ahn, J. Jung, Y. Lee, H. Lee, S. Haddadin, and J. Park, "Efficient computation of whole-body control utilizing simplified whole-body dynamics via centroidal dynamics," *ArXiv*, vol. abs/2409.10903, 2024. 22, 33

[111] G. Romualdi, S. Dafarra, Y. Hu, P. Ramadoss, F. J. A. Chavez, S. Traversaro, and D. Pucci, "A benchmarking of dcm based architectures for position, velocity and torque controlled humanoid robots," *Int. J. Humanoid Robotics*, vol. 17, pp. 1 950 034:1–1 950 034:25, 2019. 22

[112] A. Jain, L. Chan, D. S. Brown, and A. Dragan, "Optimal cost design for model predictive control," *ArXiv*, vol. abs/2104.11353, 2021. 23

[113] S. Christen, S. Stevšić, and O. Hilliges, "Demonstration-guided deep reinforcement learning of control policies for dexterous human-robot interaction," *2019 International Conference on Robotics and Automation (ICRA)*, pp. 2161–2167, 2019. 24

[114] S. Mason, N. Rotella, S. Schaal, and L. Righetti, "An mpc walking framework with external contact forces," *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1785–1790, 2017. 24

[115] Q. Zhang, P. Cui, D. Yan, J. Sun, Y. Duan, A. Zhang, and R. Xu, "Whole-body humanoid robot locomotion with human reference," *ArXiv*, vol. abs/2402.18294, 2024. 25

[116] T. He, Z. Luo, X. He, W. Xiao, C. Zhang, W. Zhang, K. Kitani, C. Liu, and G. Shi, "Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning," *ArXiv*, vol. abs/2406.08858, 2024. 26, 27

[117] M. Meser, A. Bhatt, B. Belousov, and J. Peters, "Mujoco mpc for humanoid control: Evaluation on humanoidbench," *ArXiv*, vol. abs/2408.00342, 2024. 26, 28, 30

[118] M. Giftthaler, M. Neunert, M. Stäuble, and J. Buchli, "The control toolbox — an open-source c++ library for robotics, optimal and model predictive control," *2018 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots (SIMPAR)*, pp. 123–129, 2018. 26

[119] T. A. Howell, N. Gileadi, S. Tunyasuvunakool, K. Zakka, T. Erez, and Y. Tassa, "Predictive sampling: Real-time behaviour synthesis with mujoco," *ArXiv*, vol. abs/2212.00541, 2022. 28, 32

[120] Y. Ding, A. Pandala, C. Li, Y.-H. Shin, and H. won Park, "Representation-free model predictive control for dynamic motions in quadrupeds," *IEEE Transactions on Robotics*, vol. 37, pp. 1154–1171, 2020. 28, 33

[121] T. He, Z. Luo, W. Xiao, C. Zhang, K. Kitani, C. Liu, and G. Shi, "Learning human-to-humanoid real-time whole-body teleoperation," *ArXiv*, vol. abs/2403.04436, 2024. 29, 30, 33

[122] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn, "Humanplus: Humanoid shadowing and imitation from humans," *ArXiv*, vol. abs/2406.10454, 2024. 30

[123] L. Bärmann, R. Kartmann, F. Peller-Konrad, A. H. Waibel, and T. Asfour, "Incremental learning of humanoid robot behavior from natural interaction and large language models," *Frontiers in Robotics and AI*, vol. 11, 2023. 30

[124] D. Baek, A. Purushottam, and J. Ramos, "Hybrid lmc: Hybrid learning and model-based control for wheeled humanoid robot via ensemble deep reinforcement learning," *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 9347–9354, 2022. 31

[125] D. Ferigo, R. Camoriano, P. M. Viceconte, D. Calandriello, S. Traversaro, L. Rosasco, and D. Pucci, "On the emergence of whole-body strategies from humanoid robot push-recovery learning," *IEEE Robotics and Automation Letters*, vol. 6, pp. 8561–8568, 2021. 32

[126] S. Faraji and A. Ijspeert, "3lp: A linear 3d-walking model including torso and swing dynamics," *The International Journal of Robotics Research*, vol. 36, pp. 436 – 455, 2016. 32

[127] D. Marew, N. Perera, S. Yu, S. Roelker, and D. Kim, "A biomechanics-inspired approach to soccer kicking for humanoid robots," *ArXiv*, vol. abs/2407.14612, 2024. 32

[128] Q. Rouxel, K. Yuan, R. Wen, and Z. Li, "Multicontact motion retargeting using whole-body optimization of full kinematics and sequential force equilibrium," *IEEE/ASME Transactions on Mechatronics*, vol. 27, pp. 4188–4198, 2022. 32, 33