Controlling the Robustness, Energy Consumption, and Speed of Terrain-Traversing Dynamic Bipedal Robots

Christian Hubicki[1], Professor Keith Buffinton[1], C. David Remy[2]

Department of Mechanical Engineering[1], Bucknell University, Lewisburg, PA

Swiss Federal Institute of Technology[2], Zurich, Switzerland

Abstract

The field of legged robotics has been long anticipated in the popular media to herald a revolution in both civilian and military life. From mechanical fire fighters barreling through burning apartments with minimal regard for self-preservation to nimble explorers bounding up Martian ridges who never complain about the cold, finding applications for bipedal machines requires little imagination. Despite their promised dexterity and overall popular appeal, in the early 21st century, bipedal robots are seldom sighted outside of university research labs or cutting-edge technology firms.

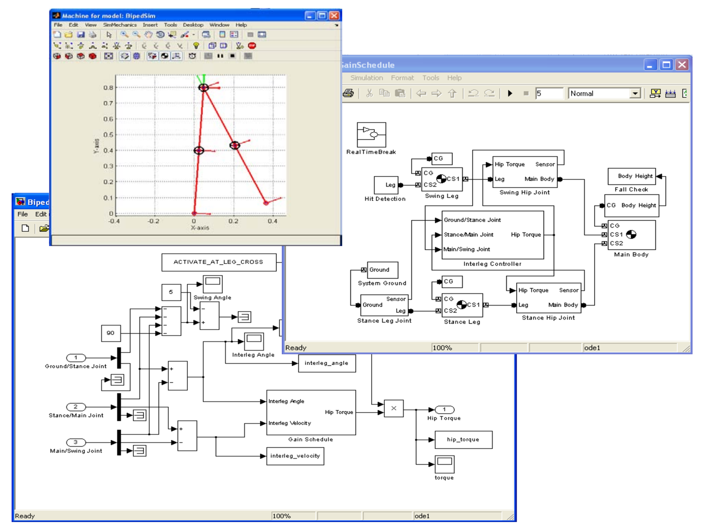
The absence of these legged machines in our daily lives can be attributed to significant technical barriers in performance. The largely untold flaw of Honda's flagship robotic humanoid, ASIMO, is that its exorbitant energy consumption drains its generously sized battery pack in roughly 30 minutes, nullifying its utility outside of relatively short public demonstrations. Recognizing that this energy limitation is not unique to ASIMO but common among current-generation walking robots, academic researchers have recently pushed to develop highly efficient bipeds. The consequence was a series of prototypes which trade an abundance of actuation and control authority for an under-actuated approach dubbed Dynamic Walking. Specifically, Cornell University developed two internationally publicized walking machines; one which boasted energy efficiency on par with human walking (for short distances) and the Cornell Ranger which set a world record for walking 5.6 miles on a single battery charge.

While delivering such significant advances in energy efficiency, dynamic walking robots have still largely fallen short in applications with high speed requirements or rough terrain. This investigation uses simulation to explore the inherent tradeoffs of controlling high-speed and highly robust walking robots while minimizing energy consumption. Using a novel controller which optimizes robustness, energy efficiency, and speed of a simulated robot on rough terrain, the user can adjust their priorities between these three outcome measures and systematically generate a characteristic performance curve. This curve represents the entire spectrum of performance for the given robot, revealing necessary energy costs for various demands of speed and robustness.

The novel robot controller is a two-tiered hierarchical system consisting of a heuristically-driven single-step controller and an overseeing Artificial Intelligence algorithm. The single-step controller is generated by tuning a set of control parameters in order to closely approximate an optimal performance curve produced by a series of genetic optimizations. This single step is calculated to produce the desired step speed for the least amount of energy expenditure. The Artificial Intelligence algorithm, more precisely described as a Value Iteration Reinforcement Learning Algorithm, is tasked to optimally plan for future steps. This step-planning algorithm decides which actions are taken by the single-step controller, thereby seeking conditions conducive to superior walking performance and avoiding unfavorable situations which the single-step controller lacks the foresight to evade.

Simulation

SimMechanics Model and Visualization



Terrain Modeling

Terrain is modeled as a stochastically generated variation in ground height. Using a Gaussian distribution, increasing the standard deviation raises the relative terrain roughness.

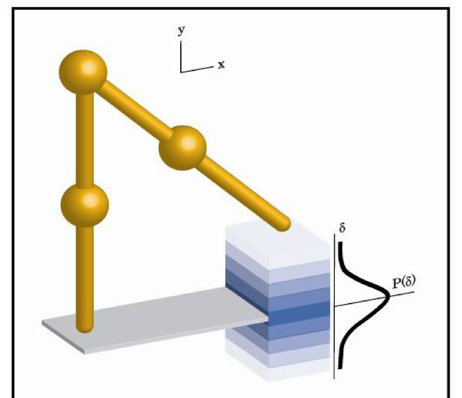
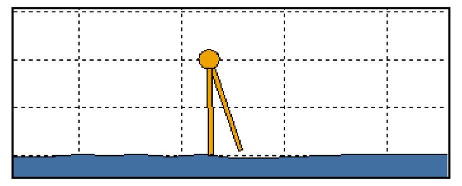
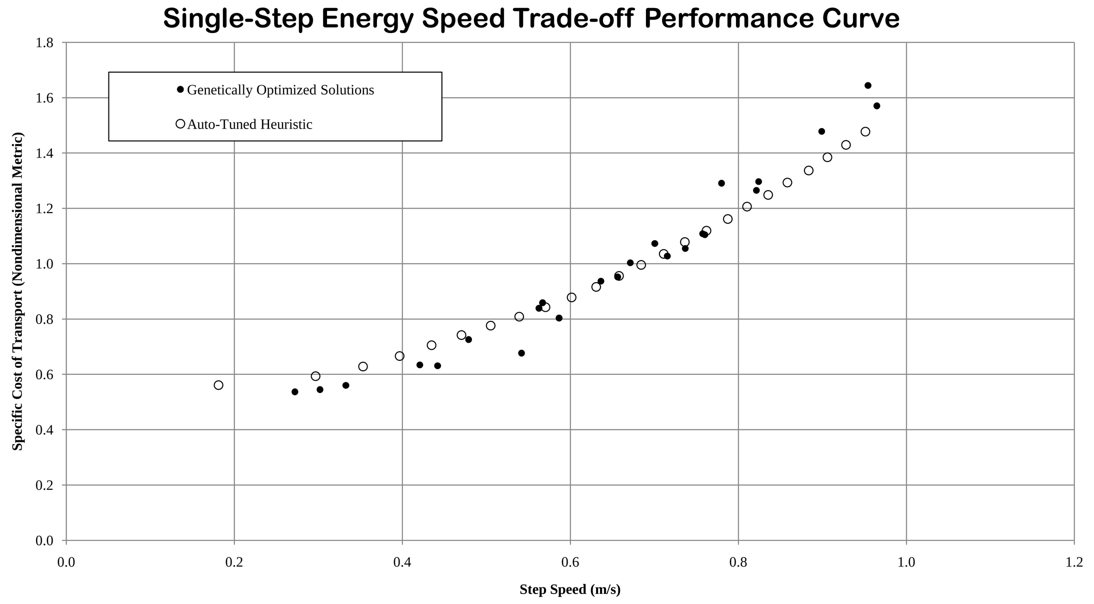


Figure 1: Visualization of compass gait model walking on stochastic terrain

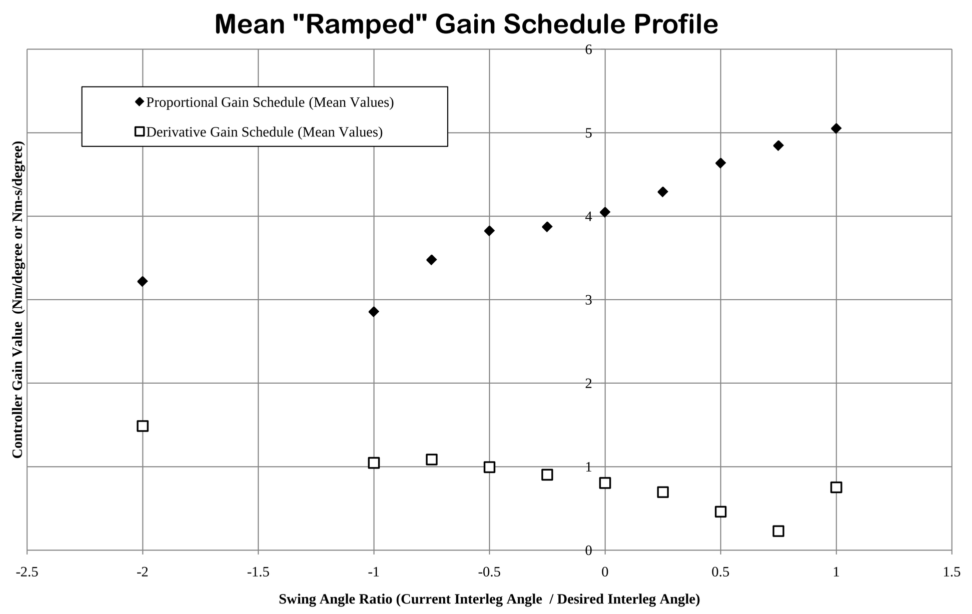


Simulation Results for Single-Step Controller



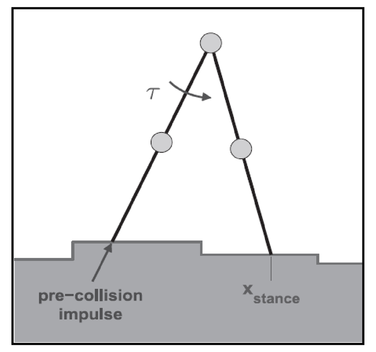
The results of a series of genetically optimized controllers were compared to a linear-scaling heuristic with its parameters tuned by a simple gradient-descent algorithm. The results were computed for a sparse sampling of the state space (an example pictured left) and the heuristic controller performance strongly predicted the genetically optimized performance (and vice versa).

The base gain schedule profile (pictured right) is the mean profile of the controllers resulting from 50+ genetic optimizations. This profile is linearly scaled to produce the heuristically driven controller and departs significantly from traditionally employed gain profiles.



Basic Actuation and Control

- **Pre-collision Impulse**
  - An instantaneous push-off of the stance leg just prior to swing leg collision
  - Efficient means of imparting energy to the system
- **Hip Torque**
  - A pure torque applied at the hip joint
  - A PD (Proportional-Derivative) Controller regulates the interleg angle



MIT Model and Testbed Robot (Byl, Tedrake 2009)

Gain Scheduled Control: Genetically Optimized vs. Heuristically Scaled

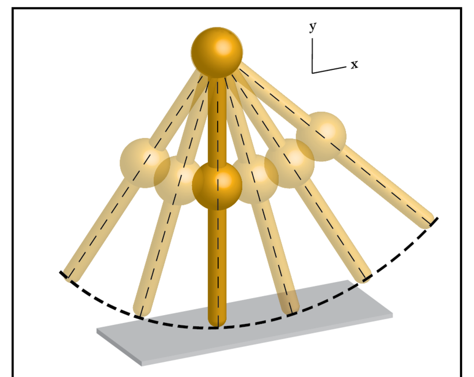
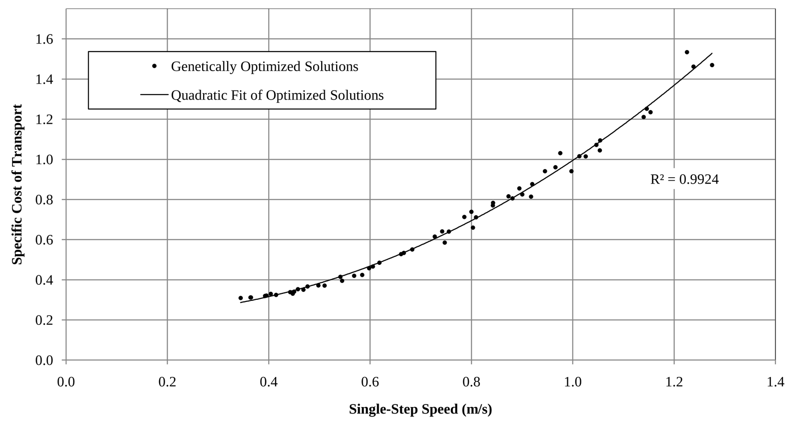


Figure 2: Gain-scheduled PD controller gains discretized by interleg angle

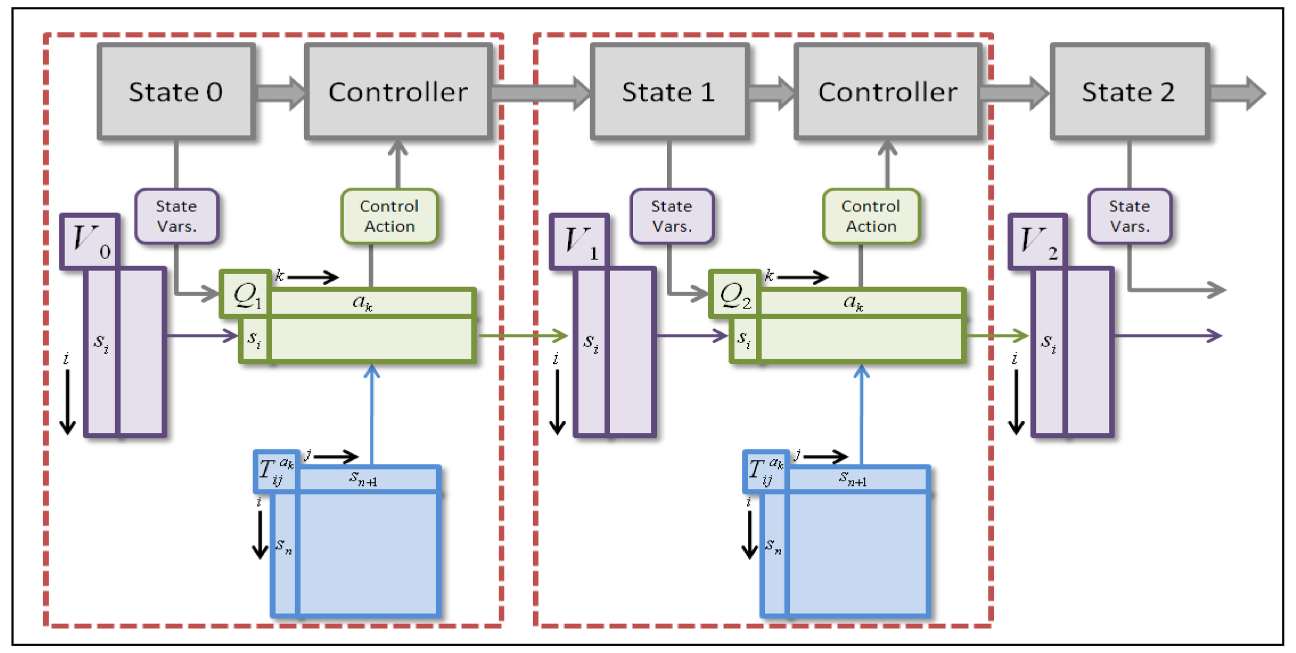
Various Genetically Optimized Solutions



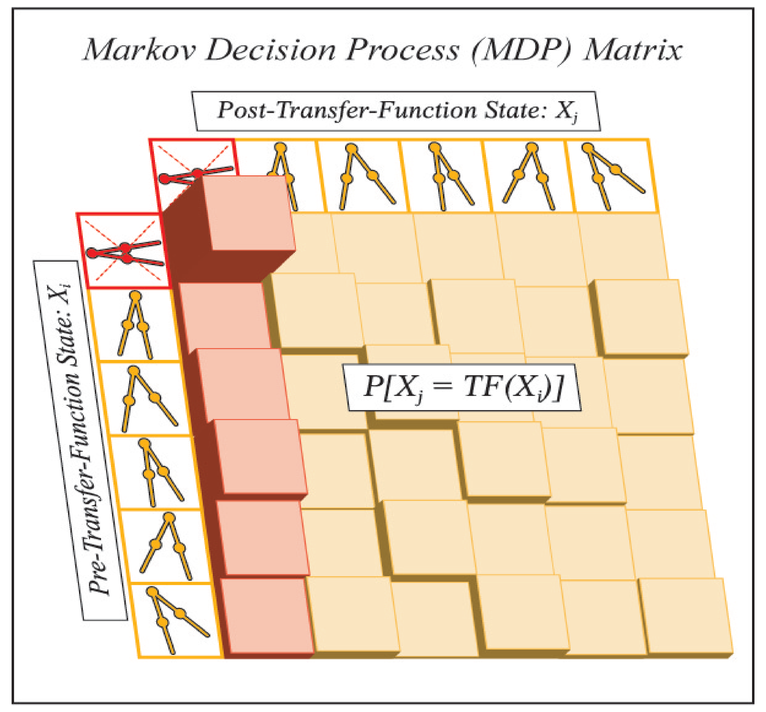
- **Gain-Scheduled Control**
  - Divides the swing motion into several sectors (ten)
  - Sector applies its own independent set of controller gains as the swing leg crosses each partition
  - Allows for any individual portion of the swing to "relax" or "tighten" control as necessary
- **Genetic Optimization**
  - Treats all gains for each sector (as well as the magnitude of the pre-collision impulse) as free parameters (21 total)
  - Each generation slightly "mutates" the free parameters and selects the best solutions of the new generation
  - The "best" solution is determined by a cost function which is set to prefer speed or energy efficiency
  - Relatively computationally "expensive" to run
- **Linear Scaling Heuristic**
  - Takes a base gain schedule (profile) and scales it based upon the demand for either speed or energy efficiency
  - Facilitates a simple means of producing tradeoffs
  - Requires the tuning of significantly fewer parameters (six)

Value Iteration Reinforcement Learning Algorithm and Subsequent Analysis

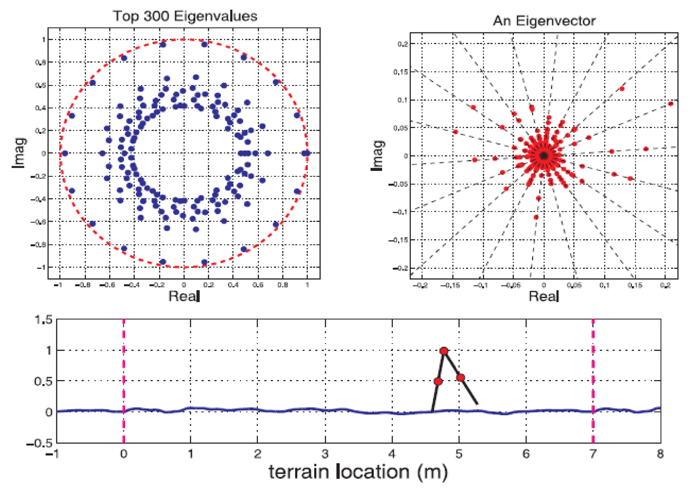
A subset of Artificial Intelligence, the *Value Iteration Reinforcement Learning Algorithm* operates by selecting actions likely to optimize future reward, seeking maximal *value*. In preparation for each step, an *Action-Value Function* (*Q*) is generated by consulting the current robot state, *State-Value Function* (*V*), and an estimation of likely future states (*Transition Matrix* T_{ij}). After executing the best action (*max(Q)*), the aforementioned functions are updated in preparation for the next decision, facilitating learning.



The *mean first-passage time* (MFPT) is a metric for stability in metastable stochastic systems. After the learning algorithm has converged upon an action policy, a *Markov Decision Process* (MDP) is generated which characterizes the probability of one robot state transitioning to another. Computing and ranking the eigenvalues of this matrix (λ_n), the second largest eigenvalue is used to calculate the MFPT which approximates the number of steps to failure.



$$M \approx \tau_2 = \frac{-1}{log(\lambda_2)} \approx \frac{1}{1-\lambda_2}$$



MIT Results published by Katie Byl and Russ Tedrake (2009) using the Value Iteration Reinforcement Learning Algorithm

Conclusion and Future Work

For the single-step case, a "trade-off conducive" controller was successfully devised. Informed by the results of a series of genetic optimizations, a comparatively simple heuristically-driven gain-scheduled controller was developed which approximates the performance of genetically optimized controllers, even when the heuristic parameters are tuned in an automated or blinded fashion. When compared to traditional profiles, the optimization inspired "ramp" profile has been shown to yield superior performance over a range of speed-energy demands.

Full implementation of this heuristic single-step controller requires the overseeing Value Iteration Algorithm. This method already has precedent in use with bipedal robots on rough terrain, however, only with a single metric and a relatively limited action space. Using the "trade-off conducive" heuristic, the learning algorithm has access to a highly diverse action space with which to achieve a variety of demands for robustness, energy efficiency, and speed on rough terrain. Ongoing work will investigate the efficacy of this mesh of methods, objectives, and application.

References

1. Byl, K., 2009, "Metastable Walking Machines," *Int. J. Robotics Res.*, **8**, 1040-1064.
2. Collins, S., 2002, "Efficient Bipedal Robots Based on Passive-Dynamic Walkers," *Science.*, **307**, 1082-1085.
3. McGeer, T., 1990, "Passive Dynamic Walking," *Int. J. Robotics Res.*, **9**, 62-82.

Acknowledgements

The authors gratefully acknowledge support for this work provided by the Office of Naval Research under Grant No. N000140810953.