B^3H: Black Box for Black Hole

— *Using AI to Bridge Spatial and Temporal Scale Separation*

Introduction

In the recently ushered in era of multimessenger astronomy, current and upcoming radio, optical, and X-ray surveys (e.g., VLASS, SKA; Vera Rubin; eROSITA) promise to unleash a flood of observational data that will uncover an unprecedented number of astrophysical transients and overwhelm the ability of the scientific community to interpret it. There is an urgent need for detailed models that describe the central engines behind the various transients, ranging from the collisions (or mergers) of neutron stars, which power short gamma-ray bursts (sGRBs), to deaths of massive stars, which produce long GRBs (lGRBs), to supermassive black holes (BHs), which feast at the centers of galaxies and power active galactic nuclei (AGN). There will likely be many brand new, yet to be discovered, classes of transients. BHs not only eat gas, but also energize and expel it. This effect, known as BH feedback, can affect the BH gas reservoir and, in turn, limit the feeding of the BH itself, thereby closing the feedback loop. In the active galaxy context, understanding AGN feedback is critical, because it can transform the parent galaxy. Similarly, BH feedback transforms dying stars in lGRBs and the merger ejecta in sGRBs, and thereby shapes the observed emission.

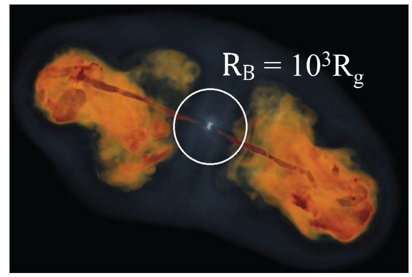

Figure 1: Largest scale 3D GRMHD simulation carried out by collaborator Lalakos of a supermassive black hole vacuuming up magnetized ambient medium (black) and producing collimated outflows, or jets (dark red), that exert feedback on the medium [1].

To understand the feedback, it is critical to describe its entire cycle: gas inflow from the feeding to BH scales and gas outflow from the hole out to the feeding scales. Doing so has been a long-standing problem because it involves a vast, disparate range of scales that so far has been impossible to bridge. For instance, a typical size of the central supermassive BH, or its gravitational radius, $R_{\rm g} \sim 1$ light-hour, is million times smaller than the size of its feeding region, or the BH sphere of influence known as the Bondi radius, $R_{\rm B} \sim 100$ light-years (ly): $R_{\rm g}/R_{\rm B} \sim 10^{-6}$. Because gas at $R_{\rm B}$ moves much slower than light, $v/c \sim \sqrt{R_{\rm g}/R_{\rm B}} \sim 10^{-3}$, the feeding timescale $T_{\rm B}$ much exceeds $R_{\rm B}/c \sim 100$ years: $T_{\rm B} \sim (R_{\rm B}/c)/(v/c) \sim (R_{\rm B}/R_{\rm g})^{3/2} T_{\rm g} \sim 10^9 T_{\rm g} \sim 0.1$ Myr, where $T_{\rm g} = R_{\rm g}/c$ is the BH timescale. This brings the temporal scale separation to 9 orders of magnitude (OOM; GRBs have a similar scale separation.) Further compounding the difficulty is that the long-term effect of the feedback occurs over the galaxy evolution timescale, ~ 100 Myr, or 12 OOM in temporal scale separation. Techniques incorporating AGN feedback over such enormous timescales come down to crude sub-grid models, which inject energy and/or momentum at the position of AGN and typically do not come close to resolving the crucial Bondi scale (see e.g., [2], for a recent attempt). Recently, PI Tchekhovskoy's former graduate student Aris Lalakos (now Caltech postdoc and collaborator (CR) on this proposal), simulated magnetized gas inflow from the Bondi scale onto a spinning supermassive BH, at the largest scale separation to date. He found that large-scale magnetic fields dramatically change the BH feedback. They do so by launching relativistic collimated outflows,

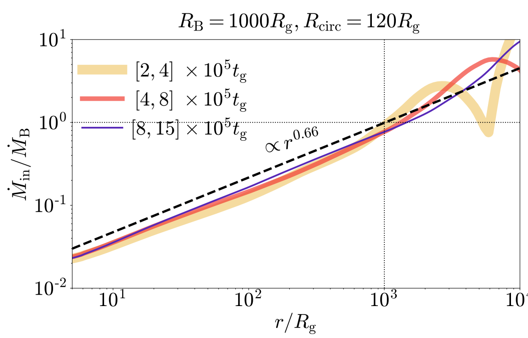

Figure 2: We discovered a *universal slope of mass inflow rate*, $\dot{M}_{\rm in} \propto r^{-0.66}$, in the largest temporal (> 6 OOM) and spatial (> 3 OOM) scale separation simulation to date of a magnetically arrested disk in the context of an idealized active galactic nucleus. GPU acceleration and unique features our GRMHD code, H-AMR, such as adaptive mesh refinement in *both* space and time, have enabled this breakthrough (Lalakos et al., in prep.).

seen in Fig. 1, which suppress the mass inflow rate, $\dot{M}_{\rm in}$, into the BH by 2 orders of magnitude, compared to the Bondi feeding rate, $\dot{M}_{\rm B}$, as seen in Fig. 2. This complex behavior complicates the construction sub-grid models. Aris discovered that the slope of the power-law scaling, $\dot{M}_{\rm in} \propto r^{0.66}$, is *universal*, i.e., is independent of the scale separation (for $10^2 \leq R_{\rm B}/R_{\rm g} \leq 10^4$). This universality motivates our approach to attacking the long-standing problem of modeling AGN feedback.

Task 1: Physics Informed Neural Network (PINN) Design

Motivation In order to understand the evolution of a galaxy, the BH dynamics play an important role. The state of the art BH simulators perform many fine temporal resolution steps to predict the outcome at an intermediate or large time scale, and require simulating the motion of magnetized ionized gas (plasma) on a curved space-time of a spinning black hole, e.g., using general relativistic magnetohydrodynamic (GRMHD) simulations. The resulting simulator can thus be prohibitively time consuming when used for larger time scale, which are the time scales of interest for galaxy level simulations. The black hole and galaxy time scales can often be separated by a factor of 10^{12}, which limits the applicability of BH simulators. In this task, we aim to build a ***Black Box for Black Hole*** **(*B^3H*)** to predict the outcome of GRMHD simulations based on a neural network (NN). Once properly trained, this neural network will be able to (1) provide rapid predictions without requiring the complex numerical calculations of traditional simulators, significantly speeding up BH simulations, and (2) serve as an efficient data generation policy for **Task 2**, which further refines the policy by reducing computation and data costs while maintaining predictive accuracy.

Training Dataset PI Tchekhovskoy's group has already collected extensive GRMHD time series data. To construct a physically tractable BH simulation training set, PI Tchekhovskoy's group considered an idealized setup: spinning BH in a uniform gas threaded with weak uniform vertical magnetic field. This setup is described by 4 physical parameters: (i) length scale separation, $10^2 \leq R_{\rm B} \leq 10^4$, (ii) BH rotation, expressed via the dimensionless spin, $0 \leq a \leq 0.99$, (iii) ambient medium rotation, expressed via the circularization radius, $0 \leq R_{\rm circ} \leq 120$, and (iv) magnetic field strength, expressed via the ratio of gas to magnetic pressure, $\beta = 100$. Fig. 2 shows that parameter (iii), $R_{\rm circ}$, does not affect the steady state solution, and our preliminary results indicate that parameter (iv) similarly has no effect. This leaves us with a 2-parameter, (i) and (ii), family of 3D GRMHD simulations, which have already been completed.

Simulation data is a volume set $V = (R_{\rm g}, 10^5 R_{\rm g}) \times (0, \pi) \times (0, 2\pi)$ in spherical polar coordinates, r, θ, φ, with the spatial resolution of $N_r \times N_\theta \times N_\varphi \simeq 300^3$ cells, uniformly spaced in $\log_{10} r, \theta, \varphi$ coordinates. We save simulation snapshots at a temporal resolution of $T = 100 T_{\rm g}$ and run the simulations for unprecedentedly long durations, out to the final time of $t_{\rm f} = (10^5 - 10^6) T_{\rm g}$. Hereafter, we adopt the units such that $R_{\rm g} = T_{\rm g} = 1$. Each simulation saves all 8 "primitive" quantities for each cell: $p^i = [\rho, T, u^i, B^i]$, where ρ is density, T temperature, and u^i and B^i the spatial contravariant components of proper velocity and magnetic field vectors ($i = 1, 2, 3$). Each of time steps from t to $t + 100$ serves as a pair of input-output for the training of the neural network. If necessary, more data can be readily generated using simulators developed by Tchekhovskoy's lab. How to minimize the cost of such data generation is the subject of **Task 2**.

Formulation We use $x^i \in \mathbb{R}^3$ to denote the coordinate position in Cartesian coordinates with BH at origin and $z-$axis being its rotation axis. The primitive quantity $p^i(t, x^i)$ is a function of time and position. Our goal is build a predictor that given a volume specified by a set of positions, $V = \{x^i\}$ and the associated $p^i(t, x^i)$ for all x^i in V at some t outputs the values of $p^i(t + T, x^i)$. In other words, we aim to learn the evolutionary mapping f_T with

$$f_T(p^i(t, V)) = p^i(t + T, V).$$

As an initial step, we aim to learn the mapping with $T = 10^2$ by using data generated by a simulator with a much finer temporal resolution of $T = 0.01$. This will *enable the bypass of small timescale simulation and obtain a significant speedup*. Once successful, we aim to combine this with techniques in **Task 2** to achieve longer-time-scale prediction with $T = 10^5$.

Challenges and Proposed Approaches While many NN can serve as universal function approximators [3], the proper data representation and architecture heavily influences the training time and accuracy [4]. We aim to design a custom architecture leveraging the physics underlying BH for this project. Specifically we plan to include the following components into a physics informed neural network (PINN).

The input and output data should exhibit *symmetry across $z = 0$ plane*. Specifically, for two points at positions (x^1, x^2, x^3) and $(x^1, x^2, -x^3)$, their associated f_T should be the same (with the exception of magnetic field, which includes an additional minus sign). To explore the symmetry, we plan to design a 3-D symmetry preserving neural network based on [5] and/or replicate data using symmetry to generate more data without incurring additional cost via data augmentation.

Another important issue is the high dimensionality for both our input and output. We will design autoencoder and decoder for the dimension reduction. We also note that the physics interaction of the primitive quantities mostly follow spatial locality. Therefore, we plan to follow the recent advancement of transformers and design a *3-D attention based transformer*, similar to [6], to exploit locality to help focus the model and to effectively mitigate the curse of dimensionality. Consider the $3 - D$ structure of our data input, unlike text stream, there is no clear ordering. We need to design *an order in which we input the data*. We suspect gradually increasing the distance to BH followed by the z value might be the best, as it preserves much of the locality. We will experiment and compare as we train the model using different approaches.

Given the fundamental physics description of the system we aim to *incorporate conservation laws into the design of neural networks*. For instance, energy and flux within a region can be computed based on the primitive quantities. The energy conservation law should hold for all time steps and at all regions. Similarly, mass, angular momentum and magnetic field should all satisfy their corresponding conservation laws. To enforce them, we can (1) add a penalty in the loss function to penalize violation of these laws, resulting in an effective soft constraint representation of the physical laws; or (2) add a projection layer before the output layer to force constraint feasibility, resulting in hard constraints. We will start without either of these approaches and evaluate the constraint violations of the trained model. If the violation is significant, which we suspect will be the case, then we will include these approaches. Co-PI Wei's expertise in large scale learning and optimization will guide the design of an efficient algorithm in this training process [7, 8].

Task 2: Cost-effective Active Imitation and Reinforcement Learning

Motivation Simulating black hole dynamics with high fidelity requires immense computational resources due to the fine temporal and spatial resolutions necessary to capture complex phenomena over vast scale separations. Our objective is to develop a *cost-effective* approach that maps initial conditions to future states while minimizing reliance on expensive full-resolution simulations.

Unlike traditional active learning, which samples independent data points, our challenge involves highly non i.i.d., time series-based simulations. We need a framework that can strategically decide when to sample partial trajectories, balancing the trade-off between exploration (using ML predictions) and the expense of running full high-fidelity simulations. Leveraging active imitation learning principles, we aim to maximize learning efficiency under tight computational constraints.

Formulation We represent the state at time t as $s_t = p^i(t, x^i) \in \mathbb{R}^8$ for all x^i, where $p^i(t, x^i)$ encompasses primitive quantities like density, velocity, magnetic fields, and temperature. The set of possible actions a_t includes: (1) Rolling out an expert policy: Using the high-fidelity simulator to compute the next state. (2) Deploying different versions of ML models, e.g., from an ensemble trained in **Task 1**. (3) Random exploration: Sampling unvisited regions of the state space.

The trajectory $\tau = (s_0, a_0, s_1, a_1, \ldots, s_T)$ describes the evolution of the system over time, with an associated cost $C(\tau)$ for using high-fidelity simulations. Our objective is to minimize the cumulative prediction error while respecting a computational budget B:

$$\min_{\pi} \mathbb{E}_{\tau\sim\pi}\left[\text{Cumulative Error}(\tau)\right] \quad \text{s.t.} \quad \sum_{\tau} C(\tau) \leq B.$$

This formulation necessitates a principled approach to selecting the optimal action a_t at each state s_t, balancing the trade-off between model prediction accuracy and computational efficiency.

Challenges and Prior Art The high cost of simulating black hole dynamics makes traditional reinforcement learning (RL) approaches infeasible, as they require extensive exploration. Imitation learning (IL) methods, such as behavioral cloning [9] and interactive strategies like DAgger [10] and AggreVaTe(D) [11], improve efficiency by learning from expert demonstrations. However, these methods often assume access to a single, near-optimal oracle and do not account for the cost of querying, making them impractical for resource-constrained environments. In realistic settings, access to a perfect oracle is often unavailable or prohibitively expensive. Instead, we may have multiple suboptimal oracles. Naively imitating a single oracle can yield suboptimal results, as it fails to utilize the complementary utilities of different oracles based on the current state. Although this problem remains largely unexplored, recent studies [12, 13] highlight the potential of leveraging multiple oracles through active policy selection and exploration.

Proposed Approach To tackle the computational challenges inherent in high-fidelity black hole simulations, we introduce a novel framework that combines *active imitation learning* and *reinforcement learning*, strategically minimizing the need for expensive full-resolution queries.

Our method begins with an *active imitation learning* component, where, instead of relying on exhaustive expert trajectories, we selectively query specific states. The high-fidelity simulator provides precise responses only at these critical points, allowing us to build an efficient policy with minimal expert intervention. By focusing on states that yield the most informative feedback, we reduce both computational and data burdens while ensuring that essential dynamics are accurately learned.

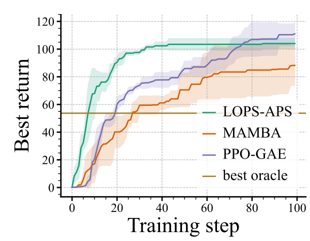

Figure 3: Activizing state and action exploration (in green) allows a sample advantage over and imitation learning baselines [13].

We extend this approach into *active reinforcement learning* to further optimize the learning process. Our active RL strategy dynamically explores the state and action spaces, making informed decisions about when to use ML models from the ensemble trained in **Task 1** or when to perform random exploration to discover uncharted regions of the state space. We frame this decision-making as an *active policy selection* problem, where the goal is to maximize information gain while minimizing unnecessary computational costs. This builds upon recent advances, such as PI Chen's work [13, 14] on active RL with access to multiple suboptimal policies (see Fig. 3), and enables accelerated policy optimization by jointly reasoning over which states to explore and which "oracle" (e.g., expert policy or ML models) to follow. A key innovation in our approach is the use of *active state exploration*, which strategically targets states with high value estimate uncertainty. By focusing on these uncertain regions, we improve the model's ability to generalize while optimizing the allocation of computational resources. In the context of black hole simulations, this approach means avoiding exhaustive simulations at every timestep and instead identifying critical points that are most informative for maintaining physical fidelity. Our method ensures that laws of conservation, such as mass, energy, and angular momentum, are respected while reducing computational load.

Benefit to Research Community We plan to make *the data used for training the model*, the architecture of the model and the model itself (architecture together with trained weights), the open-source active learning software package publicly available. The research community can then use data to train additional models to study spinning black holes, use the architecture designed for this study in other learning tasks, use the active learning software for cost-effective simulation, or simply use B^3H to predict and generate time series data for research purposes.

References

[1] Lalakos, A; Tchekhovskoy, A; Bromberg, O; Gottlieb, O; Jacquemin-Ide, J; et al. "Jets with a Twist: The Emergence of FR0 Jets in a 3D GRMHD Simulation of Zero-angular-momentum Black Hole Accretion," A*strophysical Journal*, v. 964(1), 2024, p. 79. `https://ui.adsabs.harvard.edu/abs/2024ApJ...964...79L`

[2] Su, KY; Natarajan, P; Cho, H; Narayan, R; Hopkins, PF; et al. "Bridging Scales: Coupling the galactic nucleus to the larger cosmic environment," *arXiv e-prints*, 2024, p. arXiv:2410.13235. `https://ui.adsabs.harvard.edu/abs/2024arXiv241013235S`

[3] Hornik, K; Stinchcombe, M; White, H. "Multilayer feedforward networks are universal approximators," *Neural networks*, v. 2(5), 1989, p. 359–366

[4] Krizhevsky, A; Sutskever, I; Hinton, GE. "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, v. 25, 2012

[5] Esteves, C; Allen-Blanchette, C; Makadia, A; Daniilidis, K. "Learning so (3) equivariant representations with spherical cnns," In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, p. 52–68

[6] Fuchs, F; Worrall, D; Fischer, V; Welling, M. "Se (3)-transformers: 3d roto-translation equivariant attention networks," *Advances in neural information processing systems*, v. 33, 2020, p. 1970–1981

[7] Mansoori, F; Wei, E. "A fast distributed asynchronous newton-based optimization algorithm," *IEEE Transactions on Automatic Control*, v. 65(7), 2019, p. 2769–2784

[8] Niu, X; Wei, E. "Fedhybrid: A hybrid federated optimization method for heterogeneous clients," *IEEE Transactions on Signal Processing*, v. 71, 2023, p. 150–163

[9] Pomerleau, DA. "ALVINN: An autonomous land vehicle in a neural network", 1988

[10] Ross, S; Gordon, G; Bagnell, D. "A reduction of imitation learning and structured prediction to no-regret online learning", 2011, p. 627–635

[11] Ross, S; Bagnell, JA. "Reinforcement and imitation learning via interactive no-regret learning," *arXiv preprint arXiv:1406.5979*, 2014

[12] Cheng, CA; Kolobov, A; Agarwal, A. "Policy improvement via imitation of multiple oracles", 2020, p. 5587–5598

[13] Liu, X; Yoneda, T; Wang, C; Walter, M; Chen, Y. "Active policy improvement from multiple black-box oracles," In *Proc. International Conference on Machine Learning (ICML)*, 2023

[14] Liu, X; Yoneda, T; Stevens, R; Walter, M; Chen, Y. "Blending imitation and reinforcement learning for robust policy improvement," In *The Twelfth International Conference on Learning Representations (ICLR)*, 2024