Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Einstein-Net: Modelling Black Holes using Error-Bounded Neural Networks

Hassan Ugail

h.ugail@bradford.ac.uk

University of Bradford

Research Article

Keywords: numerical relativity, black holes, initial data, constraint equations, physics-informed neural networks, error bounds, conformal prediction, apparent horizons

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11776989/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

EINSTEIN-NET: MODELLING BLACK HOLES USING ERROR-BOUNDED NEURAL NETWORKS

Hassan Ugail
Centre for Visual Computing and Intelligent Systems
University of Bradford
United Kingdom

November 17, 2026

ABSTRACT

In this paper, we investigate the use of neural networks for generating black-hole initial data that satisfy the constraint equations of general relativity. Existing neural approaches to solve this problem often enforce the constraints via loss penalties without providing error estimates for individual predictions. It is noteworthy that any solver in this domain must compute new initial data for each choice of mass, momentum, and spin. Here we introduce Einstein-Net, in which the Bowen–York momentum constraint is satisfied exactly, and the puncture structure, positivity, asymptotic flatness and Schwarzschild limit are incorporated into the output formula. Thus, in our case, only the Hamiltonian constraint remains to be learnt. To do this, we represent the Arnowitt–Deser–Misner mass explicitly using the Gauss-law identity. We also derive an analytical error bound from the Hamiltonian residual, evaluate it numerically, and assess a separate acceptance rule using independent reference cases. Our experiments on single black holes with parallel momentum and spin provide median mass errors of a few parts in ten thousand. Furthermore, the generated data reproduce the expected apparent horizons and physical mass and spin relations. Thus, we show that the neural network setup we present here can generate black-hole initial data accurately and effectively.

***K*eywords** numerical relativity · black holes · initial data · constraint equations · physics-informed neural networks · error bounds · conformal prediction · apparent horizons

1 Introduction

In numerical relativity, to simulate the evolution of two black holes towards a merger, one must provide data for an initial spatial configuration. However, one cannot prescribe this configuration arbitrarily. Of the ten Einstein equations available, four impose constraints on the initial data. This leads to a nonlinear elliptic problem [1, 2, 3, 4]. Various approaches can be adopted to solve this problem. The conformal decomposition introduced by York separates the spatial metric into a conformal factor and background geometry, while the extrinsic curvature is divided into its trace and trace-free components [3]. For puncture data, under the usual choices, this leaves a linear momentum constraint and a nonlinear equation to be solved. The construction can be simplified further using the Bowen–York solution for the momentum and spin [5]. Brandt and Brügmann further showed that the singular part of the conformal factor can be prescribed explicitly, leaving a regular correction term to be determined [6]. This is the standard approach for the moving-puncture simulations of binary black holes [7, 8]. Though such spectral methods are generally accurate [9], the difficulty with them is that for a new solution, one is always required to recompute the mass, momentum, or spins. It must be pointed out that for gravitational-wave studies involving large catalogues of configurations, this repeated calculation becomes an undesirable computational burden.

Artificial neural learning offers an alternative mechanism for computing such families of initial data. In a physics-informed neural network, the coordinates relating to the initial data are provided as input, and the resulting field is automatically differentiated to obtain the residual of the governing equation [10, 11, 12]. By treating the physical

parameters as additional input, a given model can represent a range of configurations. Similar methods have been utilised for studying matter-sourced metrics [13], four-dimensional systems with scalar fields [14], vacuum equations using domain decomposition [15], simulation of puncture evolution in spherical symmetry [16], and solving holographic equations [17]. There is also work that uses the representation of Einstein's equations as continuous-time recurrent networks [18]. However, one difficulty remains when the constraints are introduced only as terms in the loss function. A small penalty does not necessarily guarantee that a constraint has been satisfied exactly, and the residual itself does not provide a useful estimate of the error in a global quantity examined. In numerical relativity, this distinction is particularly important. This is because constraint violations must routinely be monitored before initial data are accepted for the simulation of a given evolution.

Here, we investigate whether a neural model can generate black-hole initial data while satisfying as many physical requirements as possible and independently of the training. We consider a single puncture with parallel linear momentum and spin, which provides a useful starting point for establishing our method. The approach we adopt here uses known hard constraints in neural trial functions [10, 19] along classical comparison principles for elliptic operators [20]. Although these ideas have been individually established, we believe it is the combination with an explicit mass representation and a residual-based error estimate that provides the proposed novel solution for puncture data considered here.

A crucial part of our method relates to the form of the output the network may produce. Thus, instead of asking a network to learn every component of the initial data, we require the Bowen–York tensor to satisfy the momentum constraint directly. For this purpose, we incorporate into the output formulation the puncture singularity, positivity of the conformal factor, its asymptotic behaviour, and the Schwarzschild solution at zero momentum and spin. These properties therefore hold for the weights of an arbitrary network. Thus, the only constraint to be solved numerically is that corresponding to the Hamiltonian.

The Arnowitt–Deser–Misner (ADM) mass is also treated separately from the shape of the conformal factor. Its coefficient is fitted using the integral identity obtained from the Hamiltonian constraint, which we refer to as Gauss's law. This is useful to note because a small local residual does not necessarily imply that a global quantity such as mass can be determined accurately. The ablation experiments we carried out indeed show that a model with a smaller residual can still possess a substantially larger mass error.

The final stage of the construction relates to the accuracy of the recovered solution. To do this, we use the comparison principle for the Hamiltonian residual. Using this, we derive a pointwise estimate for the conformal-factor error together with a corresponding bound on the ADM mass. These results are indeed the analytical properties of the underlying continuum problem. In order to assess the reliability of these computations, we study the convergence behaviour and complement it with a leading-order mass consistency check and an acceptance criterion which is then tested on independent classical solutions. This acceptance criterion can be considered as an empirical statistical procedure rather than part of the analytical error analysis.

Throughout the paper, we therefore distinguish between our theoretical results, the corresponding numerical evaluation, and the empirical validation procedure. Related work on such interpretation and assessment of numerical diagnostics can also be found in our previous studies [22, 23].

The remainder of the paper is organised as follows. In Section 2, we introduce the constraint equations and the quantities required to characterise a single black hole. In Section 3, we present the Einstein-Net architecture and its training methodology. In Section 4, we derive the error estimate and introduce the three acceptance checks. In Section 5, we describe the independent numerical solvers and the horizon-finding procedure used for validation. The numerical results, including those for the conformal factor, ADM mass, apparent horizons, error estimates, and computational cost, are reported in Section 6. The limitations of the proposed approach are discussed in Section 7, and the paper is concluded in Section 8. Additional mathematical derivations and numerical implementation details are provided in the appendices. Information on the associated code and datasets is given at the end of the manuscript.

2 The constraint equations for a single black hole

Throughout this work, we use geometrised units, $G = c = 1$, and follow the conventions in [24, 25]. A spatial hypersurface is characterised by its metric γ_{ij} and extrinsic curvature K_{ij}. For vacuum initial data, these quantities must satisfy the Hamiltonian constraint $R + K^2 - K_{ij}K^{ij} = 0$ and the momentum constraint $D_j(K^{ij} - \gamma^{ij}K) = 0$. Here, R denotes the scalar curvature of the spatial metric, K is the trace of the extrinsic curvature, and D is its covariant derivative. The conformal decomposition takes the form $\gamma_{ij} = \psi^4\tilde{\gamma}_{ij}$ and $K_{ij} = \psi^{-2}\tilde{A}_{ij} + \frac{1}{3}\gamma_{ij}K$, where $\tilde{A}_{ij}$ is trace free. Choosing a conformally flat background, $\tilde{\gamma}_{ij} = \delta_{ij}$, and maximal slicing, $K = 0$, reduces the constraints to,

$$\partial_j \tilde{A}^{ij} = 0, \qquad \Delta\psi + \tfrac{1}{8}\,\tilde{A}_{ij}\tilde{A}^{ij}\,\psi^{-7} = 0. \tag{1}$$

Here Δ is the Laplacian associated with the flat background and indices on $\tilde{A}_{ij}$ are raised using the flat metric. Thus, $\tilde{A}_{ij}\tilde{A}^{ij}$ represents its ordinary squared Frobenius norm. The momentum equation is linear and, for a puncture at the origin with momentum P^i and spin S^i, it admits the Bowen–York solution [5] such that,

$$\tilde{A}^{ij} = \frac{3}{2r^2}\Big[P^i n^j + P^j n^i - \big(\delta^{ij} - n^i n^j\big)P_k n^k\Big] + \frac{3}{r^3}\Big[\epsilon^{ikl}S_k n_l\, n^j + \epsilon^{jkl}S_k n_l\, n^i\Big], \tag{2}$$

where $n^i = x^i/r$. In the particular family considered here, both P and S lie along the z axis. The contraction can then be written explicitly as,

$$\tilde{A}_{ij}\tilde{A}^{ij} = \frac{9P^2}{2r^4}\big(1 + 2\cos^2\theta\big) + \frac{18S^2}{r^6}\sin^2\theta. \tag{3}$$

There is no momentum–spin cross term in this expression. Thus, the remaining Hamiltonian constraint is a single nonlinear elliptic equation for the conformal factor considered here. If we adopt the puncture method of Brandt and Brügmann [6], we can now write,

$$\psi = 1 + \frac{m}{2r} + u. \tag{4}$$

Note that, in this expression, m is the minimal puncture mass and u is a regular correction which leads to zero at infinity. Near the puncture, the source, we see that $\tilde{A}_{ij}\tilde{A}^{ij}\psi^{-7}$ behaves as $r^7\tilde{A}_{ij}\tilde{A}^{ij}$. This provides a contribution of order r^3 from linear momentum and order r from spin. Therefore, the source can remain bounded at the puncture. As a result, a bounded regular solution can be obtained without treating the puncture as a separate inner boundary. This can also lead to a second asymptotically flat end of the slice, as required by the puncture description of a black hole.

We also note that the ADM mass provides a particularly useful test of our solution. It can be obtained either from the asymptotic coefficient of the conformal factor or by integrating the Hamiltonian constraint, giving,

$$M_{\mathrm{ADM}} = m + 2\lim_{r\to\infty} r\,u = m + \frac{1}{16\pi}\int \tilde{A}_{ij}\tilde{A}^{ij}\,\psi^{-7}\,\mathrm{d}V. \tag{5}$$

The first expression reads the mass from the $1/r$ term of ψ, while the second follows from the divergence theorem. Here, when the Hamiltonian constraint is satisfied, the agreement is exact. We also note that the ADM mass is a geometric invariant of the asymptotically flat hypersurface, i.e., the volume integral is conformally flat in the Cartesian coordinates which we have adopted here. Henceforth, we refer to this relation as Gauss's law for the Hamiltonian constraint. It provides an efficient way of determining the mass during the neural network training. One must also note that the momentum of ADM, both linear and angular, is already specified by the Bowen–York quantities P^i and S^i [5].

A further check can be derived for relatively small values of momentum and spin. Substituting the unperturbed conformal factor $\psi_0 = 1 + m/2r$ into the integral in Eq. (5) gives the leading-order change in mass. The angular integrals of Eq. (3) are $\int(1 + 2\cos^2\theta)\,\mathrm{d}\Omega = 20\pi/3$ and $\int \sin^2\theta\,\mathrm{d}\Omega = 8\pi/3$. With the substitution $s = m/2r$, the corresponding radial integrals become $\int_0^\infty (1+s)^{-7}\mathrm{d}s = 1/6$ and $\int_0^\infty s^2(1+s)^{-7}\mathrm{d}s = 1/60$. We therefore obtain the following result.

Proposition 1 (Leading-order mass)**.** *For a single puncture of bare mass m with parallel momentum P and spin S,*

$$M_{\mathrm{ADM}} = m + \frac{5P^2}{8m} + \frac{2S^2}{5m^3} + O\big(P^4, S^4, P^2S^2\big). \tag{6}$$

Appendix A gives the derivation. The coefficients given there provide an independent small-parameter prediction for the mass. This prediction is used as a second acceptance check, as described in Section 4. This procedure also allows us to test the classical reference solver before using it to evaluate the trained network.

The location of the black hole is obtained from its apparent horizon, defined as the outermost closed surface for which the expansion of outgoing null rays vanishes [26, 27]. If s^i is the unit normal to the surface, its expansion is $\Theta = D_i s^i + K_{ij}s^i s^j - K$. In the present conformally flat maximal setting, the physical and flat unit normals are related by $s^i = \psi^{-2}\hat{s}^i$, giving,

$$\Theta = \psi^{-2}\big(\partial_i \hat{s}^i + 4\,\hat{s}^i\partial_i \ln\psi\big) + \psi^{-6}\,\tilde{A}_{ij}\hat{s}^i\hat{s}^j. \tag{7}$$

For Schwarzschild data, $\tilde{A}_{ij} = 0$ and $\psi = 1 + m/2r$, the horizon is shown to be located at $r = m/2$. Here, the area A determines the irreducible mass $M_{\mathrm{irr}} = \sqrt{A/16\pi}$, from which the Christodoulou mass $M_H^2 = M_{\mathrm{irr}}^2 + S^2/(4M_{\mathrm{irr}}^2)$ and dimensionless spin $\chi = S/M_H^2$ follow [28, 27] can be obtained. One must also note that for axisymmetric vacuum data, the horizon angular momentum is S. With this setup and the quantities considered there, we are in a position to assess the physical properties of the generated black holes.

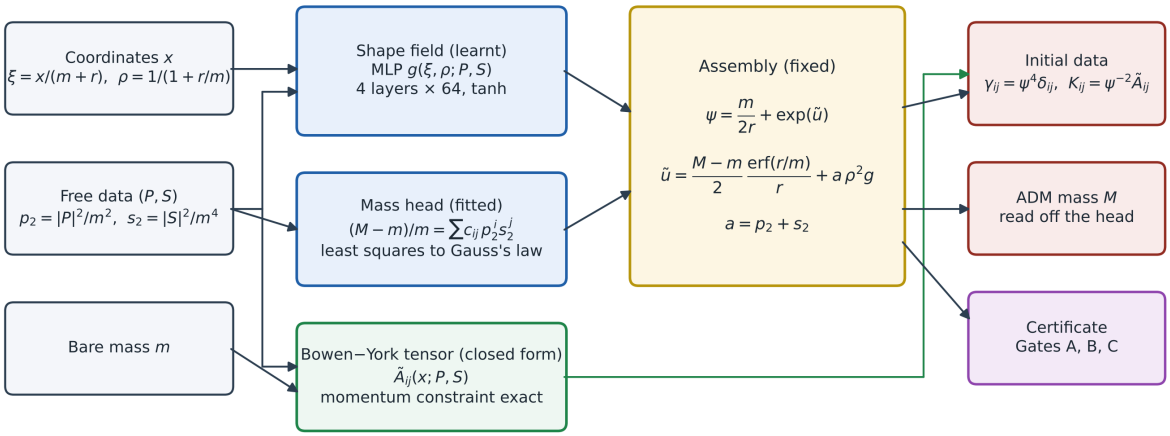


Figure 1: Einstein-Net for a single puncture. The coordinates are defined through the variables ξ and ρ, and the free data through p_2 and s_2. The Bowen–York tensor is a closed-form function which is not learned. The shape field is the multilayer perceptron considered, and the mass head is a tiny polynomial function whose coefficients are then fitted to Gauss's law. The assembly is carried out using fixed arithmetic, which then determines various properties independently of the weights. The output from the network comprises the initial data, the ADM mass read off from the head, and the certificate as described in Section 4.

3 Einstein-Net

The Einstein-Net architecture is designed to avoid directly predicting the metric components or the conformal factor, but instead produces a regular scalar field along with a set of coefficients related to the ADM mass. This output is then fed into a pre-defined expression in order to assemble the conformal factor. As shown in Fig. 1, the modular nature of this framework allows one to enforce specific properties of the initial data.

To define the inputs to the neural network, we take $r = |x|$ and represent the radial and position coordinates in a compact form, such that,

$$\rho = \frac{1}{1 + r/m}, \qquad \xi = \frac{x}{m + r}. \tag{8}$$

Here, $\rho \in (0, 1]$ and ξ belongs to the unit ball. The dimensionless input parameters are $p_2 = |P|^2/m^2$ and $s_2 = |S|^2/m^4$, with $a = p_2 + s_2$. The regular shape field can be modelled by the multilayer perceptron $g(\xi, \rho; P, S)$ having ten input units, four hidden layers containing sixty-four tanh units per layer and one output unit. The mass contribution is represented by a separate polynomial function such that,

$$\frac{M-m}{m} = \sum_{1 \le i+j \le 3} c_{ij}\, p_2^i\, s_2^j. \tag{9}$$

The nine coefficients c_{ij} are determined during training. Next, the conformal factor is then assembled using,

$$\psi(x) = \frac{m}{2r} + \exp(\tilde{u}), \qquad \tilde{u} = \frac{M-m}{2} \frac{\mathrm{erf}(r/m)}{r} + a\,\rho^2\, g(\xi, \rho; P, S). \tag{10}$$

The initial data make use of $\gamma_{ij} = \psi^4 \delta_{ij}$ and $K_{ij} = \psi^{-2} \tilde{A}_{ij}$, along with the Bowen–York tensor from Eq. (2). From the above equation (Eq. (10)), the contribution of each term can be seen clearly, which shows the part each term plays. The term $m/2r$ is preserved. Hence, the correct puncture singularity is retained. Similarly, the exponential term ensures that $\psi > m/2r > 0$ away from the puncture.

The metric tensor one can derive is positive definite. The function $\mathrm{erf}(r/m)/r$ is well behaved at the singularity, and for large radii, it converges to the asymptotic form $1/r$ very rapidly. The coefficient of this function provides the isotropic asymptotic term, from which the ADM mass can be computed.

As discussed earlier, our method can extract the mass from the network directly without having to solve the underlying field equations for this purpose. The ρ^2 factor involved in the residual shape correction determines its decay as r^{-2}. Furthermore, when $P = S = 0$, the condition $a = 0$ is automatically satisfied, and hence the solution reduces to the Schwarzschild solution regardless of the learned weights. In agreement with Proposition 1, the correction starts from the quadratic order in both momentum and spin. Since the variables are bounded, we do not need to represent the coordinates using arbitrary magnitudes, and hence the input space of the shape network is also constrained.

Theorem 1 (Properties by construction)**.** *Let the shape-field weights and the mass coefficients take any set of values. The initial data assembled by Eq.* (10) *satisfy the momentum constraint exactly. The metric γ_{ij} is also positive definite everywhere away from the puncture, and the data reduce to the Schwarzschild data $\psi = 1 + m/2r$ when $P = S = 0$. They are also asymptotically flat with,*

$$\psi = 1 + \frac{M}{2r} + O(r^{-2}), \qquad M = m + m \sum_{1 \le i+j \le 3} c_{ij}\, p_2^i s_2^j,$$

and have ADM linear momentum P^i and ADM angular momentum S^i. The only equation that the data can violate is the Hamiltonian constraint.

The relevant physical properties can be inferred directly using the assembly formula. The ADM momenta, as well as the momentum constraint, are fixed by the Bowen–York tensor [5]. Due to the exponential term being positively valued, the overall positivity is ensured. When $P = S = 0$, we obtain $M = m$, $a = 0$, which gives $\psi = 1 + m/2r$. As can be seen, for sufficiently large radii the finite tanh network has a desirable property, namely its magnitude is bounded and $a\rho^2 g = O(r^{-2})$. Similarly, since $\tilde{u} = (M - m)/(2r) + O(r^{-2})$, the expansion of the exponential gives the asymptotic form. Thus, the theorem holds regardless of the effectiveness of the training process. This approach has the advantage of properties, which would normally be approximated via penalty terms, now being ensured by the output definition itself.

The training protocol was based on the Hamiltonian constraint, with the ADM mass being determined separately. An important part of the training procedure was the optimisation of the shape field, which aims to minimise an estimate of the squared L^2 norm of the Hamiltonian residual obtained by Monte Carlo integration,

$$\mathcal{L}_{\mathrm{res}} = \mathbb{E}\big[\, r^3\, R(x; P, S)^2 \,\big], \qquad R = \Delta\psi_N + \tfrac{1}{8}\, \tilde{A}_{ij} \tilde{A}^{ij}\, \psi_N^{-7}. \tag{11}$$

One can compute the Laplacian in the residual R via automatic differentiation of the assembled conformal factor ψ_N. The sampling protocol uses a log-uniform distribution to sample radii between $0.02\, m$ and $200\, m$. In the context of angular space, directions are selected randomly with equal probability, following a uniform distribution. Furthermore, the two free parameters are drawn uniformly from $P/m \in [0, 1]$ and $S/m^2 \in [0, 0.8]$. The vectors P and S are aligned with the z axis. The factor r^3 compensates for the log-uniform radial density distribution, which gives a probability density of $1/[r \ln(r_1/r_0)]$ for a radius that falls within the interval $[r_0, r_1]$. However, the angular density is $1/4\pi$. The expectation of the quantity $4\pi \ln(r_1/r_0) r^3 f$ is the volume integral of f over the sampled shell. The optimisation process is insensitive to the constant factor, so this factor can be neglected from further consideration. Thus the objective estimates the integral $\int R^2 \,\mathrm{d}V$ over the sampled shell, in the chosen Cartesian conformal representation.

During the optimisation, the mass coefficients do not change and therefore require periodic recalculation. Specifically, these coefficients are refitted every fifty optimisation iterations through a linear least squares regression on sixty-four different parameter sets using the spherical product quadrature method, to the Gauss-law identity (Eq. (5)). In our approach, the learning of the local shape correction function and the global mass coefficients is kept separate. An earlier joint optimisation of the two terms stalled and resulted in an inaccurate mass estimate, though the residual was relatively small. The incorrect mass estimate is understandable because the source of the equation is concentrated near the black hole, whereas the mass is mainly determined by the long range $1/r$ coefficient. As a result, we abandoned the joint optimisation and simply use the Gauss-law fit, which constrains the latter coefficient.

On the other hand, we have not introduced additional penalty terms to the main functional as is usually done in the conventional approaches. Neither have we used supervised reference solutions as a means of obtaining a desired solution. The shape field in our case comprise of $13,249$ trainable weights. Adam was used to optimise the model [29]. The model was initially trained for a few thousand iterations, and in each iteration we used 2048 collocation points. Towards the latter part of the optimisation, we also used the limited-memory BFGS (L-BFGS) optimisation algorithm, as described by Liu and Nocedal [30]. All implementations were carried out using the PyTorch library [31].

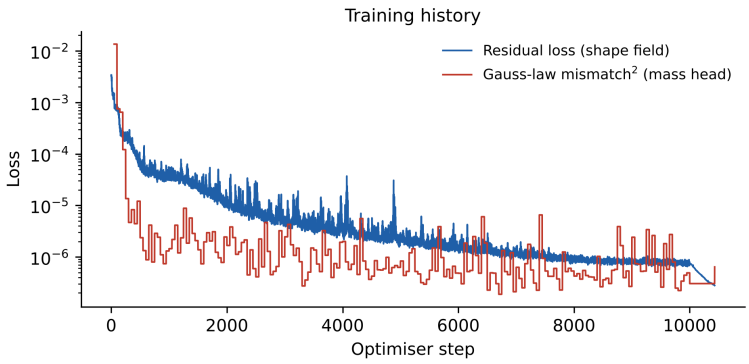


Figure 2: Training history. The residual loss of the shape field, Eq. (11), and the squared Gauss-law mismatch of the mass head is shown.

As can be seen from the plot in Figure 2, there is a drop in the residual loss following the optimisation by L-BFGS. One can also observe that the optimised residual loss reaches 2.8×10^{-7}, which is about four orders of magnitude lower than its initial value.

4 The error bound and the three gates

Once trained, the network can then produce a conformal factor ψ_N for any of the specified parameter pairs. To test the accuracy, our analysis uses a battery of three checks.

For the first check, we let $e = \psi_N - \psi$ denote the error relative to the true conformal factor, and we let R be the residual of Eq. (11). Using the exact Hamiltonian constraint, we subtract it from the network residual. Then, by applying the mean value theorem for the function x^{-7}, we have the following error equation,

$$\Delta e + \bar{V} e = R, \qquad \bar{V} = -\tfrac{7}{8}\, \tilde{A}_{ij}\tilde{A}^{ij}\, \bar{\psi}^{-8} \leq 0. \tag{12}$$

Here, the intermediate value $\bar{\psi}$ lies between the two conformal factors at each point, with one of the factors being exact and the other predicted. It is especially noteworthy that the quantity $\bar{V}$ is nonpositive. The elliptic operator $\Delta + \bar{V}$ consequently satisfies the comparison principle [20]. Furthermore, comparison with the Newton potential of $|R|$, which satisfies the equation $\Delta\mathcal{N} = -|R|$, gives the following result.

Proposition 2 (Analytically certified bound)**.** *Let ψ_N be any positive function with $\psi_N - 1 - m/2r \to 0$ at infinity, let R be its Hamiltonian residual, and let $\mathcal{N}(x) = \frac{1}{4\pi}\int |R(y)|\,|x-y|^{-1}\,\mathrm{d}y$. Then the error $e = \psi_N - \psi$ against the true solution satisfies, at every point of the slice,*

$$|e(x)| \leq \mathcal{N}(x), \tag{13}$$

and the ADM mass $M_N = m + 2\lim r(\psi_N - 1 - m/2r)$ read off ψ_N satisfies

$$|M_N - M_{\mathrm{ADM}}| \leq \frac{1}{2\pi}\Big(\Big|\int R\,\mathrm{d}V\Big| + \int \tfrac{7}{8}\,\tilde{A}_{ij}\tilde{A}^{ij}\,\big(\psi_N - \mathcal{N}\big)^{-8}\,\mathcal{N}\,\mathrm{d}V\Big). \tag{14}$$

wherever $\psi_N > \mathcal{N}$.

A proof is provided in Appendix B. This is an estimate about the continuum equation, made without the need for linearisation. It is noteworthy that the intermediate value $\bar{\psi}$ does not require explicit determination, despite its appearance in the error equation. Eq. (13) gives the inequality $\bar{\psi} \geq \psi_N - \mathcal{N}$. The Newton potential and residual can be recovered from the predicted solution itself, without needing a classical reference solution.

In addition to the above, the numerical implementation requires a further qualification. In the experiments, the integrals are approximated using a spherical product quadrature of $30,720$ nodes covering $10^{-4}\,m < r < 10^{5}\,m$. We compute the potential at each node in the discretised domain, and the maximum of the computed potentials is used as a surrogate for the continuum supremum. The positivity condition $\psi_N > \mathcal{N}$ is likewise checked numerically. Given that they cover only a part of the radial domain and are based on quadrature approximation, the results produced by these methods are not rigorous enclosures of the analytical bound. In order to conclude that, one would need a fully validated numerical integration scheme with interval bounds. In Section 6 and Appendix C, we quantify the differences between the three

levels of quadrature resolution adopted in this work. We have also utilised these techniques to study the radial tails and to compare the results with independently generated random points.

Next, we compare the calculated mass to the theoretically expected expansion behaviour described in Proposition 1. To do this, we first estimate the quartic terms from sweeps of the classical reference solver. An output passes if its mass lies within a band consisting of twice the fitted quartic contribution plus $10^{-3}\,m$. The above band is not a rigorous estimate of the neglected terms. This approach is intended to identify grossly flawed solutions and to provide a simple physical validation against an analytical benchmark. The mechanism has a deliberately broader tolerance than the other two checks.

It is also important to evaluate the frequency with which accepted outputs have mass errors above a given tolerance. We now compute the numerical value of the bound in Equation (14), and we denote it by $\lambda(P, S)$. Based on the above discussion, we need to consider two important parameters, the specified mass error tolerance (τ) and the predefined risk level (δ). We first generate a total of n pairs of parameters by random sampling from a uniform distribution in the defined training box, and obtain accurate reference masses for the parameter pairs. Based on the above analysis, we record details of all candidate thresholds chosen from the set of bound values λ. Here, n_λ is the number of accepted calibration instances, while k_λ is the number of those instances for which the error is beyond the specified tolerance τ. Writing $U(k_\lambda, n_\lambda)$ for the one-sided Clopper–Pearson upper confidence limit [32] at confidence $1 - \alpha$, we select the largest threshold $\lambda^\star$ satisfying $U(k_\lambda, n_\lambda) \leq \delta$.

The procedure is similar to risk-controlling prediction-set methods [21], but we do not prove the monotonicity of the conditional failure risk. Raising the threshold value can admit cases with either higher or lower probabilities of failure. Given the limitations of our selection procedure, we do not assign a finite-sample coverage guarantee to the selection itself. Once the threshold value, say $\lambda^\star$, is frozen, we then build an independent test set. When a total of n_t instances are observed to pass, and k_t of these are found to have errors beyond the specified tolerance τ, the one-sided Clopper–Pearson limit $U(k_t, n_t)$ gives a valid confidence bound for the conditional failure probability of the fixed acceptance criterion. The validity of this assertion is guaranteed because the accepted test cases are independently sampled from the underlying population, subject to the condition that they have passed the relevant criteria. The guarantee is based on the assumption that future inputs are drawn from the same probability distribution as during training, namely the distribution defined by the training box. The guarantee does not extend to extrapolated parameters outside of the specified range. In such situations, the residual based assessment can be utilised, subject to its numerical evaluation.

5 Reference solutions and the horizon finder

To verify the accuracy of our network, we have used both analytical checks and independent numerical solutions. To measure the performance of the network, we use the Schwarzschild limit as a baseline. Furthermore, Proposition 1 gives a small-parameter mass expansion. Finally, two conventional numerical solvers, implemented independently of Einstein-Net, have provided the necessary data to perform our experiment.

The more accurate solver takes into account the axisymmetric nature of a single puncture with parallel momentum and spin. The regular correction can be expanded in Legendre polynomials, i.e., $u(r, \mu) = \sum_\ell u_\ell(r) P_\ell(\mu)$, where $\mu = \cos\theta$. Only even modes appear, because the source in Equation (3) is even in μ. The radial coordinate is compactified using $r = s/(1 - s)$, where $0 < s < 1$, so that infinity is represented by the endpoint $s = 1$. The radial equations are discretised via a second order difference scheme on a cell centred grid. The Newton–Raphson method, which is also called Newton iteration, is used to solve the resultant nonlinear system of equations. Gauss–Legendre projection provides the coupling between the angular modes. Three radial resolutions are used in order to estimate the accuracy of the solutions using Richardson extrapolation. The computed ADM mass has a small radial discretisation error, estimated to be less than 2×10^{-8} in all the configurations we have discussed above. The combined effect of the angular truncation, compactification scale and Newton tolerance on the solution is less than $10^{-7}\,m$. Details of the numerical parameters we have used can be found in the appendix (see Appendix C).

Furthermore, we have solved the same problem using a fully three dimensional Cartesian finite difference scheme - using the Newton iteration method combined with algebraic multigrid preconditioning and conjugate gradients. This solver was different from the previous one in that it did not involve axisymmetry, and it had a Robin boundary condition on the faces of a finite cube. As the results using this approach are not very accurate, our cost comparison (described in Section 6) is restricted to this method.

Using an independent apparent-horizon finder [26, 33], we examined the geometry of the generated data. The surface is mathematically represented by the function $r = h(\theta) = \sum_\ell a_\ell P_\ell(\cos\theta)$, where both even and odd modes contribute to the sum. This is so because a moving black hole may not have a horizon centred on the puncture. A level set function

$F(x) = |x| - h(\theta(x))$ was used to determine the flat normal $\hat{s} = \nabla F/|\nabla F|$ as well as its divergence, which was evaluated using automatic differentiation. The derivatives of the natural logarithm of the conformal factor, $\ln \psi$, were computed for both the network solution and the corresponding reference solution using different numerical techniques. In contrast to the reference solution, the derivatives of the network were obtained via automatic differentiation. For the reference solution, the gradient was approximated via the use of central differences. The expansion given by Equation (7) was fitted with forty Gauss–Legendre angular nodes and twelve modes. For Schwarzschild, the resulting surface is $r = m/2$. The expansion on this surface is approximately 10^{-15}. The convergence tests in Appendix C are the results we have obtained to verify the accuracy of the solution. Horizon agreement is a numerical check of the initial data, but it is not an analytical certificate. This is because Proposition 2 does not bound the gradient needed to certify the horizon. Additionally, the surface area gives the irreducible and horizon masses and, together with S, the dimensionless spin.

6 Results

Here we discuss some of the properties guaranteed by our architecture. We show the accuracy of our approach within the training range. We consider the apparent horizons and the physical quantities derived from the initial data. We examine the error bound, including when the network is extrapolated beyond the training regime. And finally, we consider the computational cost of our approach.

6.1 Exactness and accuracy across the family

The Schwarzschild limit provides a means of checking the network directly. At $P = S = 0$, the largest difference between the network output and $1 + m/2r$ was 1.8×10^{-15} over five thousand randomly selected points which cover four decades in radius. To check the effect of initialisation, we repeated the above experiment, but this time all the weights were randomly initialised. We obtained the same result. Similarly, the anisotropy of the asymptotic $1/r$ term at $r = 10^6\, m$ was below 10^{-9} in all parameter cases we have studied. These results confirm that they arise from the form of the output formula, rather than the training procedure.

To verify the accuracy of the classical solver, we have compared the results obtained using it to those obtained using the small parameter mass expansion. At $P = 0.05\, m$, it gives $M_{\mathrm{ADM}} - m = 1.5613 \times 10^{-3}$, compared with 1.5625×10^{-3} from Proposition 1. For a spin of $S = 0.05\, m^2$, the associated values are 9.980×10^{-4} and 1.0000×10^{-3}. The differences of 1.2×10^{-6} and 2.0×10^{-6} are consistent with quartic order corrections. The differences are much larger than the estimated Richardson errors of 2×10^{-11} and 2×10^{-14}. This is as expected.

In Fig. 3, we compare the regular conformal-factor correction with the reference solution (boosted, spinning and combined scenarios) at the symmetry axis and equatorial plane, where the differences between the two are shown. The difference is typically of the order 10^{-3} when the correction is of the order 10^{-1}. The largest differences occur within the horizon where the diverging puncture term greatly reduces the effect of the source term $\tilde{A}_{ij}\tilde{A}^{ij}\psi^{-7}$. As a result, in this region, the interior solution is mainly controlled by the outer solution. The spatial distribution of the Hamiltonian residual is shown in Fig. 4, alongside the source for the combined scenario. It is clear that, in general, the residual is consistently smaller than the source by at least two orders of magnitude, except in the vicinity of the puncture.

One can examine the mass accuracy independently of the local conformal-factor profiles. As can be seen in Fig. 5, the ADM mass is computed using three different methods, namely the explicit mass head, Gauss-law volume integral and the classical solver, which is used as the reference.

Here it is important to note that, since the mass head is fitted jointly in both momentum and spin, accuracy along the individual coordinate axes does not guarantee accuracy across the wider parameter space. In order to thoroughly examine the characteristics of the system, we have analysed an eleven by nine grid, containing ninety-nine configurations as shown in Fig. 6. The median mass error was $3.5 \times 10^{-4}\, m$, with a 95th percentile of $8.6 \times 10^{-4}\, m$. The largest error, $3.0 \times 10^{-3}\, m$, occurred at $P = m$ and $S = 0.8\, m^2$. In relative terms the median was 2.9×10^{-4} and the maximum 1.9×10^{-3}. Ninety-six out of the tested cases had a mass error less than $10^{-3}\, m$, and all three exceptions are located on the high momentum edge. The supremum error in ψ had a median of 9.4×10^{-4} and maximum of 3.5×10^{-3}. In general, the accuracy of this approach is substantially better than one part in a thousand, though there are exceptions. As a caveat, this description should not be applied to the worst cases. From the above results, for each of the ninety-nine grid points checked, we observe that the computed mass bounds and the pointwise bounds are always greater than the measured errors. The median ratios of bound to error were 7.0 and 3.8. The lowest ratios recorded were 1.04 and 1.19.

The problem is axially symmetric. The main network is not constrained to have this symmetry, however. We undertook measurements in sixty-four directions for a range of radii (from $0.1\, m$ to $10\, m$) to study the variation of the regular correction in the main network with azimuth. Our results show that the standard deviation relative to the mean was

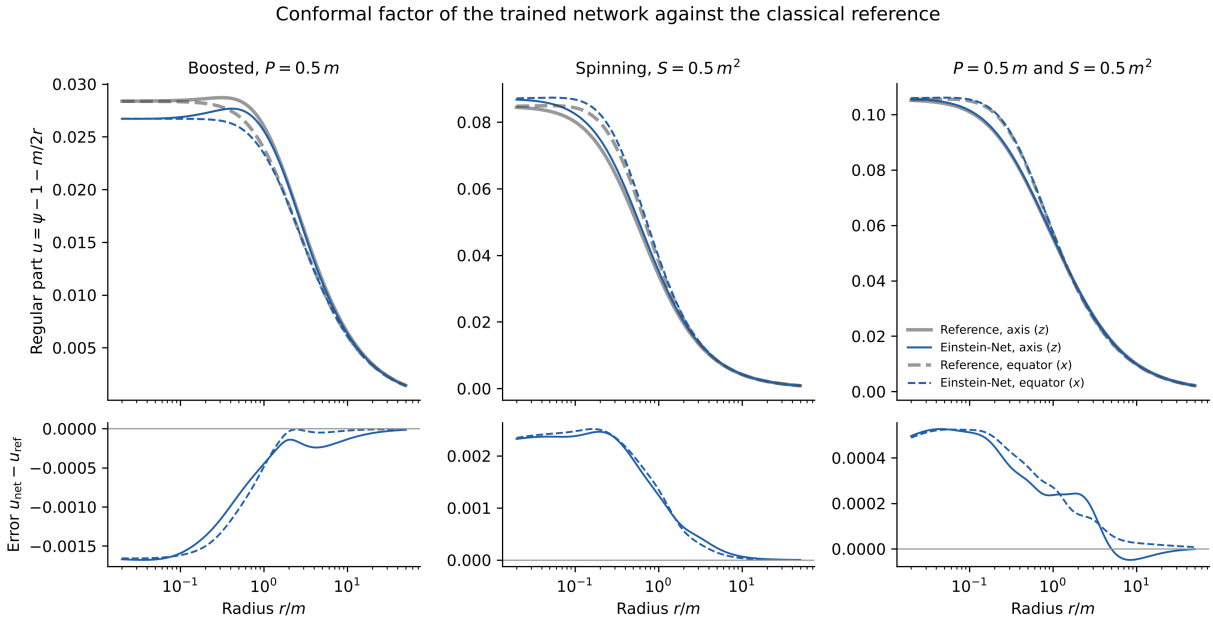


Figure 3: The regular part $u = \psi - 1 - m/2r$ of the conformal factor from the trained network against the classical reference, along the axis and in the equatorial plane, for a boosted hole, a spinning hole and a hole that is both. The lower panels show the difference. The largest errors sit inside the horizon, at $r < m/2$, where the source of the constraint is weakest, and the solution is fixed only by matching to the exterior.

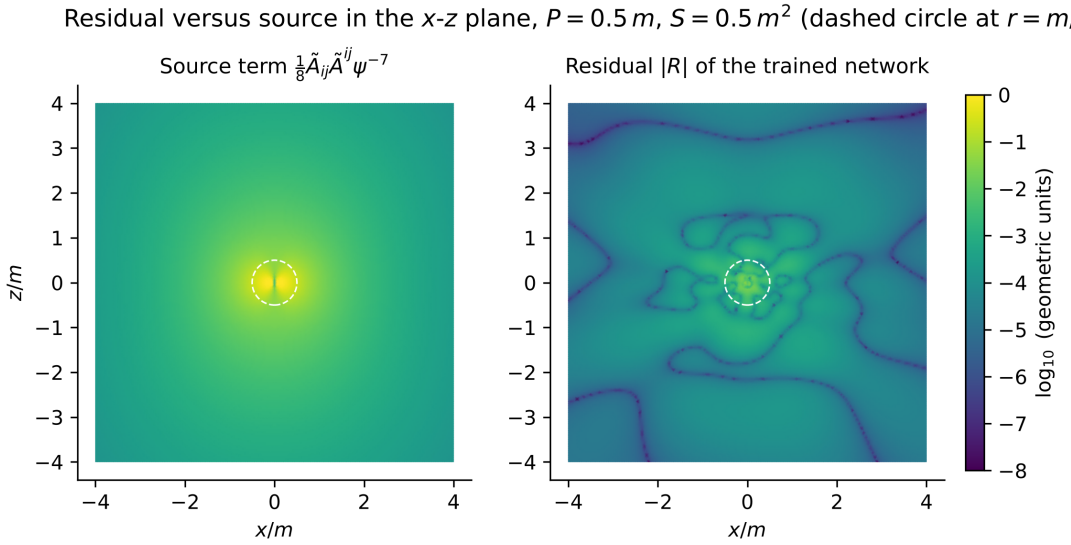


Figure 4: The source term of the Hamiltonian constraint and the residual of the trained network in the x–z plane through the puncture, for $P = 0.5\,m$ and $S = 0.5\,m^2$, on a logarithmic scale. The dashed circle is $r = m/2$.

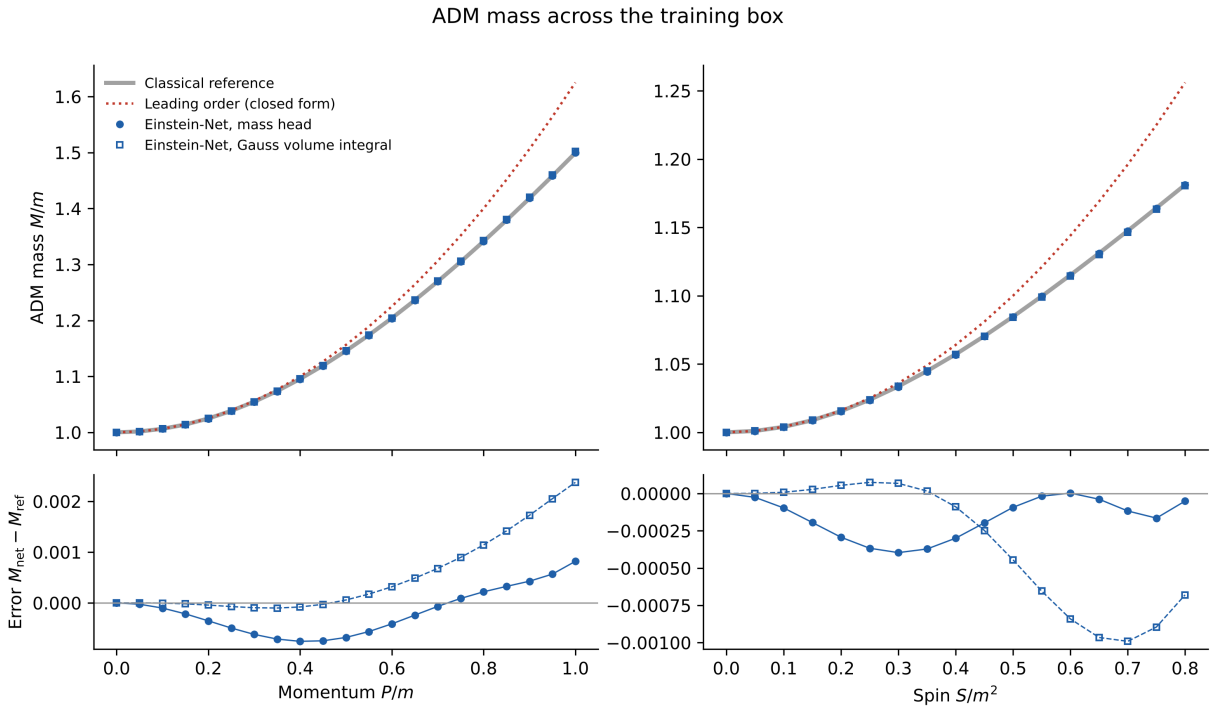


Figure 5: The ADM mass across the training box. The classical reference, the leading-order closed form of Proposition 1, and the network's mass read from the head and from the Gauss volume integral of its own solution. The lower panels show the error against the reference.

between 3×10^{-4} and 2×10^{-3}. This variation is comparable to the overall network error, which suggests that part of the network is dedicated to approximately reproducing a symmetry already available analytically.

In Table 1, we have compared changes made to the mass-fitting procedure and the input representation. Given the variation in the choice of the fitting interval, the results show only slight variations, and so the difference is not considerable. Even though a slightly smaller residual loss is achieved, we observe that by updating both the mass head and shape field using the residual, there is a poor result with $1.9 \times 10^{-2}\, m$ and $0.15\, m$ as the median and maximum mass error, respectively. It is crucial to emphasise the difference between minimising the residual error in the numerical solution of a differential equation and the accuracy of a physical parameter, e.g. mass. The minimisation of the residual of the differential equation is the first step in estimating the mass, but this does not ensure an accurate mass determination alone.

We also constructed an axisymmetric version of the shape field. We can introduce a coordinate system where the axial component ξ_z and the magnitude of the transverse component $|\xi_\perp|$ can be used as surrogates for the Cartesian input vector. Interestingly, the anisotropy of the system is about 10^{-15}. Nevertheless, the model was not more accurate than the main model when both were trained with the same computational budget. With full training, the residual loss was 4.2×10^{-7} compared with 2.8×10^{-7} for the main model. Its median and maximum mass errors were $4.9 \times 10^{-4}\, m$ and $4.8 \times 10^{-3}\, m$ respectively. In the case of the main model, the corresponding mass errors of $3.5 \times 10^{-4}\, m$ and $3.0 \times 10^{-3}\, m$ show that the main model has better accuracy. The results for the supremum error of the function ψ were mixed. One possible cause is the non-smooth nature of $|\xi_\perp|$ on the symmetry axis. Removing the symmetry error did not necessarily improve the overall approximation accuracy, as seen under the specific encoding and optimisation procedure used.

6.2 Black holes and their horizons

Here we show six different solutions that we have obtained using the Einstein-Net framework, Fig. 7. The solutions we consider are accompanied by two distinct visualisations - the apparent horizon and the contours of $K_{ij}K^{ij}$, the latter showing the extrinsic curvature patterns dictated by the imposed linear momentum and spin. A notable feature of boosted black holes is the displacement of the apparent horizon in the direction of motion. For example, at a momentum of $P = 0.5\, m$, the radius of the apparent horizon is calculated to be about $0.460\, m$ in the direction of motion and

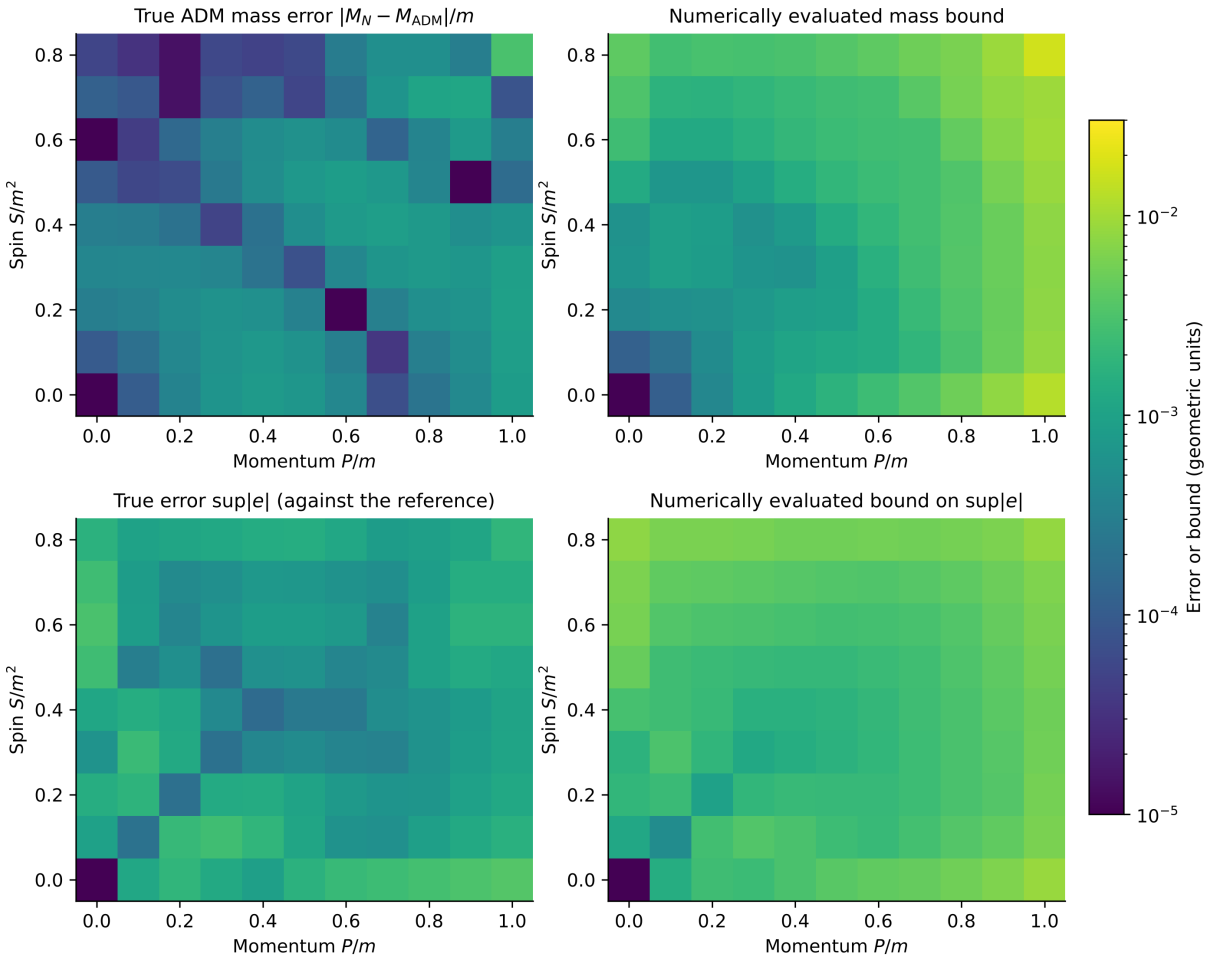


Figure 6: The error and its bound over the training box, on a grid of eleven momenta and nine spins. Top row, the true ADM mass error against the reference and the numerically evaluated mass bound of Eq. (14). Bottom row, the true supremum error in ψ and the numerically evaluated bound of Eq. (13). All four panels share the colour scale.

Table 1: Ablations at an equal budget of fifteen hundred optimiser steps from the same seed, and the symmetric variant at the full budget.

Scheme	residual loss	median $\lvert\Delta M\rvert/m$	max $\lvert\Delta M\rvert/m$	anisotropy
Alternating, refit every 50 steps	4.0×10^{-5}	1.6×10^{-3}	1.8×10^{-2}	2.1×10^{-3}
Alternating, refit every 200 steps	5.1×10^{-5}	1.2×10^{-3}	1.1×10^{-2}	2.8×10^{-3}
Alternating, refit every 500 steps	5.5×10^{-5}	2.4×10^{-3}	1.4×10^{-2}	3.2×10^{-3}
Joint gradient training of the head	2.3×10^{-5}	1.9×10^{-2}	1.5×10^{-1}	1.5×10^{-3}
Axisymmetric field, refit every 50 steps	2.2×10^{-4}	1.1×10^{-3}	2.0×10^{-2}	7×10^{-16}
Main model, full budget	2.8×10^{-7}	3.5×10^{-4}	3.0×10^{-3}	1.9×10^{-3}
Axisymmetric field, full budget	4.2×10^{-7}	4.9×10^{-4}	4.8×10^{-3}	$< 10^{-14}$

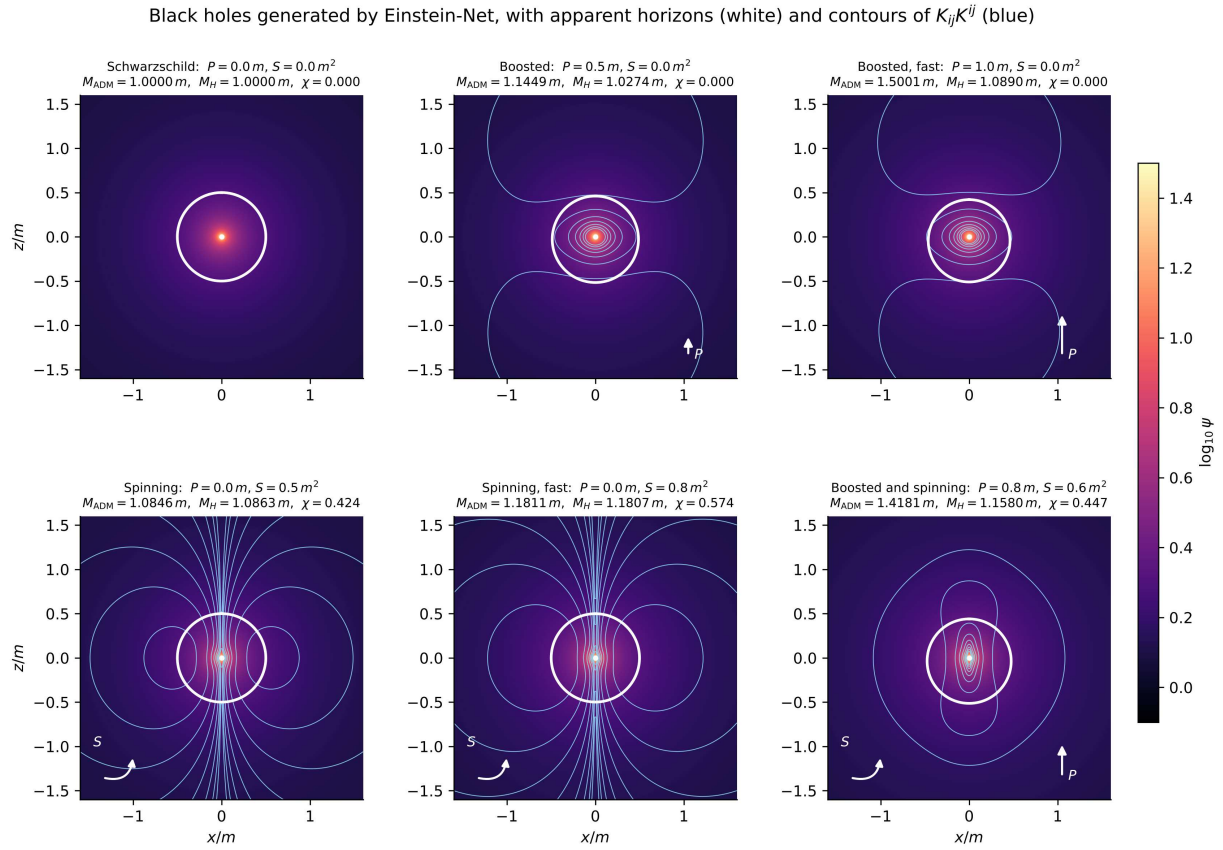


Figure 7: Black holes generated by Einstein-Net. Each panel shows the conformal factor in the x–z plane on a logarithmic scale, contours of $K_{ij}K^{ij}$ in blue, the puncture as a dot, and the apparent horizon found in the data in white, with the ADM mass, the horizon mass and the dimensionless spin above.

Table 2: Invariants of the black holes of Figure 7, all from the network's own data. The horizon radius is given in the direction of motion, in the equatorial plane and behind, and the last column is the largest value of the expansion on the final surface.

Case	P/m	S/m^2	$M_{\rm ADM}/m$	$M_{\rm irr}/m$	M_H/m	χ	$h(0)/m$	$h(\pi/2)/m$	$h(\pi)/m$	$\max\lvert\Theta\rvert$
Schwarzschild	0	0	1.0000	1.0000	1.0000	0	0.5000	0.5000	0.5000	7×10^{-16}
Boosted	0.5	0	1.1449	1.0274	1.0274	0	0.4602	0.4865	0.5165	3×10^{-7}
Boosted, fast	1.0	0	1.5001	1.0890	1.0890	0	0.4198	0.4599	0.5103	8×10^{-6}
Spinning	0	0.5	1.0846	1.0604	1.0863	0.424	0.4999	0.4998	0.4996	2×10^{-6}
Spinning, fast	0	0.8	1.1811	1.1260	1.1807	0.574	0.4993	0.4991	0.4989	5×10^{-6}
Boosted, spinning	0.8	0.6	1.4181	1.1270	1.1580	0.447	0.4403	0.4737	0.5138	4×10^{-6}

$0.517\,m$ in the opposite direction. At a momentum of $P = m$, we note that the radius varies between approximately $0.420\,m$ in the direction of motion and $0.510\,m$ in the opposite direction. The spinning configurations produce the expected quadrupole curvature pattern and maintain a horizon near to $r = m/2$, due to the inversion symmetry in the spin component of the Bowen–York data used here [5, 34]. Our results show that the expansion on the fitted surface is small, with magnitudes less than 10^{-5} in all cases we have discussed. The computed horizon coefficients and irreducible masses are in agreement with the reference values, with differences of about 10^{-3}. The results are summarised in Table 2.

The calculated horizon values are used in two main physical checks. First, the mass of a moving black hole should approximately satisfy $M_{\rm ADM}^2 \simeq M_H^2 + P^2$, where the deviation from this relationship is partly due to the gravitational radiation contained in Bowen–York initial data [34, 35]. A comparison of the network and reference solutions is

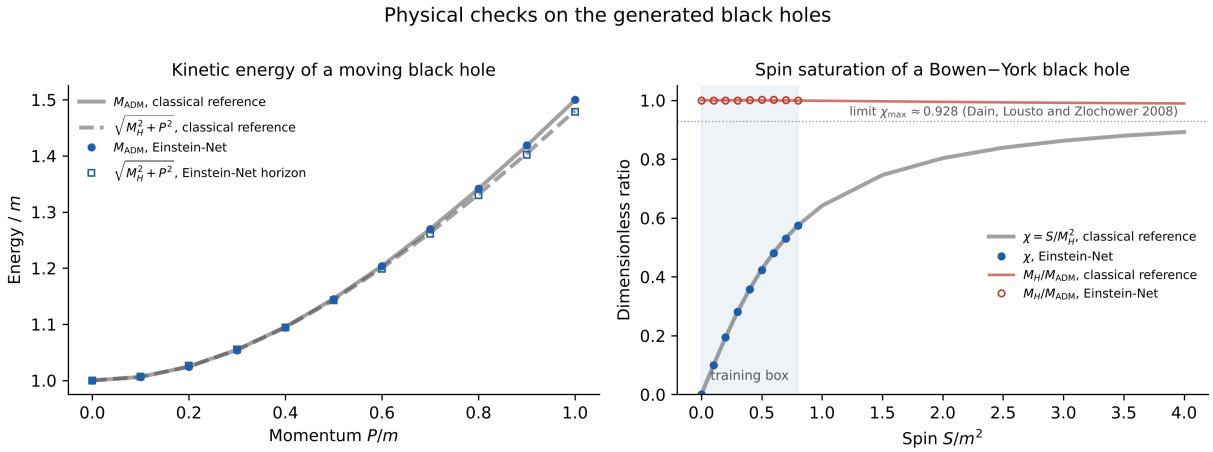


Figure 8: Two physical checks. Left, the kinetic energy of a moving hole, with M_{ADM} and $\sqrt{M_H^2 + P^2}$ from the network's own horizon against the same quantities from the reference. Right, the spin saturation of a spinning hole, with $\chi = S/M_H^2$ and M_H/M_{ADM} from the network inside its training box (shaded) against the reference, which is continued beyond the box towards the known limit.

depicted in Figure 8, where we show the relevant quantities. At $P = m$, the excess relative to M_{ADM} was 1.44 per cent for Einstein-Net and 1.25 per cent for the reference. The difference is comparable with the mass errors discussed above.

The second check concerns the spin of Bowen–York punctures. Increasing the prescribed spin does not result in an arbitrarily large dimensionless horizon spin. The quantity $\chi = S/M_H^2$ approaches a limit below unity, around 0.93 [34, 36, 37]. The Einstein-Net values at all points within the training box follow the reference curve very closely, with a maximum deviation of 2.1×10^{-3}. At a spin of $S = 4\,m^2$, to which the reference solution was continued, χ reaches 0.892 and approaches the known limiting behaviour.

6.3 The bound and the gates inside and outside the family

The first acceptance check was examined against seven representative reference solutions (see Table 3). The numerically derived bounds on the supremum of the absolute error ($\sup |e|$) in all cases are higher than the measured supremum errors, and the ratios are between 1.3 and 4.6. Similarly, the computed mass bounds are higher than the actual mass errors, and the ratios are between 1.1 and about 80. It is also noteworthy that in some of the spinning configurations, the large ratios are mainly due to the very low mass errors of the system. Figure 9 shows these comparisons. The aim of this approach is to obtain an upper bound without having to compute a new reference solution for every single query. Note that this estimate is not expected to coincide closely with the actual error in every case. The results for the full ninety-nine-point grid were described earlier.

In this case, we have tested the numerical evaluation by comparing quadrature schemes using 3840, $30,720$ and $103,680$ nodes (see Table 5). The pointwise error bound varies by up to 0.27 per cent between the default and the finest quadrature rules across all seven cases studied. Similarly, the mass error bound has varied by up to 0.8 per cent in these scenarios. In addition, the least refined rule appeared to be less reliable, with up to a 40 per cent change in the mass bound, and therefore we did not use it to derive the main results.

We tried to quantify the effects of the radial regions which were not taken into consideration in our quadrature scheme. As can be observed near the singularity (puncture), the residual behaves approximately as c/r, where c is a constant coefficient, and r is the radial coordinate. This behaviour of the measured residual corresponds to an integrable singularity associated with the compactified radial coordinate. In particular, if $|R| \leq c/r$, for $r < \epsilon$, the contribution of this region to the Newton potential at its maximum can be upper bounded by $c\epsilon$. Estimating the values of c between $10^{-7}\,m$ and $10^{-4}\,m$ for the parameter settings $P = S = 0.5$ gives 0.037 and for the parameter-box corner gives 0.074. Thus, the contributions would be 3.7×10^{-6} and 7.4×10^{-6} respectively, which are relatively small compared to the evaluated bounds. As one reaches the far field, the asymptotic behaviour of the residual takes the form of r^{-4}. The measured coefficients between $10^5\,m$ and $10^7\,m$ were 0.13 and 0.69. Given the above, if such decay persists, the unaccounted contribution beyond a radius of $10^5\,m$ would be below 10^{-10}. The conditional estimates described above are based on the sampled residual rather than a full and rigorous global enclosure.

Table 3: The numerically evaluated bound of Proposition 2 against the true error on seven holes, at the default quadrature. The true supremum of $|e|$ is measured against the reference on the quadrature nodes with $r < 50\,m$, and the true mass error against the Richardson-extrapolated reference mass. The last column is the L^2 norm of the residual.

P/m	S/m^2	evaluated bound on $\sup|e|$	true $\sup|e|$	evaluated bound on $\lvert\Delta M\rvert/m$	true $\lvert\Delta M\rvert/m$	$\lvert R\rvert_{L^2}$
0.2	0	2.49×10^{-3}	1.91×10^{-3}	4.05×10^{-4}	3.54×10^{-4}	6.75×10^{-3}
0.5	0	3.72×10^{-3}	1.68×10^{-3}	1.70×10^{-3}	6.81×10^{-4}	9.87×10^{-3}
0	0.5	4.59×10^{-3}	2.52×10^{-3}	1.31×10^{-3}	9.29×10^{-5}	1.53×10^{-2}
0.5	0.5	2.10×10^{-3}	5.74×10^{-4}	1.92×10^{-3}	7.24×10^{-4}	7.41×10^{-3}
1.0	0	8.63×10^{-3}	3.53×10^{-3}	1.23×10^{-2}	8.18×10^{-4}	1.73×10^{-2}
0	0.8	7.63×10^{-3}	1.67×10^{-3}	4.12×10^{-3}	5.10×10^{-5}	2.92×10^{-2}
1.0	0.8	8.08×10^{-3}	1.89×10^{-3}	1.66×10^{-2}	3.02×10^{-3}	2.11×10^{-2}

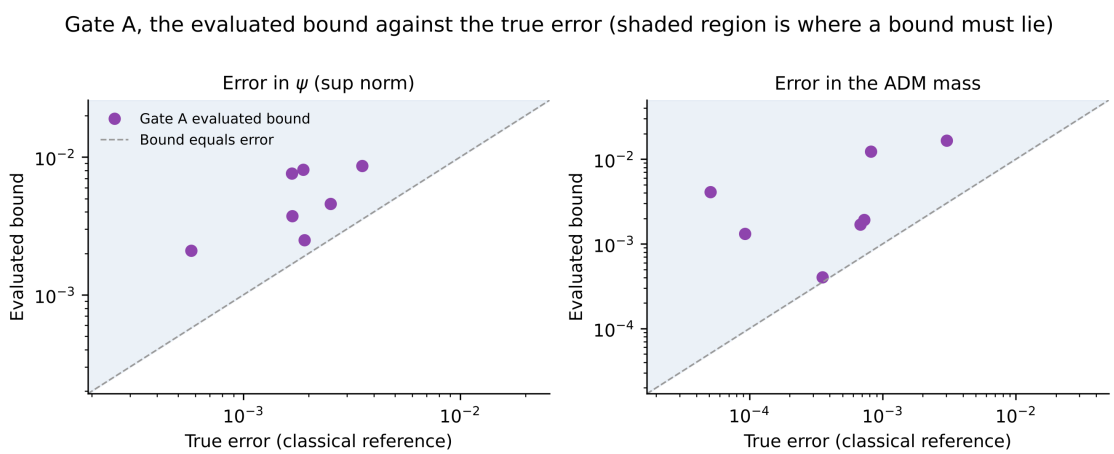


Figure 9: The numerically evaluated bound against the true error for the cases of Table 3, for the error in ψ and for the error in the ADM mass. A valid bound lies in the shaded region above the diagonal.

To verify the validity of the inequality $\psi_N > \mathcal{N}$ and the magnitude of the Newton potential, we have used a dataset of twenty thousand independent points where the radii are log-uniformly distributed. The smallest measured values of $\psi_N - \mathcal{N}$ were greater than unity for both experimental arrangements. The largest potentials on these points agreed with the corresponding quadrature-node maxima to four significant figures, with 2.0954×10^{-3} versus 2.0952×10^{-3} for the first configuration and 8.0829×10^{-3} versus 8.0826×10^{-3} for the second. From the aggregate data obtained from the diagnostic tests, the numerical evaluation errors in tested cases are likely to be one per cent or less. It is crucial to highlight that the verification procedures discussed here do not convert the obtained estimates into rigorous numerical bounds. All seven different configurations passed the second gate. The relevant quartic coefficients used in this check, -0.192 and -0.311, were fitted from the reference mass sweeps.

To select the statistical acceptance rule, we first generated forty different parameter pairs based on a uniform distribution. The mass-error tolerance was $\tau = 2 \times 10^{-3}\,m$, with risk target $\delta = 0.15$ and confidence level 95 per cent. The median mass error was $3.9 \times 10^{-4}\,m$ and the maximum error was $1.1 \times 10^{-3}\,m$. As can be seen from all forty samples, the proxy values were all below the defined threshold of $9.1 \times 10^{-3}\,m$. All cases were within the error tolerance. The Clopper–Pearson upper limit of 0.072 calculated from those same cases was used in selecting the threshold, and it cannot independently validate that selection.

For validation, we then generated an additional forty test cases using a different random number seed, while keeping the decision threshold frozen. Thirty-eight passed. From the successful cases, we found that none of the resulting solutions exceeded the specified mass-error tolerance. Furthermore, the forty instances gave a median mass error of $4.2 \times 10^{-4}\,m$ and a maximum of $9.0 \times 10^{-4}\,m$. The fixed rule had zero failures among the thirty-eight accepted independent cases. Hence, the one-sided Clopper–Pearson confidence limit for this rule at 95 per cent confidence level is given by $U(0, 38) = 0.076$. Thus, under the same parameter distribution, the conditional probability that a future accepted output will exceed the given mass tolerance has an upper confidence limit of 0.076. The two rejected cases were nevertheless accurate, with proxy values up to $1.05 \times 10^{-2}\,m$. This illustrates the conservative behaviour of the threshold. Figure 10 shows the calibration observations.

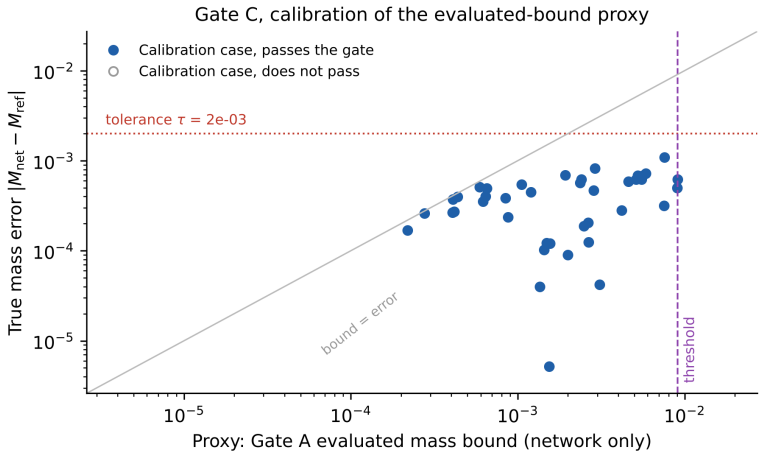


Figure 10: Calibration of the third gate. Each point is a calibration case, with the numerically evaluated mass bound of Proposition 2 as the proxy on the horizontal axis and the true mass error on the vertical axis. The dotted line is the tolerance, the dashed line the calibrated threshold, and the diagonal marks where the bound would equal the error.

Table 4: The bound outside the training region. The network is required to provide the spinning holes with S up to $3\,m^2$. Training was only done up to $0.8\,m^2$. The true error is measured against the reference, and the evaluated bound is Eq. (14). In the case that the sufficient condition $\psi_N > \mathcal{N}$ fails, no bound is returned.

S/m^2	M_N/m	M_{ADM}/m (reference)	true $\lvert\Delta M\rvert/m$	evaluated bound on $\lvert\Delta M\rvert/m$	in box
0.4	1.0569	1.0572	3.0×10^{-4}	6.0×10^{-4}	yes
0.8	1.1811	1.1812	5.1×10^{-5}	4.1×10^{-3}	yes
1.0	1.2582	1.2504	7.8×10^{-3}	2.0×10^{-2}	no
1.5	1.8273	1.4230	0.40	1.3	no
2.0	5.1258	1.5863	3.5	no bound	no
3.0	56.886	1.8807	55	no bound	no

Extrapolation gives a further test because the statistical rule was calibrated only for $S \leq 0.8\,m^2$. The architecture can still return values for larger spins, but they need not be accurate. From the results presented in Table 4, it is clear that when the spin parameter is set to $S = m^2$, the mass error increases to $7.8 \times 10^{-3}\,m$. The evaluated bound in this case is $2.0 \times 10^{-2}\,m$. At a spin of $S = 1.5\,m^2$, the error is approximately $0.40\,m$. The associated error bound increases to $1.3\,m$. For $S = 2\,m^2$ and $3\,m^2$, the predicted masses are far from the reference values, and the sufficient condition $\psi_N > \mathcal{N}$ fails. In cases where the above criterion is not satisfied, no mass bounds can be derived. This failure can be seen as an appropriate warning that the predictions made by the model cannot be supported by the present estimate. Interestingly, the residual based check does not explicitly depend on the boundaries of the training box, and its numerical value increases as the prediction deteriorates.

6.4 Cost and practical use

The output from the model is the initial-data fields along with the relevant physical parameters, and validity checks. Furthermore, the metric and extrinsic curvature are computed on the specified grid, along with the numerical bound supplied by the residual. The other two checks assess the calculated mass. The archive contains the metric tensor γ_{ij} and the extrinsic curvature tensor K_{ij}, along with the grid and parameters used, conventions and acceptance information. In the specific case where the physical parameters are $P = 0.4\,m$ and $S = 0.3\,m^2$, the Einstein-Net algorithm gives a calculated mass of $M_{\mathrm{ADM}} = 1.12460\,m$, compared to the reference value of $1.12481\,m$. The evaluated mass bound is $5.8 \times 10^{-4}\,m$, and all three validity checks accept the result.

The computational cost depends strongly on whether the error estimate and the horizon are required. For example, when the problem is solved on a 64^3 grid, it takes about 0.30 seconds to compute the conformal factor using two CPU cores. Similarly, the construction of the metric γ_{ij} and extrinsic curvature tensor components K_{ij} can be done in a relatively shorter time of 0.14 seconds. The horizon finder takes around 0.80 seconds to run, whereas the bound evaluation takes around 3.4 seconds. Thus, the initial data took less than half a second to compute, while the total, including the horizon

and the evaluated bound, took about 4.6 seconds. As part of our study, we tested fifty different random configurations and ran the three diagnostics on each of them, taking a total of 289 seconds. The lowest evaluated mass bound was $4 \times 10^{-4}\, m$ and the highest $8.6 \times 10^{-3}\, m$.

Comparison with the conventional solvers depends on the specific formulation being used. For example, for axisymmetric problems, the reference solver takes advantage of the symmetry of the problem and quickly generates an accurate grid in approximately 1.5 seconds with an accuracy of order 10^{-8}. It is consequently faster and considerably more accurate than the network. However, there are a few key advantages of Einstein-Net to note in this comparison. Firstly, Einstein-Net models an entire family of parameters through a single model; secondly, its output depends differentiably on those parameters, and thirdly, it provides a residual based assessment.

In the case of the three-dimensional finite-difference (FD) method that does not take into account the axisymmetric nature of the problem, we have performed a study of its computational efficiency. Using a 64^3 grid and a cube with a half-width of $8\, m$, the computational time is 12 seconds, and the maximum error is 6.6×10^{-4} in the ψ variable, which is worse than the 4.0×10^{-4} obtained using the network. Even after we refined the grid and kept the domain dimensions unchanged, the error remains roughly stable at 6.5×10^{-4}. Therefore, we undertook further investigation to understand the effects of the resolution and domain size on the solution accuracy. Increasing the half-width to $16\, m$ on a 64^3 grid, the error in the region $r < 8\, m$ increases to 1.0×10^{-3}, because the grid is less dense. Furthermore, increasing the half-width to $32\, m$ on 128^3 cells but keeping a consistent spacing, the error decreases to 9.4×10^{-4}.

Here, it is clear that the boundary no longer dominates the error. A 128^3 solve on the $16\, m$ domain takes 98 seconds and the error is 7.0×10^{-4}. From the above results, we note that the behaviour near the puncture in the maximum norm converges at approximately first order. Based on these results, one would need to reduce the spacing to roughly $0.12\, m$, or around 256^3 cells on the $16\, m$ domain, to achieve the network's 4.0×10^{-4} error. However, due to a lack of memory (seven gigabytes), we did not perform the computation, and the estimated time of ten minutes or more for the calculation is an extrapolation. This comparison is specifically applicable to our second-order implementation but not the best general-purpose numerical relativity codes. The network generated its own grid faster and with smaller error when compared to the three-dimensional cases we were able to run. Additionally, the use of automatic differentiation through the trained network may prove to be useful in optimising the initial data parameters without having to differentiate through an iterative solve.

7 Discussion

We have applied the method to a single puncture with parallel momentum and spin. The numerical results from our simulations are very consistent with those obtained using the reference solver. Furthermore, we are also able to recover the Schwarzschild solution, recover the apparent horizons and pass both imposed physical checks. Note, though, that experiments that lie outside the parameter range tested provide only limited information. In particular, we note that while residual based analysis can give some limited information, this analysis may fail to give any bound at all, especially in cases where the network extrapolates poorly.

Compared to the existing neural approaches to Einstein's equations [13, 14, 15, 16], our method relates to a narrower physical problem. This is because the results discussed here are only related to the problem of generating the initial data for the puncture.

For the puncture data, we use the hard trial functions and comparison principles. Thus, a specific formulation has to be derived for the physical system involved. This is an obvious limitation, but at the same time it is a vital part of the approach used here. This makes the approach different from existing neural network architectures. The York decomposition and the Brandt–Brügmann puncture representation provide a structure which a generic neural network architecture would otherwise have to learn.

Note that the residual based bound produced in this work may have a wider applicability than just for the black hole initial data considered here. The bound is based on the fact that the derivative of the nonlinear source term with respect to the unknown has a definite sign and hence similar lines of reasoning can be extended to other semi-linear elliptic equations with decreasing nonlinearities subject to appropriate boundary conditions where such comparison principles still hold.

The estimates thus obtained are independent of the methods used to obtain the approximate solutions. Any general numerical solver can use the same techniques described above to compute the residuals. However, one must check the accuracy of the numerical integration in practice. This can be done by forming an interval around the quadrature cells and the neglected radii using interval arithmetic, to give a rigorous bound on the computed quantities. Similarly, a fixed-rule statistical test can be deployed to control the mass-error risk within a given population of parameters. This

statistical verification can also help one decide if the prediction is good enough to be used in an evolution code or if one must resort to a classical solve.

There are some limitations in this work. Currently, the bound controls the solution ψ and the ADM mass, but this does not include the gradient. One would need the gradient to derive an error bound on the horizon location or the horizon mass. This could be achieved by extending the approach and using the derivatives of the Newton potential. However, this has not been established here. Also, the neural network architecture we used does not come with an inherent axisymmetry. Applying symmetric ablation resulted in the disappearance of measurable azimuthal variation, but this did not improve the accuracy with the encoding we used. More suitable equivariant models that can handle general momenta and spins may be of use here, as well as diagnostic tools for identifying residual symmetry breaking such as those discussed in our work on quantum networks [23].

Following the results discussed above, an interesting question would be to generalise these results to a binary system with two punctures, but this time with both punctures having their own momenta and spins. The linearity of the Bowen–York momentum solution means that such a superposition would be possible, with the conformal factor requiring an additional singular term and a corresponding regular correction. The static Brill–Lindquist solution [38] and the spectral solver of Ansorg et al. [9] can be utilised as useful references. Moreover, identifying a common apparent horizon for a binary system is a challenging task. Moving beyond conformal flatness, or introducing matter and a cosmological constant on constant-mean-curvature slices [22], can change the governing constraint problem drastically. Such complex scenarios would allow one to test the validity of the residual based arguments in different settings.

8 Conclusions

In this paper, we have introduced the Einstein-Net, a neural-network framework for constructing constraint-compatible initial data for simulating single black holes with linear momentum and spin. For this purpose, we combine the Bowen–York formulation with an analytically prescribed conformal-factor representation. The method we propose here satisfies the momentum constraint, reproduces the puncture structure, preserves positivity and asymptotic flatness, and recovers the Schwarzschild limit independently of the learning process. The neural network is therefore required to only approximate the remaining Hamiltonian constraint. Furthermore, the ADM mass is represented explicitly and determined through a Gauss-law formulation.

We show that the methods provide accurate results for the Hamiltonian residual alone. Furthermore, the ablation study discussed shows that a lower residual can be accompanied by significantly larger mass errors when the ADM mass is not treated separately. This indicates the importance of incorporating physical constraints and conservation identities directly into the learning framework instead of relying exclusively on residual minimisation. Across the range of parameters considered, the proposed approach shows that the mass estimates with median relative errors of only a few parts in 10^4 can be achieved, and these remain in close agreement with independent classical solutions.

The framework we present here thus can be used for assessing the quality of solutions. Thus, an argument relating to the comparison-principle argument yields an analytical error estimate which is derived from the Hamiltonian residual. Additional acceptance checks can provide complementary numerical and statistical evidence of reliability. Within the training regime, our assessments were generally consistent with the accuracy required. Beyond that regime, some deterioration in performance was noticed - this was mainly in the residual-based estimates and the acceptance criteria, indicating that the proposed diagnostics still remain informative even in such cases.

Computationally speaking, the solution can be generated efficiently, and most of the computational cost comes from the assessment procedures rather than field evaluation itself. Future work could be focused on strengthening the validation framework. Further work can also be carried out to generalise the approach to binary black-hole configurations and more complex families of initial data.

Data and code availability

The Einstein-Net library, trained weights, reference tables, and executed notebook are available at `https://github.com/ugail/Einstein-Net`.

A The leading-order mass

To derive the leading-order mass, we insert $\psi_0 = 1 + m/2r$ into the volume form of Eq. (5). The correction u begins at second order in momentum and spin, so its contribution to the integrand starts at fourth order. Using Eq. (3), the

calculation separates into angular and radial integrations. The angular factors are

$$\int_{S^2} \left(1 + 2\cos^2\theta\right) \mathrm{d}\Omega = 4\pi + \frac{8\pi}{3} = \frac{20\pi}{3}, \qquad \int_{S^2} \sin^2\theta \, \mathrm{d}\Omega = \frac{8\pi}{3}.$$

For the radial part, set $s = m/2r$, giving $r = m/2s$ and $\mathrm{d}r = -m\,\mathrm{d}s/(2s^2)$. The momentum contribution becomes

$$\int_0^\infty \frac{r^2\,\mathrm{d}r}{r^4}\Big(1 + \frac{m}{2r}\Big)^{-7} = \frac{2}{m}\int_0^\infty (1+s)^{-7}\,\mathrm{d}s = \frac{2}{m}\cdot\frac{1}{6} = \frac{1}{3m},$$

Similarly, the spin contribution is

$$\int_0^\infty \frac{r^2\,\mathrm{d}r}{r^6}\Big(1 + \frac{m}{2r}\Big)^{-7} = \frac{8}{m^3}\int_0^\infty s^2(1+s)^{-7}\,\mathrm{d}s = \frac{8}{m^3}\cdot\frac{1}{60} = \frac{2}{15m^3},$$

Here $\int_0^\infty s^2(1+s)^{-7}\mathrm{d}s = B(3,4) = 2!, 3!/6! = 1/60$. Combining the angular and radial factors gives

$$M_{\mathrm{ADM}} - m = \frac{1}{16\pi}\Big[\frac{9P^2}{2}\cdot\frac{20\pi}{3}\cdot\frac{1}{3m} + 18S^2\cdot\frac{8\pi}{3}\cdot\frac{2}{15m^3}\Big] = \frac{5P^2}{8m} + \frac{2S^2}{5m^3},$$

This establishes Proposition 1. It is clear from Eq. (3) that there is no cross term at second order. The corrections at the fourth order include the effect of the variable u within the integral. The quartic terms used for the empirical mass check were fitted from the reference solver as $-0.192\,P^4/m^3$ and $-0.311\,S^4/m^7$.

B Proof of the analytically certified bound

Let ψ satisfy the exact constraint in Eq. (1), with $\psi - 1 - m/2r \to 0$ at infinity, and let ψ_N be the approximation of Proposition 2. Both are positive. Applying the mean value theorem to t^{-7} gives $\psi_N^{-7} - \psi^{-7} = -7\bar{\psi}^{-8}e$, where $e = \psi_N - \psi$ and $\bar{\psi}(x)$ lies between the two conformal factors. Thus e satisfies Eq. (12), with $\bar{V} = -\frac{7}{8}\tilde{A}_{ij}\tilde{A}^{ij}\bar{\psi}^{-8} \le 0$. The singular puncture terms in the expression e cancel each other out, leaving a bounded difference which tends to zero at infinity. The puncture is therefore removable for the error equation.

The Newton potential of $|R|$ is denoted by the symbol $\mathcal{N}$. It is nonnegative, tends to zero at infinity and satisfies $\Delta\mathcal{N} = -|R|$. For $w = \mathcal{N} - e$, we obtain

$$\Delta w + \bar{V}w = -|R| - R + \bar{V}\mathcal{N} \le 0,$$

The right-hand side is nonpositive since $-|R| - R \le 0$ and $\bar{V}\mathcal{N} \le 0$. Hence w is a supersolution of $L = \Delta + \bar{V}$, with $\bar{V} \le 0$. Applying the weak maximum principle [20] on successively larger balls gives $w \ge 0$, because its boundary values approach zero. Therefore $e \le \mathcal{N}$. Applying the same argument to $-e$ gives $-e \le \mathcal{N}$, which proves Eq. (13).

To compute the mass estimate, we need to integrate the error equation (Eq. (12)) over the specified slice. The flux through the infinitesimal spherical surface around the puncture is zero, as a result of the cancellation of the singular terms and the fact that the error is bounded. At large radius the flux equals -4π times the $1/r$ coefficient of e, namely $(M_N - M_{\mathrm{ADM}})/2$. Thus

$$-2\pi\,(M_N - M_{\mathrm{ADM}}) = \int \big(R - \bar{V}e\big)\,\mathrm{d}V,$$

Taking absolute values gives $|M_N - M_{\mathrm{ADM}}| \le (2\pi)^{-1}(|\int R\,\mathrm{d}V| + \int |\bar{V}|\,|e|\,\mathrm{d}V)$. From the pointwise estimate, $\psi \ge \psi_N - \mathcal{N}$, and consequently $\bar{\psi} \ge \psi_N - \mathcal{N}$ wherever the latter is positive. Therefore $|\bar{V}| \le \frac{7}{8}\tilde{A}_{ij}\tilde{A}^{ij}(\psi_N - \mathcal{N})^{-8}$. Combining this with $|e| \le \mathcal{N}$ yields Eq. (14).

C Numerical details

The training procedure involves drawing collocation points at each iteration. Radii of the sampling sites are drawn from a log-uniform distribution on $[0.02\,m, 200\,m]$ and directions are uniformly distributed on a sphere. Momentum and spin of the sampling sites are also drawn uniformly from the specified bounds. Two reverse-mode automatic differentiation passes are used to compute the Laplacian of the assembled conformal factor. For the purpose of the Gauss-law mass integral, we use spherical quadrature with 192 logarithmic radial cells from $10^{-3}\,m$ to $10^4\,m$, sixteen Gauss–Legendre angular nodes in $\cos\theta$ and eight azimuthal nodes. Increasing the counts to 384, 24 and 8 renders the integral mass slightly different, with an absolute difference of less than 3×10^{-4}. The nine coefficients of the mass head are then determined via a 64×9 linear least-squares fit to the integral masses of sixty-four sampled parameter pairs.

Table 5: The evaluated bounds of Proposition 2 at three quadrature resolutions. The rules have 120, 240 and 360 log-spaced radial cells on $10^{-4}\,m < r < 10^{5}\,m$ with 8, 16 and 24 Gauss–Legendre nodes in $\cos\theta$ and 4, 8 and 12 azimuthal nodes, that is $3,840$, $30,720$ and $103,680$ nodes. The default is the middle rule.

		bound on $\sup\lvert e\rvert$			bound on $\lvert\Delta M\rvert/m$		
P/m	S/m^2	coarse	default	fine	coarse	default	fine
0.2	0	2.496×10^{-3}	2.490×10^{-3}	2.490×10^{-3}	3.61×10^{-4}	4.05×10^{-4}	4.05×10^{-4}
0.5	0	3.741×10^{-3}	3.722×10^{-3}	3.725×10^{-3}	1.854×10^{-3}	1.705×10^{-3}	1.708×10^{-3}
0	0.5	4.637×10^{-3}	4.594×10^{-3}	4.582×10^{-3}	1.516×10^{-3}	1.313×10^{-3}	1.303×10^{-3}
0.5	0.5	1.993×10^{-3}	2.095×10^{-3}	2.091×10^{-3}	1.104×10^{-3}	1.919×10^{-3}	1.909×10^{-3}
1.0	0	8.714×10^{-3}	8.627×10^{-3}	8.626×10^{-3}	1.250×10^{-2}	1.226×10^{-2}	1.221×10^{-2}
0	0.8	7.877×10^{-3}	7.629×10^{-3}	7.616×10^{-3}	4.31×10^{-3}	4.12×10^{-3}	4.09×10^{-3}
1.0	0.8	8.366×10^{-3}	8.083×10^{-3}	8.062×10^{-3}	1.669×10^{-2}	1.658×10^{-2}	1.650×10^{-2}

The default quadrature used to evaluate the error bound has 240 log-spaced radial cells from $10^{-4}\,m$ to $10^{5}\,m$, together with the same angular rule, giving $30,720$ nodes. Here, for each node in the system, the Newton potential is computed using both the contributions of the other nodes and a separate self-cell term. For a cell of volume w_i, the integral of $|x-x_i|^{-1}$ is no greater than that for a ball of equal volume centred at x_i, namely $2\pi a^2$ with $a=(3w_i/4\pi)^{1/3}$. This value is used for the self-cell contribution. In contrast to the self-cell term, the other cell contributions are conventional quadrature estimates and not validated enclosures. It is crucial to distinguish between the numerical values that appear as bounds in the results and the analytical statement. Table 5 shows the comparison of the three different quadrature resolutions for the cases in Table 3.

The default routine takes around four seconds to compute on two CPU cores; however, the finest quadrature takes about a minute. The bulk of the computation time is spent in calculating the pairwise distances, which are subsequently required for the computation of the Newton potential.

The problem was solved using an axisymmetric reference calculation which includes 10 even Legendre modes and 40 Gauss–Legendre nodes in $\cos\theta$. Within the radial grids, we used a compact coordinate form with 1000, 2000, and 4000 cells. Newton iterations then converged at a residual of 10^{-12}. The order of radial convergence was observed to be approximately 2.00. The Richardson error estimates come from the difference between the extrapolated and finest-grid masses.

References

[1] Arnowitt, R., Deser, S., Misner, C. W. (2008). Republication of: The dynamics of general relativity. *General Relativity and Gravitation*, 40, 1997–2027. https://doi.org/10.1007/s10714-008-0661-1

[2] York, J. W. (1971). Gravitational degrees of freedom and the initial-value problem. *Physical Review Letters*, 26, 1656–1658. https://doi.org/10.1103/PhysRevLett.26.1656

[3] York, J. W. (1972). Role of conformal three-geometry in the dynamics of gravitation. *Physical Review Letters*, 28, 1082–1085. https://doi.org/10.1103/PhysRevLett.28.1082

[4] Cook, G. B. (2000). Initial data for numerical relativity. *Living Reviews in Relativity*, 3, 5. https://doi.org/10.12942/lrr-2000-5

[5] Bowen, J. M., York, J. W. (1980). Time-asymmetric initial data for black holes and black-hole collisions. *Physical Review D*, 21, 2047–2056. https://doi.org/10.1103/PhysRevD.21.2047

[6] Brandt, S., Brügmann, B. (1997). A simple construction of initial data for multiple black holes. *Physical Review Letters*, 78, 3606–3609. https://doi.org/10.1103/PhysRevLett.78.3606

[7] Campanelli, M., Lousto, C. O., Marronetti, P., Zlochower, Y. (2006). Accurate evolutions of orbiting black-hole binaries without excision. *Physical Review Letters*, 96, 111101. https://doi.org/10.1103/PhysRevLett.96.111101

[8] Baker, J. G., Centrella, J., Choi, D.-I., Koppitz, M., van Meter, J. (2006). Gravitational-wave extraction from an inspiraling configuration of merging black holes. *Physical Review Letters*, 96, 111102. https://doi.org/10.1103/PhysRevLett.96.111102

[9] Ansorg, M., Brügmann, B., Tichy, W. (2004). Single-domain spectral method for black hole puncture data. *Physical Review D*, 70, 064011. https://doi.org/10.1103/PhysRevD.70.064011

[10] Lagaris, I. E., Likas, A., Fotiadis, D. I. (1998). Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks*, 9, 987–1000. https://doi.org/10.1109/72.712178

[11] Raissi, M., Perdikaris, P., Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

[12] Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3, 422–440. https://doi.org/10.1038/s42254-021-00314-5

[13] Li, Z.-H., Li, C.-Q., Pang, L.-G. (2023). Solving Einstein equations using deep learning. *arXiv preprint*, arXiv:2309.07397. https://doi.org/10.48550/arXiv.2309.07397

[14] Carretero, J. A., Urbán, J. F., Abalos, F., Massó, J., Reula, O. A., Pons, J. A., Palenzuela, C. (2026). Towards solving General Relativity with Physics-Informed Neural Networks. *arXiv preprint*, arXiv:2609.38331. https://doi.org/10.48550/arXiv.2609.38331

[15] Bayona, E., Quevedo, H. (2026). Solving Einstein's vacuum equations with physics-informed neural networks: Boundary conditions and domain decomposition. *arXiv preprint*, arXiv:2608.08846. https://doi.org/10.48550/arXiv.2608.08846

[16] Shukla, E., Suriyanarayanan, A., Radice, D., Maulik, R., Bernuzzi, S. (2026). Puncture evolution with physics-informed neural networks. *arXiv preprint*, arXiv:2610.06776. https://doi.org/10.48550/arXiv.2610.06776

[17] Tarancón-Álvarez, P., Tejerina-Pérez, P., Jimenez, R., Protopapas, P. (2025). Efficient PINNs via multi-head unimodular regularization of the solutions space. *Communications Physics*, 8, 335. https://doi.org/10.1038/s42005-025-02248-1

[18] Kohli, I. S. (2020). Einstein's field equations as continuous-time recurrent neural networks. *arXiv preprint*, arXiv:2001.02515. https://doi.org/10.48550/arXiv.2001.02515

[19] Lu, L., Pestourie, R., Yao, W., Wang, Z., Verdugo, F., Johnson, S. G. (2021). Physics-informed neural networks with hard constraints for inverse design. *SIAM Journal on Scientific Computing*, 43, B1105–B1132. https://doi.org/10.1137/21M1397908

[20] Gilbarg, D., Trudinger, N. S. (2001). *Elliptic Partial Differential Equations of Second Order*. Springer, Berlin. https://doi.org/10.1007/978-3-642-61798-0

[21] Bates, S., Angelopoulos, A., Lei, L., Malik, J., Jordan, M. (2021). Distribution-free, risk-controlling prediction sets. *Journal of the ACM*, 68, 43. https://doi.org/10.1145/3478535

[22] Ugail, H. (2026). Reliable Weyl diagnostics for an inhomogeneous universe in numerical relativity. *arXiv preprint*, arXiv:2608.06383. https://doi.org/10.48550/arXiv.2608.06383

[23] Ugail, H., Howard, N. (2026). A channel-level diagnostic for symmetry breaking in noisy equivariant quantum neural networks. *IEEE Access*, 14, 95152–95167. https://doi.org/10.1109/ACCESS.2026.3706394

[24] Baumgarte, T. W., Shapiro, S. L. (2010). *Numerical Relativity: Solving Einstein's Equations on the Computer*. Cambridge University Press. https://doi.org/10.1017/CBO9781139193344

[25] Alcubierre, M. (2008). *Introduction to 3+1 Numerical Relativity*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199205677.001.0001

[26] Thornburg, J. (2007). Event and apparent horizon finders for 3+1 numerical relativity. *Living Reviews in Relativity*, 10, 3. https://doi.org/10.12942/lrr-2007-3

[27] Ashtekar, A., Krishnan, B. (2004). Isolated and dynamical horizons and their applications. *Living Reviews in Relativity*, 7, 10. https://doi.org/10.12942/lrr-2004-10

[28] Christodoulou, D. (1970). Reversible and irreversible transformations in black-hole physics. *Physical Review Letters*, 25, 1596–1597. https://doi.org/10.1103/PhysRevLett.25.1596

[29] Kingma, D. P., Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations*, arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

[30] Liu, D. C., Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45, 503–528. https://doi.org/10.1007/BF01589116

[31] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32, arXiv:1912.01703. https://doi.org/10.48550/arXiv.1912.01703

[32] Clopper, C. J., Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26, 404–413. https://doi.org/10.1093/biomet/26.4.404

[33] Gundlach, C. (1998). Pseudospectral apparent horizon finders: An efficient new algorithm. *Physical Review D*, 57, 863–875. https://doi.org/10.1103/PhysRevD.57.863

[34] Cook, G. B., York, J. W. (1990). Apparent horizons for boosted or spinning black holes. *Physical Review D*, 41, 1077–1085. https://doi.org/10.1103/PhysRevD.41.1077

[35] Hannam, M., Husa, S., Ohme, F., Brügmann, B., Ó Murchadha, N. (2008). Wormholes and trumpets: Schwarzschild spacetime for the moving-puncture generation. *Physical Review D*, 78, 064020. https://doi.org/10.1103/PhysRevD.78.064020

[36] Dain, S., Lousto, C. O., Zlochower, Y. (2008). Extra-large remnant recoil velocities and spins from near-extremal-Bowen-York-spin black-hole binaries. *Physical Review D*, 78, 024039. https://doi.org/10.1103/PhysRevD.78.024039

[37] Lovelace, G., Owen, R., Pfeiffer, H. P., Chu, T. (2008). Binary-black-hole initial data with nearly extremal spins. *Physical Review D*, 78, 084017. https://doi.org/10.1103/PhysRevD.78.084017

[38] Brill, D. R., Lindquist, R. W. (1963). Interaction energy in geometrostatics. *Physical Review*, 131, 471–476. https://doi.org/10.1103/PhysRev.131.471