

Subgrid Modelling for Relativistic Magnetohydrodynamics with Machine Learning

William Cook

Sebastiano Bernuzzi, Tobias Dieselhorst, David Radice

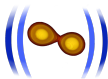
Phys. Rev. D 114, 063010

Symmetry 13 (2021) 11, 2157

Theoretisch-Physikalisches Institut FSU Jena

UKNR Machine Learning in NR & COCONUT Meeting

September 8, 2026



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

① Binary Neutron Stars

② Turbulence Modelling with Machine Learning

③ Large Eddy Simulations in SRMHD

④ Conserved-to-primitive variable inversion with ML

① Binary Neutron Stars

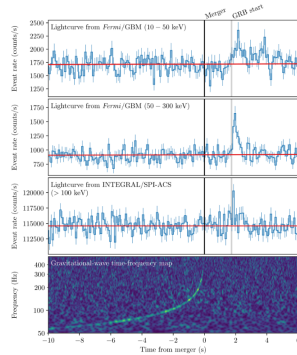
② Turbulence Modelling with Machine Learning

③ Large Eddy Simulations in SRMHD

④ Conserved-to-primitive variable inversion with ML

BNS mergers - a rich physics laboratory

- Binary Neutron Star systems probe all four forces of nature in extreme regimes.
- To date 2 BNS seen in GWs.
- GW170817 seen as multimessenger observation, with a coincident EM (gamma ray) signal, and long lived EM afterglow, the kilonova.

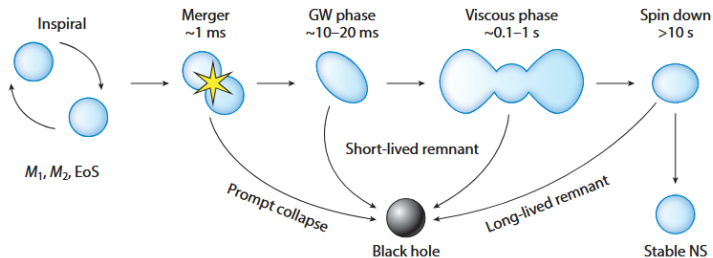


[Abbott+ '17]



BNS mergers - a rich physics laboratory: Gravity and the Strong Force

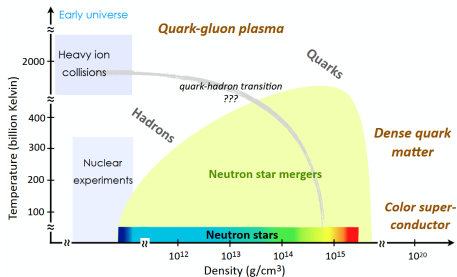
- BNS inspiral is driven by GW emission, ending in post merger remnant, either BH + disk, or NS.
- The GW signal morphology depends on the "deformability" of the nuclear matter.
- In the inspiral the GW encodes the "cold" neutron star equation of state.



[Radice+ '20]

BNS mergers - a rich physics laboratory: Gravity and the Strong Force

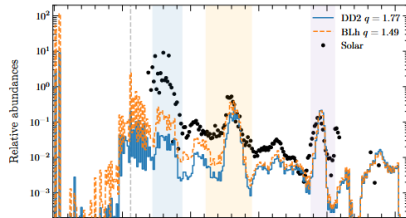
- Post merger the matter is shock heated, and the GW signal probes high temperature regions of the QCD phase diagram above nuclear saturation density.
- GW observations constrain the allowed EOS models for neutron stars.



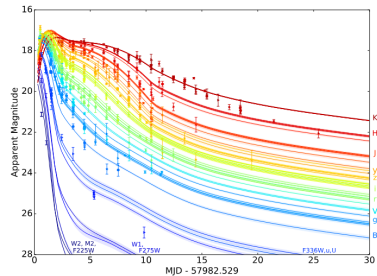
[Maggiore+ '19]

BNS mergers - a rich physics laboratory: Weak Force

- Neutrinos produced at high temperatures during the merger - radiate away energy, and can convert $n \leftrightarrow p$
- During the merger, neutron rich material ejected.
- This material undergoes r-process nucleosynthesis, decaying to stable isotopes.
- One of the main channels for forming heavy elements ($> \text{Fe}$) in the universe.
- Afterglow of this radioactive decay: the kilonova.



[Bernuzzi+ '24]

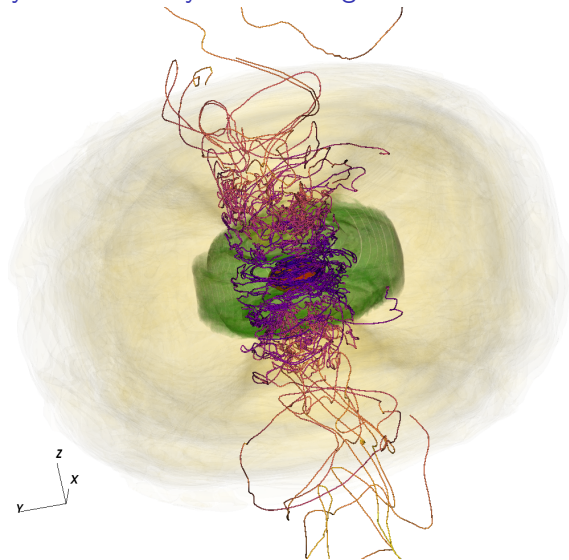


[Villar+]

'17]

BNS mergers - a rich physics laboratory: Electromagnetism

- Source NS contain magnetic fields typically $\sim 10^{10} - 10^{12} G$
- During a BNS merger magnetic and fluid instabilities amplify B field and trigger turbulent flow - structure cascades to smaller length scales.
- Large scale B field forms, triggers jet launching, and associated gamma ray burst.
- Magnetic field in remnant drives ejecta winds, impacts ejected material, associated nucleosynthesis
- Magnetic stresses transport angular momentum, impact collapse time of remnant and associated GW emission.



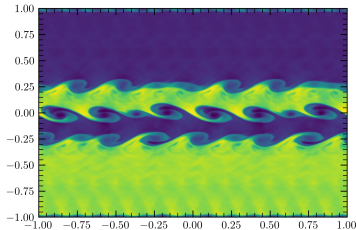
Numerical Relativity for accurate BNS modelling

- To accurately model these events we need to evolve a lot of physics: GR, finite temperature hydro, MHD, neutrino transport (M1/MC/Boltzmann)
- For accurate parameter estimation in ET we need waveform models informed by GW data from simulations with $<100\text{m}$ resolution
- Capturing magnetic field instabilities requires high resolution of merger and post merger phase.
- Simulations are expensive - aim to develop highly scaling codes on CPUs (Dendro-GR, GR-Athena++) and codes running on GPU architectures (ANRe-M1, AsterX, AthenaK, AMREX/CarpetX, GRACE, GRAMX, GRTeclyn, MHDuet, PHOEBUS, SACRA-K, Spectre...)
- Can we leverage Machine Learning techniques in parts of our codes to accelerate them, or incorporate extra physics in a cheaper simulation?

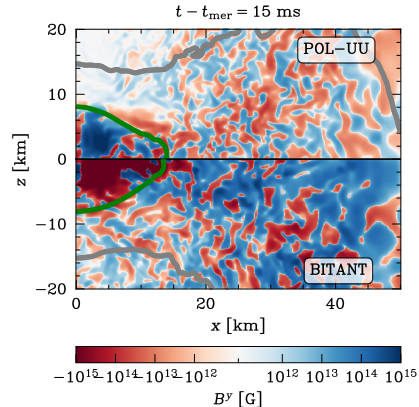
- ① Binary Neutron Stars
- ② Turbulence Modelling with Machine Learning
- ③ Large Eddy Simulations in SRMHD
- ④ Conserved-to-primitive variable inversion with ML

Magnetic field instabilities

- How does the B field amplify during BNS merger / post-merger?
- At merger Kelvin-Helmholtz instability. Fields $\sim 10^{12}$ may amplify up to $\sim 10^{16}$ [Kiuchi + '26] .
- Field amplification continues as B field undergoes turbulent cascade to small scales.
- Numerical simulation with discretisation length Δx can only capture amplification up to the point the cascade drops off the grid.

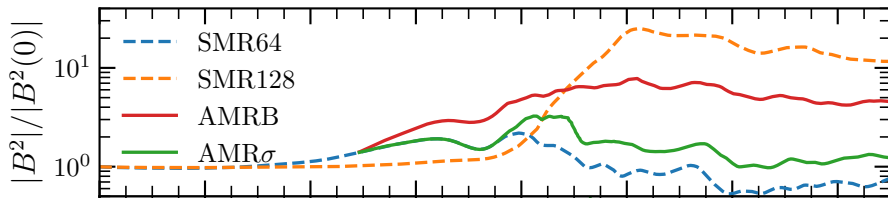


- In post-merger disk/remnant star MRI, Taylor Instability.
- Can power dynamo processes: $\alpha - \Omega$, Taylor-Spruit; and generate effective viscosity



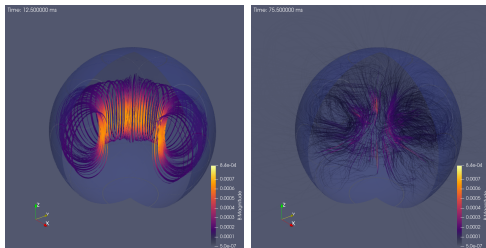
Numerical limitations

- Numerical simulation limited in its ability to capture amplification through Kelvin-Helmholtz instability by resolution. $\sigma \propto 1/\Delta x$
- How to model turbulence in numerical simulations? Cannot span the relevant range of length scales from viscous length $\mathcal{O}(1\text{cm})$ to e.g. GW wavezone $\mathcal{O}(1000\text{km})$.

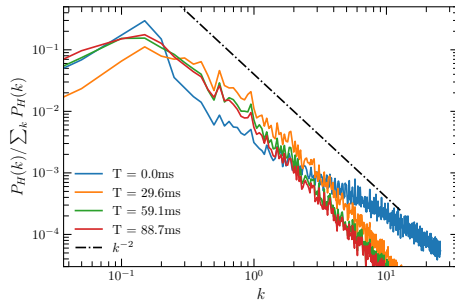
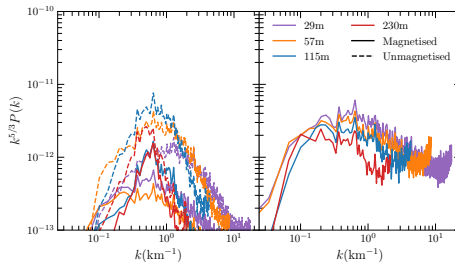


[WC+ '23]

- Modelling strategies: Learn what we can from few VHR sim - [WC+ '25] for single star, 29m, e.g. [Kiuchi+ '26] BNS, 6m.
- Runs incredibly costly e.g. 28,000 GPU node hours ; 5×10^8 CPU hours.
- Cannot explore parameter space with this approach. Use **Large Eddy Simulations + Subgrid Model** to perform low resolution simulations with model that incorporates back reaction of small scale physics on large scale evolution



Modelling strategy



Large Eddy Simulations

- In standard Finite Volume approach we discretise continuous equations on a grid of cells width Δx and evolve volume averaged variables.

$$q_t + f(q)_x = 0$$

$$Q_i^n = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} q(x, t_n) dx$$

$$F_{i+1/2}^n = \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} f(q(x_{i+1/2}, t)) dt$$

$$\frac{Q_i^{n+1} - Q_i^n}{\Delta t} + \frac{1}{\Delta x} (F_{i+1/2} - F_{i-1/2}) = 0$$

- Discretised grid acts like a “filter” on the fluid state.
- Lose information on scale $\ell < \Delta x$. Also lose back reaction of small scales onto resolved large scales.
- Numerical fluxes F are calculated as functions (indirectly) of Q , e.g. physical flux evaluated at solution to (approximate) Riemann problem at cell interface $f(Q^*)$.

Large Eddy Simulations

- Rather than having this arise from a numerical approximation, instead ask : what is the continuous equation that a volume averaged or “filtered” fluid obeys?

$$\begin{aligned}\bar{f}(x, t) &= \int_{x'_{\min}}^{x'_{\max}} G(x - x') f(x', t) dx' \\ G(x - x') &= \frac{1}{\Delta x} H\left(\frac{\Delta x}{2} - |x - x'|\right) \\ \bar{q}_t + \overline{f(q)}_x &= 0\end{aligned}$$

- Gives an **exact** equation for the evolution of the volume averaged variable $\bar{q}$ (equivalently Q).
- What's the catch? Equations are not closed, $\bar{f}$ contains unknown terms.
- Must prescribe a model for the behaviour on the subgrid scale “subgrid model”.
- Equations describe the evolution of “Large” eddies $\ell > \Delta x$, while subgrid model describes $\ell < \Delta x$
- Since generically f non-linear $\overline{f(q)} \neq f(\bar{q})$, like fluxes used in standard FV approach

Concrete example: B in Newtonian Physics

- Induction equation in ideal MHD

$$\partial_t B^k + \partial_i (B^k v^i - B^i v^k) = 0$$

- What equation does this obey when filtered on scales $> \Delta x$?
- Filtered equation:

$$\partial_t \overline{B^k} + \partial_i (\overline{B^k v^i} - \overline{B^i v^k}) = 0$$

- Problem: what is $\overline{B^k v^i}$? We know $\overline{B^k}$, $\overline{v^i}$, but filtration and non linear terms do not commute. Would need to know true continuum solution, take product, then filter.
- Traditionally we ignore this, evaluate $f(\overline{q})$.
- To capture the back reaction of small scale physics on the large scale filtered physics, we construct a subgrid model for these terms to close the system of Eqs.

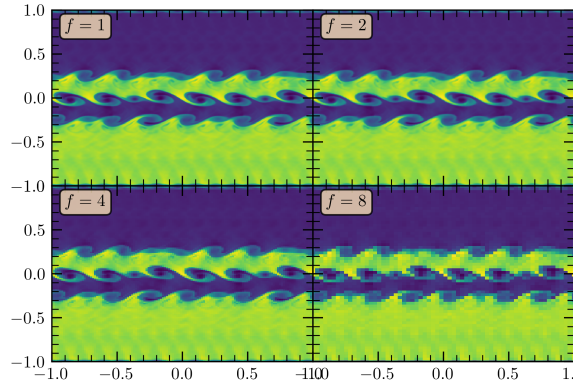
$$\begin{aligned} \partial_t \overline{B^k} + \partial_i (\overline{B^k} \overline{v^i} - \overline{B^i} \overline{v^k} - \tau_B^{ik}) &= 0 \\ \tau_B^{ik} &= (\overline{B^k} \overline{v^i} - \overline{B^i} \overline{v^k}) - (\overline{B^k v^i} - \overline{B^i v^k}) \end{aligned}$$

Subgrid models

- Numerical dissipation present in standard approximate Riemann solvers provides an implicit subgrid model - eddies of size Δx are naturally dissipated
- Various explicit closure schemes exist:
- Smagorinsky [Smagorinsky '63, Germano+ '90, Radice '17,'20] - (hydrodynamics) treat turbulent terms in the momentum transport as an effective viscosity. Free parameter of mixing length can be tuned to high resolution data [Radice '20] ,
- Gradient model [Yeo '87, Vigano+ '19, Carrasco+ '20, Palenzuela+ '21] . Physics agnostic. With various smoothness assumptions on the filtration kernel, expand $\overline{fg} \sim \overline{f}\overline{g} + \mathcal{O}(\Delta x^n)$. Introduce free parameters that can be fit to data.
- MINIT model [Miravet-Tenes+ '22,'23] - Including magnetic fields, model specific physical processes associated to MHD instabilities like KHI and MRI, set free parameters by comparing to data.

Subgrid models

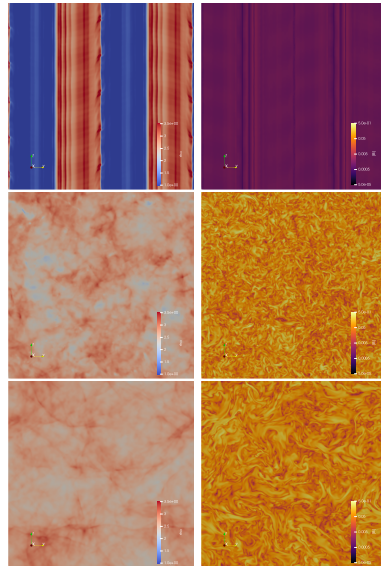
- Data driven approach: Machine Learning [Rosofsky, Huerta '19]
- Perform high resolution simulation of turbulent physics. Treat this as the “true” “continuum” solution.
- Calculate product terms like $B^i v^k$, as well as individual terms like B^i, v^i .
- Filter everything down to low resolution
 $B^i v^k \rightarrow \overline{B^i v^k}, B^i \rightarrow \overline{B^i}, v^i \rightarrow \overline{v^i}$.
- Train a neural network on the relationship between inputs: $\overline{B^i}, \overline{v^i}$ (evolved variables in our simulations) and outputs $\overline{B^i v^k}$ (the unknown subgrid terms)
- Perform a low res simulation and evaluate the neural network to add subgrid contribution.
- Low res simulation should capture the physics of the high res simulation used for training, at fraction of the cost.



- ① Binary Neutron Stars
- ② Turbulence Modelling with Machine Learning
- ③ Large Eddy Simulations in SRMHD
- ④ Conserved-to-primitive variable inversion with ML

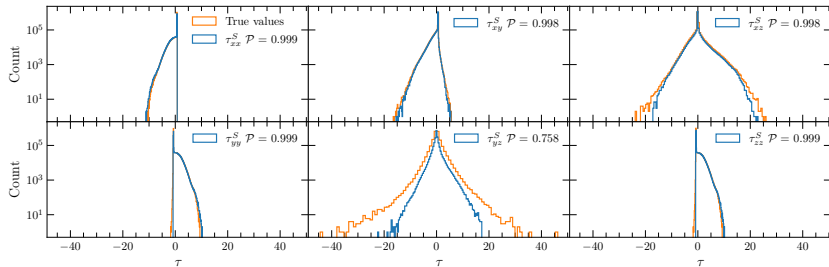
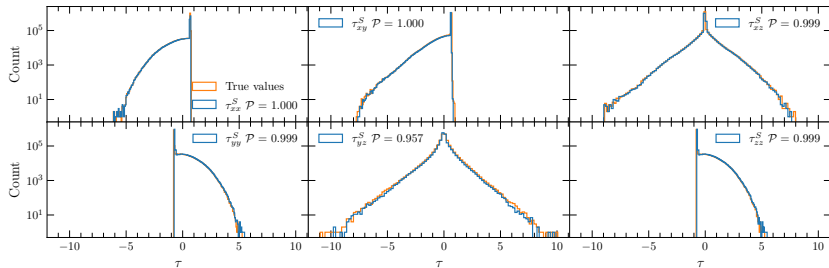
Kelvin-Helmholtz Simulation

- We use an idealised model, local Kelvin-Helmholtz instability in 3D, special relativity.
- We run 3 resolutions (LR, SR, HR)
- Filter the SR data by a factor 2 $\rightarrow$ LR
- Filter the HR data by a factor 4 $\rightarrow$ LR
- 2 sets of models f_2 , f_4
- Subgrid terms present in all evolution equations, separate model for each (D, S_j, τ, B).
- Code used GR-Athena++, modern NR code capable of solving BNS mergers in dynamical spacetime with GRMHD and M1 neutrino transport. [Daszuta (WC) + 21, WC+ '23, Daszuta (WC)+ '26]



A priori testing

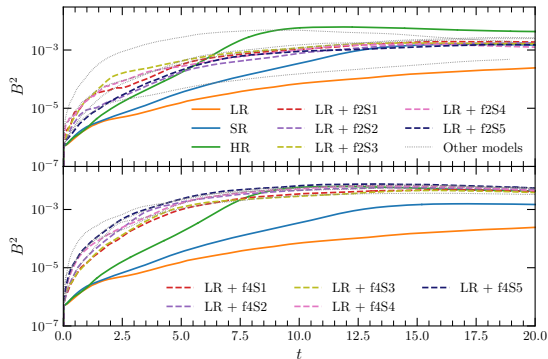
- Given a slice of data not in the training set can our model predict the correct subgrid tensors?



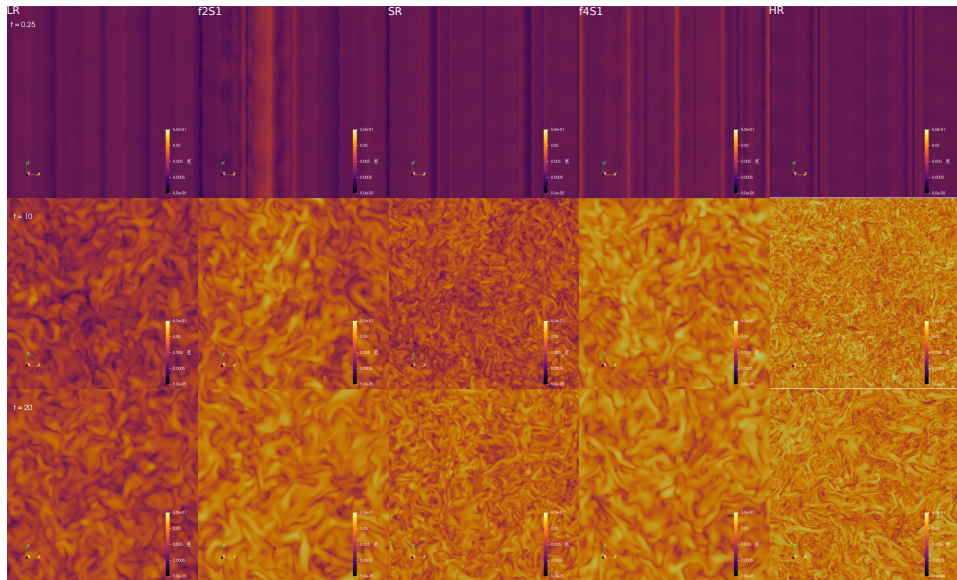
A posteriori testing

- We generate models with different hyperparameters and select those that give the target amplification at the end of the simulation.
- All models are designed to be small: all faster than the equivalent higher res run.

Name	$B^2(t = 20)$	Speed up to target
LR	2.43×10^{-4}	-
SR	1.48×10^{-3}	-
HR	4.31×10^{-3}	-
f2S1	1.88×10^{-3}	4.17
f2S2	1.75×10^{-3}	2.80
f2S3	1.62×10^{-3}	4.63
f2S4	1.26×10^{-3}	4.02
f2S5	1.52×10^{-3}	3.94
f4S1	4.02×10^{-3}	40.5
f4S2	4.75×10^{-3}	43.5
f4S3	3.81×10^{-3}	33.3
f4S4	4.46×10^{-3}	38.6
f4S5	5.46×10^{-3}	28.4



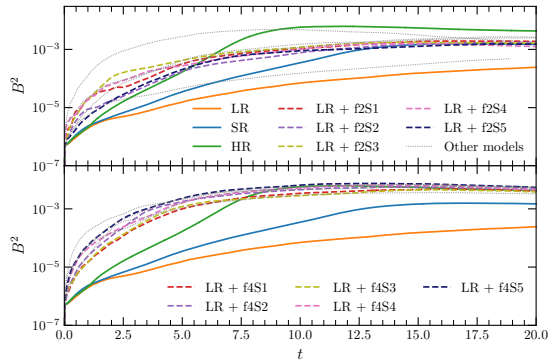
LES evolution



A posteriori testing: growth rate

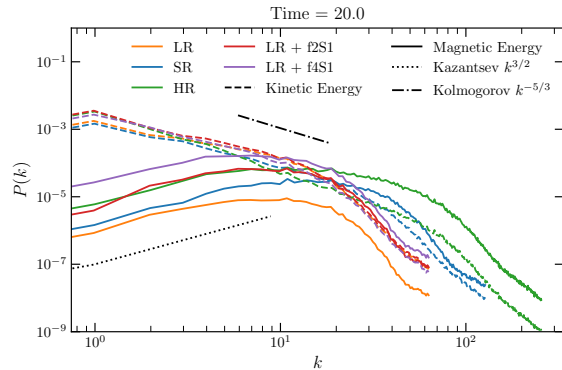
- Final B matches, but what about growth rate?
- Growth rate at finite resolution $B^2 \propto \exp(\gamma/\Delta x)$
- Can estimate γ from “no-subgrid runs” and effective $\Delta x = L/N_{\text{eff}}$.

Name	$N_{\text{mesh,eff}}$
LR	128
SR	256
HR	512
f2S1	270
f2S2	288
f2S3	159
f2S4	175
f2S5	285
f4S1	508
f4S2	541
f4S3	528
f4S4	428
f4S5	409



A posteriori testing: spectra

- How do the structures of the magnetic and kinetic energies compare?
- More energy injected on longer wavelengths.
- Match to expected Kolmogorov and Kazantsev scaling for dynamo process.
- Robust kinetic energy modelling even though model selection criterion based on magnetic fields.



Summary

- First machine learning subgrid model for relativistic MHD implemented in online simulation.
- Can be used to incorporate HR physics accurately in LR runs, for factor ~ 44 speedup.
- Can be applied to BNS simulations capturing magnetic field amplification from KHI.
- Do we need separate models for different physics, e.g. MRI vs KHI?
- Gauge invariance of model when moving to full GR?

- ① Binary Neutron Stars
- ② Turbulence Modelling with Machine Learning
- ③ Large Eddy Simulations in SRMHD
- ④ Conserved-to-primitive variable inversion with ML

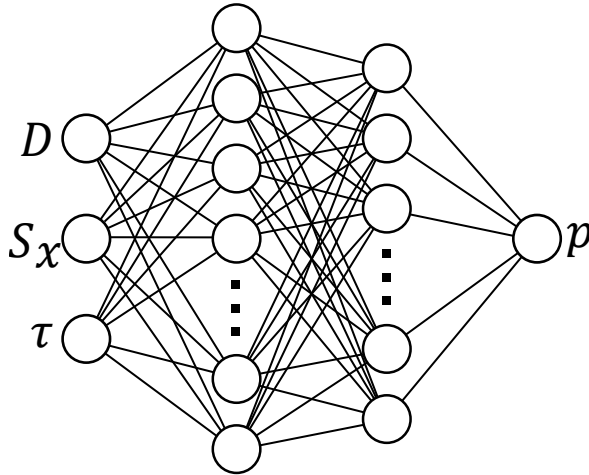
Conserved-to-Primitive with Machine Learning

- In relativistic simulations we must evaluate the conserved-to-primitive inversion $w = w(q)$.

$$D = \rho W, \quad S_i = \rho h W^2 v_i, \quad \tau = \rho h W^2 - p - D$$

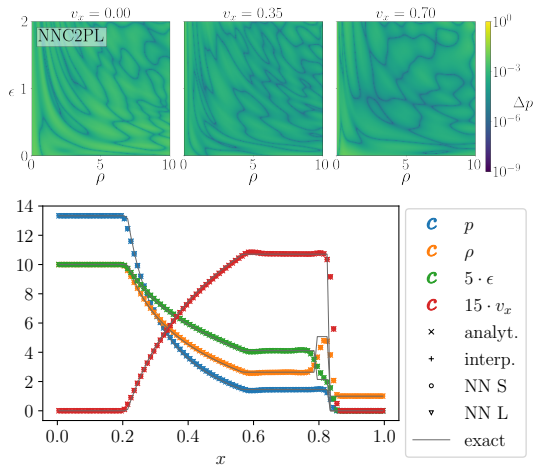
- Requires a root finding operation and evaluation of EOS $p = p(\rho, T, Y_e)$.
- Straightforward for simplified analytic EOS but for finite temperature, multicomponent fluids EOS given by tabulated data requires table interpolations
- Can take $\lesssim 10\%$ of code time.
- Train a NN on the relationship $w = w(q)$.

[Dieselhorst (WC) + '21]



Conserved-to-Primitive

- Using tabulated gamma law EOS $p = \rho\epsilon(\Gamma - 1)$, learn the relationship between conservatives and primitives
- Evaluate network in shock tube simulation to replace C2P - speeds up C2P by factor > 10 with comparable accuracy
- [Kacmaz+ '24] have extended to polytropic EOS, investigated GPU acceleration.
- First BNS application in [Mudimadugula+ '25] . Trade off between model size and speed - find little improvement due to need for small model sizes.
- Potential for further acceleration on more powerful hardware?



Conclusions

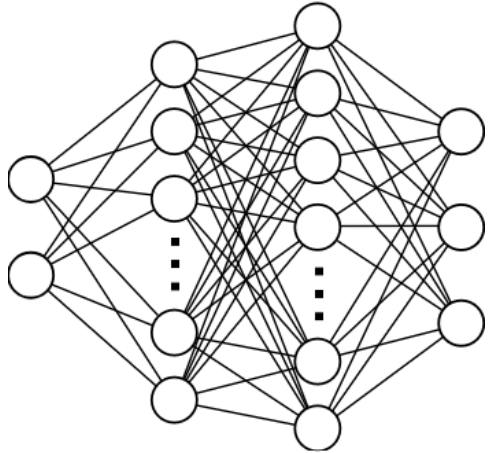
- There is a place for machine learning techniques in online numerical relativity simulations
- Clear application where models tuned on data need to be evaluated within time evolution:
- Subgrid model tuned on HR turbulent data, allows high resolution physics to be incorporated in LR simulations
- C2P network trained on EOS tables and C2P itself may accelerate this part of the code.
- ML can improve robustness of physics in simulations, and accelerate existing routines

Neural Network Architecture: Multilayer Perceptron

- Model the relationship between a vector of inputs (e.g.) $X^{(0)} = (\overline{B^i}, \overline{v^i})$ and outputs $X^{(\ell_n)} = (\overline{B^i v^k})$ as a sequence of n linear transformations from $\mathbb{R}^{m_0} \rightarrow \mathbb{R}^{m_1} \rightarrow \dots \rightarrow \mathbb{R}^{m_n}$ with a simple non linear activation function g in between:

$$X_i^{(\ell)} = g(W_{ij}^{(\ell-1)} X_j^{(\ell-1)} + b_i^{(\ell-1)})$$

- Weights W and biases b selected through stochastic gradient descent approach to minimise loss function on training dataset.
- For our purposes example activation functions include $ReLU(x) = \max(0, x)$, $\tanh(x)$, and loss function is mean squared error .
- The number of transformations or "hidden layers" n , the number of neurons per layer m_i , the activation and loss functions are hyperparameters chosen ahead of time for each particular model.



Subgrid equations in Special Relativity

$$\begin{aligned}
 \partial_t \bar{D} + \partial_i \left(\bar{D} \hat{v}^i - \tau_{\text{D}}^i \right) &= 0 & \tau_{\text{D}}^i &= \bar{D} \hat{v}^i - \overline{D v^i} \\
 \partial_t \bar{S}_j + \partial_i \left(\bar{S}_j \hat{v}^i + \delta_j^i \left(\hat{p} + \frac{\hat{b}^2}{2} \right) - \left(\frac{\hat{b}_j \bar{B}^i}{\hat{W}} \right) - \tau_{j\text{S}}^i \right) &= 0 & \tau_{j\text{S}}^i &= \bar{S}_j \hat{v}^i - \overline{S_j v^i} + \hat{p} - \bar{p} + \frac{\hat{b}^2}{2} - \frac{\bar{b}^2}{2} - \left(\left(\frac{\hat{b}_j \bar{B}^i}{\hat{W}} \right) - \overline{\left(\frac{b_j B^i}{W} \right)} \right) \\
 \partial_t \bar{\tau} + \partial_i \left(\bar{\tau} \hat{v}^i + \left(\hat{p} \hat{v}^i + \frac{\hat{b}^2 \hat{v}^i}{2} \right) - \frac{\hat{b}^0 \bar{B}^i}{\hat{W}} - \tau_{\tau}^i \right) &= 0 & \tau_{\tau}^i &= \bar{\tau} \hat{v}^i - \overline{\tau v^i} + \hat{p} \hat{v}^i - \overline{p v^i} + \frac{\hat{b}^2 \hat{v}^i}{2} - \frac{\bar{b}^2 \bar{v}^i}{2} - \left(\frac{\hat{b}^0 \bar{B}^i}{\hat{W}} - \overline{\left(\frac{b^0 B^i}{W} \right)} \right) \\
 \partial_t \bar{B}^k + \partial_i \left(\left(\bar{B}^k \hat{v}^i - \bar{B}^i \hat{v}^k \right) - \tau_{\text{B}}^{ik} \right) &= 0, & \tau_{\text{B}}^{ik} &= \bar{B}^k \hat{v}^i - \overline{B^k v^i} - \left(\bar{B}^i \hat{v}^k - \overline{B^i v^k} \right)
 \end{aligned}$$

- Added complication: conserved $\mathbf{q} = (D, S_j, \tau)$ and primitive $\mathbf{w} = (\rho, p, \epsilon, v^i)$ representation of hydro variables.
- $q = q(w)$ trivial, but $w = w(q)$ requires numerical solution of nonlinear algebraic eq.
- In standard FV simulations, we solve $w = w(\bar{q}) = \hat{w}$, not $w = \overline{w(\bar{q})} = \bar{w}$.
- We correct this with additions to the subgrid terms