
Attention-enhanced neural differential equations for physics-informed deep learning of ion transport

Danyal Rehman

Center for Computational Science and Engineering
Massachusetts Institute of Technology (MIT)
Cambridge, MA 02139, USA
drehman@mit.edu

John H. Lienhard

Department of Mechanical Engineering
Massachusetts Institute of Technology (MIT)
Cambridge, MA 02139, USA
lienhard@mit.edu

Abstract

Species transport models typically combine partial differential equations (PDEs) with relations from hindered transport theory to quantify electromigrative, convective, and diffusive transport through complex nanoporous systems; however, these formulations are frequently substantial simplifications of the governing dynamics, leading to the poor generalization performance of PDE-based models. Given the growing interest in deep learning methods for the physical sciences, we develop a machine learning-based approach to characterize ion transport across nanoporous membranes. Our proposed framework centers around attention-enhanced neural differential equations that incorporate electroneutrality-based inductive biases to improve generalization performance relative to conventional PDE-based methods. In addition, we study the role of the attention mechanism in illuminating physically-meaningful ion-pairing relationships across diverse mixture compositions. Further, we investigate the importance of pre-training on simulated data from PDE-based models, as well as the performance benefits from hard vs. soft inductive biases. Our results indicate that physics-informed deep learning solutions can outperform their classical PDE-based counterparts and provide promising avenues for modelling complex transport phenomena across diverse applications.

Introduction and Background

Modelling ion transport phenomena is a common problem observed across a host of applications that include ion-exchange through biological membranes ([Gschwend and Girault, 2020](#)), diffusing ionized gases in nuclear reactors ([Tsypin et al., 1998](#)), and the transport of metal ions through polyamide nanopores ([Ritt et al., 2020](#); [Roy et al., 2015](#)). Across these applications, different models are typically used to describe the physics of the governing transport phenomena; however, the two most common approaches, derived from irreversible thermodynamics, are the Maxwell-Stefan formulations and the Nernst-Planck (NP) equations ([Taylor and Krishna, 1993](#)). Maxwell-Stefan frameworks, although typically more accurate than the NP approach, can be used to model inter-species diffusion, yet require access to inter-species diffusion coefficients that become challenging to measure when large numbers of species are present ([Krishna and Wesselingh, 1997](#)). The NP equations, albeit simpler, introduce many assumptions and simplifications into the governing dynamics, which can adversely impact model performance making generalization a challenge ([Rehman et al., 2023](#)). Given the advent of deep learning methods in the natural sciences, there are clear opportunities to address some of the shortcomings of classical PDE-based ion transport models ([Nagy et al., 2021](#); [Rehman and Lienhard, 2023b](#)) through deep learning-based alternatives¹.

In this work, we explore the use of attention-enhanced neural differential equations to model ion

¹Relevant research covering machine learning for PDEs and ion transport is detailed in Appendix A.

transport across polyamide nanopores (Chen et al., 2018). We supplement classical neural differential equation models using the attention mechanism (Vaswani et al., 2017) and encode electroneutrality as an inductive bias (Rehman and Lienhard, 2023b) into the model architecture. Next, we pre-train the model on simulated data from PDE-based models supplemented with Gaussian noise to emulate experimental error, and then fine-tune the model on experimental data from over 750 measurements (Micari et al., 2020). Further, we highlight the importance of the attention layers by illustrating their ability to learn physically-representative ion-pairing relationships across studied solutions (Ahdab et al., 2020, 2021a). We also run ablations to ascertain the benefits of pre-training, while investigating the performance trade-offs between hard and soft inductive bias constraints. Lastly, we benchmark the performance of our approach relative to other competitive deep learning methods (LeCun et al., 1989; Ronneberger et al., 2015). Using our proposed method, we show that it is possible to learn multi-species transport across nanoporous membranes and improve predictive performance relative to conventional PDE-based solutions (Geraldes and Brites Alves, 2008).

Physics-informed Deep Learning Model

Neural Ordinary Differential Equations The hidden layer dynamics, $\mathbf{h}(J_v)$, are parameterized by a first-order ordinary differential equation (ODE) that depends on transmembrane fluid flux, J_v :

$$\frac{d\mathbf{h}(J_v)}{dJ_v} = f_\theta(\mathbf{h}(J_v), J_v; \theta) \quad (1)$$

with $J_v = \{0 \dots \mathcal{J}_v\}$, $\mathbf{h} \in \mathbb{R}^d$, and $f_\theta: [0, \mathcal{J}_v] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$. Here, d corresponds to the maximum number of charged solutes present across all datasets². Since the distribution of ions varies across the studied datasets, we mask out absent ions prior to the positional encodings and attention layer.

The outputs of the neural differential equations correspond to scalar ion concentrations, $\mathbf{h}(J_v)$, which are obtained by integrating over ODENet using the Tsitouras 5(4) numerical method (Simos and Tsitouras, 2018). In ODENet, $\theta \in \Theta$, is a set of learnable parameters from some finite dimensional parameter space, Θ (Chen et al., 2018). By learning the derivative of the output function, concentrations are uniformly Lipschitz continuous in $\mathbf{h}(J_v)$ and continuous in J_v , enabling facile pre-training on classical PDE-based transport models (Boral et al., 2023; Kidger, 2022).

The model is comprised of five linear layers, each with $\tanh(\cdot)$ non-linearities applied to the outputs. Prior to the orthogonal projector, no point-wise activations are used. The network is trained using Adam with a batch size of 32 and an initial learning rate of 10^{-3} (Kingma and Ba, 2014).

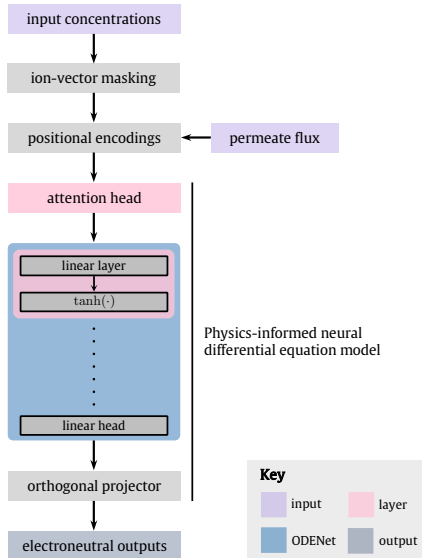


Figure 1: Physics-informed ODENet with an attention head and orthogonal projector for learning ionic context and predicting electroneutral outputs.

Attention Mechanism In language models, the attention mechanism serves as a means for learning semantic context (Vaswani et al., 2017); in the molecular or ionic setting, we can also leverage the attention mechanism to learn ionic context across diverse mixtures (Veličković et al., 2018). Using this approach, the model can identify governing ion-pairing relationships that dictate transport across polyamide nanopores (Rehman et al., 2023). Attention is calculated as follows:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^\top}{\sqrt{d_k}} \right) V \quad (2)$$

where Q , K , and V are the query, key, and value matrices, obtained from $W_Q^\top \in \mathbb{R}^{d_k}$, $W_K^\top \in \mathbb{R}^{d_k}$, and $W_V^\top \in \mathbb{R}^{d_k}$, respectively. In the reported work, we set $d_k = 8$ unless stated otherwise.

²We combine independent sets of experimental data all studying DuPont’s FilmTec™ NF270 polyamide membrane. The set of ions present across solutions is $\mathcal{S}: \{\text{Na}^+, \text{K}^+, \text{Li}^+, \text{Mg}^{2+}, \text{Ca}^{2+}, \text{Cl}^-, \text{SO}_4^{2-}, \text{NO}_3^-\}$.

Inductive Biases: Charge Conservation For dissociated ions in fluid systems, electroneutrality is typically a conserved quantity in the bulk solution (Gupta et al., 2022; Rehman and Lienhard, 2023a; Wang et al., 2021b). The conservation law can be quantified as follows:

$$\sum_{j=1}^d z_j \mathbf{h}_j(J_v) = 0, \quad \forall J_v \quad (3)$$

When treated as a hard constraint, we use the orthogonal projection of the hidden layer to ensure electroneutral outputs from the model. The projection is evaluated as follows: $z^\top \mathbf{h}_\perp = z^\top \mathbf{h} - z^\top \mathbf{h}_\parallel$, where $z \in \mathbb{R}^d$ corresponds to ion valences. During ablations, we elucidate the importance of the inductive bias by also applying it as a soft constraint. In this case, the electroneutrality term is simply appended to the loss functions as a regularization constraint to be minimized (Raissi et al., 2019).

Training Regime and Augmentations We first pre-train the neural solver on simulated data from the Donnan–Steric Pore Model with Dielectric Exclusion (DSPM–DE): a well-established PDE-based approach that involves solving the Nernst–Planck equations (Geraldes and Brites Alves, 2008). Details and derivation of the PDE-based model and regression formulation are provided in prior work (Rehman and Lienhard, 2022; Wang and Lin, 2021). The pre-training loss is expressed as follows:

$$\mathcal{L}^{\text{PDE}}(\mathbf{h}, \mathbf{h}^{\text{PDE}}) = \frac{1}{kd} \sum_{i=1}^k \sum_{j=1}^d [\mathbf{h}_j(J_{v,i}) - \mathbf{h}_j^{\text{PDE}}(J_{v,i})]^2 \quad (4)$$

Subsequently, we freeze the first three layers of the network and fine-tune the remaining two using measurement data fitted with Gaussian statistics to emulate experimental uncertainty:

$$\mathcal{L}^{\text{exp}}(\mathbf{h}, \mathbf{h}^{\text{exp}}) = \frac{1}{nd} \sum_{i=1}^n \sum_{j=1}^d [\mathbf{h}_j(J_{v,i}) - \mathbf{h}_j^{\text{exp}}(J_{v,i})]^2, \quad \mathbf{h}_j^{\text{exp}}(J_{v,i}) \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}^2) \forall i, j \quad (5)$$

where n corresponds to the number of flux measurements taken per species.

Results and Discussion

Predictive Performance For a sample ionic composition from the test set, we predict ion rejection rollouts as a function of flux: $\mathfrak{R}_j^{\text{mod}}(J_v) \triangleq (1 - [\mathbf{h}_j(J_v)/c_{j,\text{in}}])$. In Fig. 2A), we observe that the neural model outperforms the classical PDE-based method for a given rollout, while we note generally superior performance across the full test set in Fig. 2B). $\pm 10\%$ confidence bounds are included to illustrate the strong agreement achieved by the neural approach relative to the PDE-based model, which shows substantial deviations from the ground truth across a large number of test samples.

Implications of Attention In Fig. 2C), we benchmark the accuracy of our physics-informed ODENet relative to other deep learning methods and quantify the performance benefits of the attention mechanism. Our approach outperforms other ML-based approaches with the U-Nets achieving the closest MSE. This is likely due to the smooth profiles generated by ODENet for unseen fluxes that closely mirror experimental observation; other methods are unable to capture this continuity leading to inferior performance on the test data. For all conducted tests, we maintained a similar number of model parameters across benchmarks to ensure a fair comparison. In all cases, we note that the inclusion of the attention layer improves predictive performance.

Further, in the inset, we include a sample of the attention matrix with all ions reported. We note that the attention mechanism clearly learns the importance of valence and ionic size in transport, as seen by the elevated scores present for ions with the largest differences in radius and charge. Even more interestingly, we note that the importance of preserving electroneutrality is also learned: in cases of negative rejection – as demonstrated by NO_3^- in Fig. 2A) – the transport of the partner cation(s) or anion(s) is accelerated to conserve charge; the attention matrix is able to accurately identify the pertinent ion-pairs instrumental in achieving electroneutrality. In the example shown, we see that the attention given to both Cl^- and NO_3^- by Na^+ is high; this

makes physical sense as SO_4^{2-} is too large and immobile to be transported through the polyamide nanopores meaning that Cl^- and NO_3^- are the primary ions carried across to achieve electroneutrality (Ahdab et al., 2021b). This is similarly observed in Fig. 2A), where the presence of three cations means that NO_3^- transport must be expedited to ensure electroneutral outputs (this is exemplified by the negative rejection observed in NO_3^-). These findings clearly illustrate the value of the attention mechanism in learning and characterizing ion transport across polyamide nanopores.

Pre-training and Inductive Biases

In Fig. 2D), we quantify the impact of pre-training on simulated data from mechanistic models, and integrating inductive biases as hard constraints rather than soft regularization terms. By not pre-training on classical PDE-based models, we note that the resultant MSE is nearly 40% higher than when it is included. Despite the shortcomings of these PDE-based models (Wang and Lin, 2021), using them to improve the quality of the intermediate embeddings through pre-training substantially improves predictive performance on the downstream task.

In addition, our results demonstrate that pre-training on simulated data is even more impactful than treating the inductive biases as hard constraints. This is likely because in the data-limited regime, the model struggles to learn meaningful representations of the governing transport phenomena without appropriate guidance from the PDE-based methods; as a result, the inductive biases are insufficient in providing substantive signal to learn physically-representative trends. In the extreme case, when pre-training is not performed, and inductive biases are treated as soft constraints, we observe that predictive performance is even worse than the PDE-derived baseline across all deep learning-based methods investigated.

Conclusion

In this work, we propose attention-based neural differential equations for learning multi-ionic transport across nanoporous membranes. The model employs attention layers to learn physically-representative ion-pairing relationships that govern transport phenomena. We illustrate this through reported attention matrices that elucidate the role of ion valence in species transport. In addition, we run ablation studies to investigate the importance of pre-training the neural approach on classical PDE-based models. In data-limited settings, like the regime we are operating in, learning high quality embeddings through pre-training on classical models is imperative to achieving competitive performance against non deep learning alternatives. We also investigate the performance implications of treating charge conservation inductive biases as hard and soft constraints. We show that including them in either capacity outperforms classical PDE-based models across most deep learning methods (barring simple feed forward neural networks). Our results also highlight that transitioning from soft constraints to hard constraints drops MSE loss on the test set by 10-20%. These findings speak to the potential of neural methods to serve as robust alternatives to PDE-based models that often struggle to meet our performance requirements across diverse input and operating conditions.

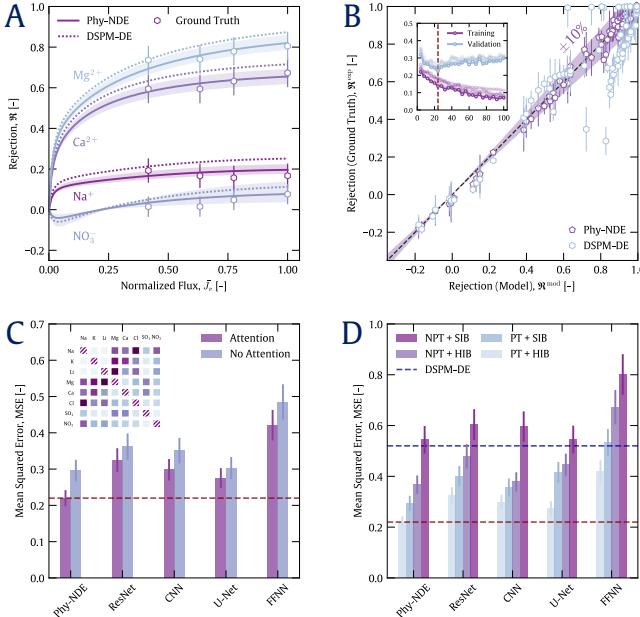


Figure 2: (A): Ion rejection predictions as a function of flux using the physics-informed ODENet and PDE-based DSPM-DE. (B): Parity plot illustrating general predictive performance across the test set using both the proposed ODENet and DSPM-DE. (C): The MSE loss achieved across deep learning methods with and without the attention mechanism included. The inset corresponds to a sample learned attention matrix for a given ionic composition from the test set. (D): The implications of pre-training and hard vs. soft inductive biases on predictive performance across a set of assessed deep learning methods. Key: **PT**: pre-trained with simulated data; **NPT**: no pre-training with simulated data; **HIB**: hard inductive bias; **SIB**: soft inductive bias.

Acknowledgments and Disclosure of Funding

The authors thank the Centers for Mechanical Engineering Research and Education at MIT and SUSTech (MechERE Centers at MIT and SUSTech) for partially funding this research. D.R. acknowledges financial support provided by a fellowship from the Abdul Latif Jameel World Water and Food Systems (J-WAFS) Lab and fellowship support from the Martin Family Society of Fellows.

References

- Y. Ahdab, D. Rehman, and J. Lienhard. Brackish water desalination for greenhouses: Improving groundwater quality for irrigation using monovalent selective electro dialysis reversal. *Journal of Membrane Science*, 610: 118072, 2020. URL <https://doi.org/10.1016/j.memsci.2020.118072>.
- Y. Ahdab, G. Schücking, D. Rehman, and J. Lienhard. Treatment of greenhouse wastewater for reuse or disposal using monovalent selective electro dialysis. *Desalination*, 507:115037, 2021a. URL <https://doi.org/10.1016/j.desal.2021.115037>.
- Y. Ahdab, G. Schücking, D. Rehman, and J. Lienhard. Cost effectiveness of conventionally and solar powered monovalent selective electro dialysis for seawater desalination in greenhouses. *Applied Energy*, 301:117425, 2021b. URL <https://doi.org/10.1016/j.apenergy.2021.117425>.
- A. Boral, Z. Wan, L. Zepeda-Núñez, J. Lottes, Q. Wang, Y. Chen, J. Anderson, and F. Sha. Neural ideal large eddy simulation: Modeling turbulence with neural stochastic differential equations, 2023. URL <https://arxiv.org/pdf/2306.01174>.
- W. Bowen and J. Welfoot. Modelling the performance of membrane nanofiltration—critical assessment and model development. *Chemical Engineering Science*, 57(7):1121–1137, 2002. URL [https://doi.org/10.1016/S0009-2509\(01\)00413-4](https://doi.org/10.1016/S0009-2509(01)00413-4).
- J. Brandstetter, M. Welling, and D. Worrall. Lie point symmetry data augmentation for neural PDE solvers. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 2241–2256. PMLR, 17–23 Jul 2022. URL <https://proceedings.mlr.press/v162/brandstetter22a/brandstetter22a.pdf>.
- J. Brandstetter, R. van den Berg, M. Welling, and J. Gupta. Clifford neural layers for PDE modeling, 2023. URL <https://arxiv.org/pdf/2209.04934>.
- R. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 31, 2018. URL <https://proceedings.neurips.cc/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf>.
- B. De Jaegher, W. De Schepper, A. Verliefde, and I. Nopens. Enhancing mechanistic models with neural differential equations to predict electro dialysis fouling. *Separation and Purification Technology*, 259:118028, 2021. ISSN 1383-5866. URL <https://doi.org/10.1016/j.seppur.2020.118028>.
- V. Geraldes and A. Brites Alves. Computer program for simulation of mass transport in nanofiltration membranes. *Journal of Membrane Science*, 321(2):172–182, 2008. URL <https://doi.org/10.1016/j.memsci.2008.04.054>.
- G. Gschwend and H. Girault. Discrete Helmholtz model: a single layer of correlated counter-ions. metal oxides and silica interfaces, ion-exchange and biological membranes. *Chemical Science*, 11(38):10304–10312, 2020. URL <https://doi.org/10.1039/D0SC03748F>.
- P. Gupta, S. Shanmukham, C. Patel, J. Lienhard, and J. Swaminathan. Replacing chloride anions in dyeing enables cheaper effluent concentration and recycling. *Desalination*, 533:115761, 2022. URL <https://doi.org/10.1016/j.desal.2022.115761>.
- O. Kedem and A. Katchalsky. Permeability of composite membranes. part 1.—electric current, volume flow and flow of solute through membranes. *Transactions of the Faraday Society*, 59:1918–1930, 1963. URL <https://pubs.rsc.org/en/content/articlepdf/1963/TF/TF9635901918>.
- P. Kidger. On neural differential equations, 2022. URL <https://arxiv.org/abs/2202.02435>.
- D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014. URL <https://arxiv.org/abs/1412.6980>.

- G. Kraaijeveld, V. Sumberova, S. Kuindersma, and H. Wesselingh. Modelling electro dialysis using the Maxwell-Stefan description. *The Chemical Engineering Journal and the Biochemical Engineering Journal*, 57(2): 163–176, 1995. URL [https://doi.org/10.1016/0923-0467\(94\)02940-7](https://doi.org/10.1016/0923-0467(94)02940-7).
- R. Krishna and J. Wesselingh. The Maxwell-Stefan approach to Mass Transfer. *Chemical Engineering Science*, 52(6):861–911, 1997.
- Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel. Handwritten digit recognition with a back-propagation network. *Advances in Neural Information Processing Systems*, 2, 1989. URL https://proceedings.neurips.cc/paper_files/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf.
- Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Fourier neural operator for parametric partial differential equations, 2021. URL <https://arxiv.org/pdf/2010.08895>.
- Z. Li, N. B. Kovachki, C. Choy, B. Li, J. Kossaifi, S. Otta, M. Nabian, M. Stadler, C. Hundt, K. Azizzadenesheli, and A. Anandkumar. Geometry-informed neural operator for large-scale 3D PDEs, 2023. URL <https://arxiv.org/pdf/2309.00583>.
- P. Lippe, B. Veeling, P. Perdikaris, R. Turner, and J. Brandstetter. Pde-refiner: Achieving accurate long rollouts with neural PDE solvers, 2023. URL <https://arxiv.org/pdf/2308.05732>.
- L. Lu, P. Jin, G. Pang, Z. Zhang, and G. Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. URL <https://doi.org/10.1038%2Fs42256-021-00302-5>.
- G. Mialon, Q. Garrido, H. Lawrence, D. Rehman, Y. LeCun, and B. Kiani. Self-supervised learning with Lie symmetries for partial differential equations, 2023. URL <https://arxiv.org/pdf/2307.05432>.
- M. Micari, D. Diamantidou, S. Heijman, M. Moser, A. Haidari, H. Spanjers, and V. Bertsch. Experimental and theoretical characterization of commercial nanofiltration membranes for the treatment of ion exchange spent regenerant. *Journal of Membrane Science*, 606:118117, 2020. URL <https://doi.org/10.1016/j.memsci.2020.118117>.
- E. Nagy, I. Hegedüs, D. Rehman, Q. Wei, Y. Ahdab, and J. Lienhard. The need for accurate osmotic pressure and mass transfer resistances in modeling osmotically driven membrane processes. *Membranes*, 11(2), 2021. URL <https://doi.org/10.3390/membranes11020128>.
- C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, A. Ramadhan, and A. Edelman. Universal differential equations for scientific machine learning, 2021. URL <https://arxiv.org/pdf/2001.04385>.
- M. Raissi, P. Perdikaris, and G. Karniadakis. 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, 2019. URL <https://doi.org/10.1016/j.jcp.2018.10.045>.
- D. Rehman and J. Lienhard. Global optimization for accurate and efficient parameter estimation in nanofiltration. *Journal of Membrane Science Letters*, 2(2):100034, 2022. URL <https://doi.org/10.1016/j.memlet.2022.100034>.
- D. Rehman and J. Lienhard. Physics-informed deep learning for multi-species membrane separations. *Under Review*, 2023a.
- D. Rehman and J. Lienhard. Physics-constrained neural differential equations for learning multi-ionic transport. *International Conference on Learning Representations (ICLR)*, 2023b. URL <https://arxiv.org/pdf/2303.04594>.
- D. Rehman, F. Sheriff, and J. Lienhard. Quantifying uncertainty in nanofiltration transport models for enhanced metals recovery. *Water Research*, 243:120325, 2023. URL <https://doi.org/10.1016/j.watres.2023.120325>.
- C. Ritt, J. Werber, M. Wang, Z. Yang, Y. Zhao, H. Kulik, and M. Elimelech. Ionization behavior of nanoporous polyamide membranes. *Proceedings of the National Academy of Sciences*, 117(48):30191–30200, 2020. URL <https://doi.org/10.1073/pnas.2008421117>.
- O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation, 2015. URL <https://arxiv.org/abs/1505.04597>.

- Y. Roy, M. Sharqawy, and J. Lienhard. Modeling of flat-sheet and spiral-wound nanofiltration configurations and its application in seawater nanofiltration. *Journal of Membrane Science*, 493:360–372, 2015. URL <https://doi.org/10.1016/j.memsci.2015.06.030>.
- T. Simos and C. Tsitouras. Fitted modifications of classical runge-kutta pairs of orders 5 (4). *Mathematical Methods in the Applied Sciences*, 41(12):4549–4559, 2018. URL <https://doi.org/10.1002/mma.4913>.
- R. Taylor and R. Krishna. *Multicomponent Mass Transfer*, volume 2. John Wiley & Sons, 1993.
- V. S. Tsypin, R. M. O. Galvão, I. C. Nascimento, A. G. Elfimov, M. Tendler, C. A. de Azevedo, and A. S. de Assis. Ion transport in tokamak plasmas with ion banana orbits squeezed by Alfvén waves. *Physical Review Letters*, 81:3403–3406, Oct 1998. URL [10.1103/PhysRevLett.81.3403](https://doi.org/10.1103/PhysRevLett.81.3403).
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017. URL <https://arxiv.org/pdf/1706.03762>.
- P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks, 2018. URL <https://arxiv.org/pdf/1710.10903>.
- Z. Wan, L. Zepeda-Núñez, A. Boral, and F. Sha. Evolve smoothly, fit consistently: Learning smooth latent dynamics for advection-dominated systems, 2023. URL <https://arxiv.org/pdf/2301.10391>.
- L. Wang, T. Cao, J. Dykstra, S. Porada, P. Biesheuvel, and M. Elimelech. Salt and water transport in reverse osmosis membranes: Beyond the solution-diffusion model. *Environmental Science & Technology*, 55(24):16665–16675, 2021a. URL <https://doi.org/10.1021/acs.est.1c05649>.
- L. Wang, D. Rehman, P. Sun, A. Deshmukh, L. Zhang, Q. Han, Z. Yang, Z. Wang, H. Park, J. Lienhard, and C. Tang. Novel positively charged metal-coordinated nanofiltration membrane for lithium recovery. *ACS Applied Materials & Interfaces*, 13(14):16906–16915, 2021b. URL <https://doi.org/10.1021/acsami.1c02252>.
- L. Wang, T. Cao, K. Pataroque, M. Kaneda, M. Biesheuvel, and M. Elimelech. Significance of co-ion partitioning in salt transport through polyamide reverse osmosis membranes. *Environmental Science & Technology*, 57(9):3930–3939, 2023. URL <https://doi.org/10.1021/acs.est.2c09772>.
- R. Wang and S. Lin. Pore model for nanofiltration: History, theoretical framework, key predictions, limitations, and prospects. *Journal of Membrane Science*, 620:118809, 2021. URL <https://doi.org/10.1016/j.memsci.2020.118809>.
- G. Wen, Z. Li, Q. Long, K. Azizzadenesheli, A. Anandkumar, and S. Benson. Real-time high-resolution CO₂ geological storage prediction using nested Fourier neural operators. *Energy & Environmental Science*, 16(4):1732–1741, 2023. URL <https://doi.org/10.1039/D2EE04204E>.
- X. Zhang, L. Wang, J. Helwig, Y. Luo, C. Fu, Y. Xie, M. Liu, Y. Lin, Z. Xu, K. Yan, K. Adams, M. Weiler, X. Li, T. Fu, Y. Wang, H. Yu, Y. Xie, X. Fu, A. Strasser, S. Xu, Y. Liu, Y. Du, A. Saxton, H. Ling, H. Lawrence, H. Stärk, S. Gui, C. Edwards, N. Gao, A. Ladera, T. Wu, E. Hofgard, A. Tehrani, R. Wang, A. Daigavane, M. Bohde, J. Kurtin, Q. Huang, T. Phung, M. Xu, C. Joshi, S. Mathis, K. Azizzadenesheli, A. Fang, A. Aspuru-Guzik, E. Bekkers, M. Bronstein, M. Zitnik, A. Anandkumar, S. Ermon, P. Liò, R. Yu, S. Günnemann, J. Leskovec, H. Ji, J. Sun, R. Barzilay, T. Jaakkola, C. Coley, X. Qian, X. Qian, T. Smidt, and S. Ji. Artificial intelligence for science in quantum, atomistic, and continuum systems, 2023. URL <https://arxiv.org/pdf/2307.08423>.
- Z. Zhang and Y. Luo. Deep spatial learning with molecular vibration, 2020. URL <https://arxiv.org/pdf/2011.07200>.

A Related Work

Deep Learning for PDEs Some recent scientific machine learning research has considered the development of neural operators, like the Fourier Neural Operator (FNO) (Li et al., 2021) and DeepONet (Lu et al., 2021), for applications to thermal-fluid sciences (Li et al., 2023), carbon capture (Wen et al., 2023), and even quantum systems (Zhang et al., 2023). These methods typically map function spaces to function spaces to learn grid-independent dynamics of diverse PDEs. Other work has focused on finite-dimensional neural alternatives to numerical methods like neural ordinary differential equations (Chen et al., 2018), universal differential equations (Rackauckas et al., 2021), physics-informed neural networks (Raissi et al., 2019), and Clifford group equivariant networks (Brandstetter et al., 2023), to name a few. Ongoing efforts have also tried to make these neural methods more accurate and/or data-efficient through the integration of symmetry-derived inductive biases (Brandstetter et al., 2022; Mialon et al., 2023), while others have focused on improving long-term prediction accuracies (Lippe et al., 2023; Wan et al., 2023). For ion transport specifically, there have been some efforts to leverage neural methods to supplement mechanistic models for fouling prediction (De Jaegher et al., 2021) and membrane monomer design (Zhang and Luo, 2020), but the development of new deep learning-based process models remains a field of ongoing research.

Mechanistic Ion Transport Models The first ion transport models across nanoporous membranes were derived from irreversible thermodynamics in the 1960s (Kedem and Katchalsky, 1963). Since then, substantial improvements have been proposed, with both the Maxwell-Stefan frameworks (Kraaijeveld et al., 1995), and the extended Nernst-Planck equations-based Donnan–Steric Pore Model with Dielectric Exclusion (DSPM–DE) (Bowen and Welfoot, 2002). Other recent models are typically extensions and variants of these models that have been further developed in recent years (Wang et al., 2021a, 2023). Despite this progress, these models typically struggle to generalize to new compositions due to the introduction of highly-constraining model simplifications and assumptions (Rehman et al., 2023). These simplifications are most often introduced as closure models to estimate parameters under nano-confinement; however, these functional relationships typically overconstrain the model, preventing accurate predictions across diverse operating conditions (Wang and Lin, 2021).