Predicting extensive properties of atomistic systems with deep neural networks

K. Ryczko, K. Mills, I. Luchak, A. Domurad, C. Homenick, I. Tamblyn

Key take-away points:

- Large nanoscale systems studied with density functional theory are computationally expensive (scale as $O(N^3)$, N = grid points)
- Deep neural networks scale as O(N), N = number of weights
- We show how to calculate properties of nanoscale systems with deep learning with and without extensivity

Publications:

Ryczko, K., Mills, K., Luchak, I., Homenick, C., & Tamblyn, I. (2018). Convolutional neural networks for atomistic systems. Computational Materials Science, 149, 134-142.

Luchak, I., Mills, K., Ryczko, K., Domurad, A., & Tamblyn, I. (2017). Extensive deep neural networks. arXiv preprint arXiv:1708.06686.

CLEAN Computational Laboratory for Energy And Nanoscience
NSERC CRSNG
uOttawa
NRC · CNRC
compute · calcul CANADA
SCIP SMART COMPUTING FOR INNOVATION
NVIDIA

Density functional theory

In the Kohn-sham DFT framework, one minimizes the total energy functional

$$E[n] = T[n] + E_{\text{ext}}[n] + E_{\text{Hartree}}[n] + E_{\text{XC}}[n] \tag{1}$$

by self-consistently solving the one particle Schrödinger equation

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{ext}}(\mathbf{r}) + \frac{1}{2}\int d\mathbf{r}' \, \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + V_{\text{XC}}(n(\mathbf{r}))\right]\psi_i = \epsilon_i \psi_i \tag{2}$$

with

$$n(\mathbf{r}) = \sum_{i=0}^{N/2} |\psi_i(\mathbf{r})|^2. \tag{3}$$

In Equation 1, T is the non-interacting kinetic energy functional, E_{ext} is the energy due to the interaction of the electrons with the external potential, E_{Hartree} is the electrostatic energy describing the electron-electron interactions, and E_{XC} is the exchange-correlation energy. Multiplying equation **??** by $\psi_i^*(\mathbf{r})$, summing over the occupied orbitals and integrating over all space, we get

$$-\frac{1}{2}\sum_i^{N/2}\langle\psi_i(\mathbf{r})|\nabla^2|\psi_i(\mathbf{r})\rangle + \int d\mathbf{r}\; V_{\text{ext}}(\mathbf{r})n(\mathbf{r})$$
$$+\int\int d\mathbf{r}d\mathbf{r}' \, \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \int d\mathbf{r} V_{\text{XC}}(n(\mathbf{r}))n(\mathbf{r}) = \sum_i^{N/2}\epsilon_i. \tag{4}$$

In equation 1, we can replace the non-interacting kinetic energy functional with

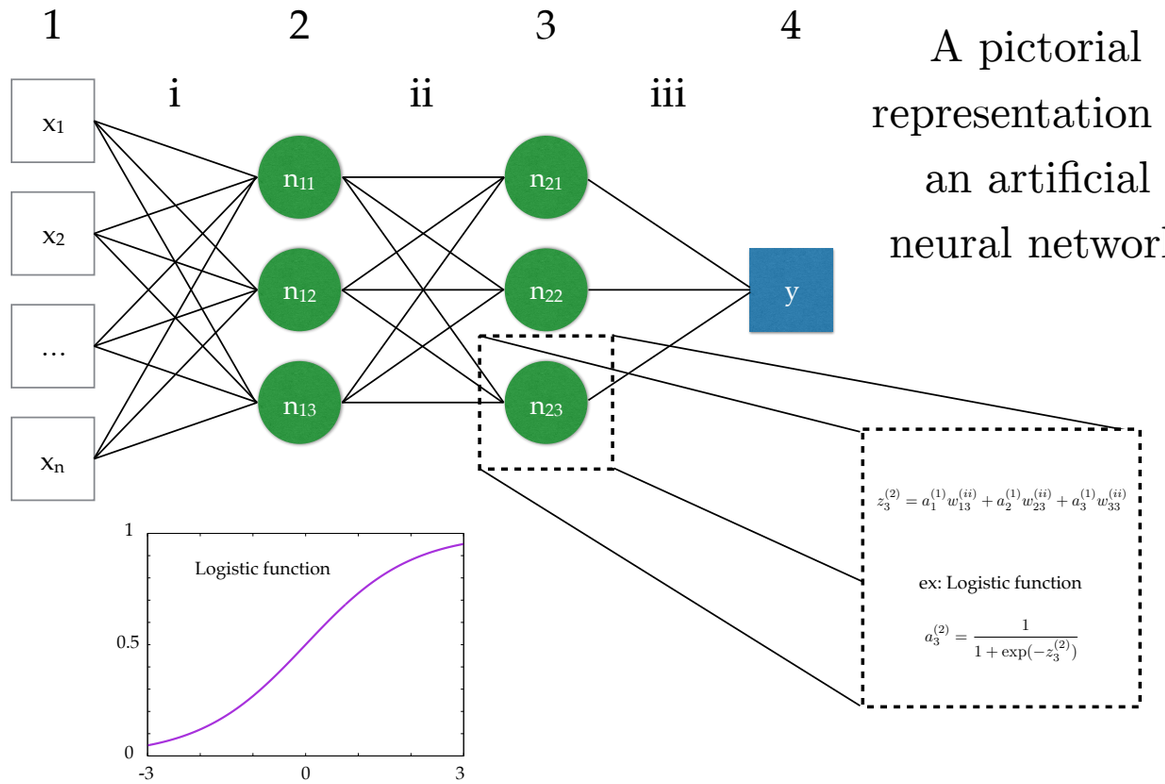
$$T[n] = \sum_i^{N/2}\epsilon_i - \int d\mathbf{r}\; V_{\text{ext}}(\mathbf{r})n(\mathbf{r}) - \int\int d\mathbf{r}d\mathbf{r}' \, \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} - \int d\mathbf{r} V_{\text{XC}}(n(\mathbf{r}))n(\mathbf{r}) \tag{5}$$

such that the total energy is

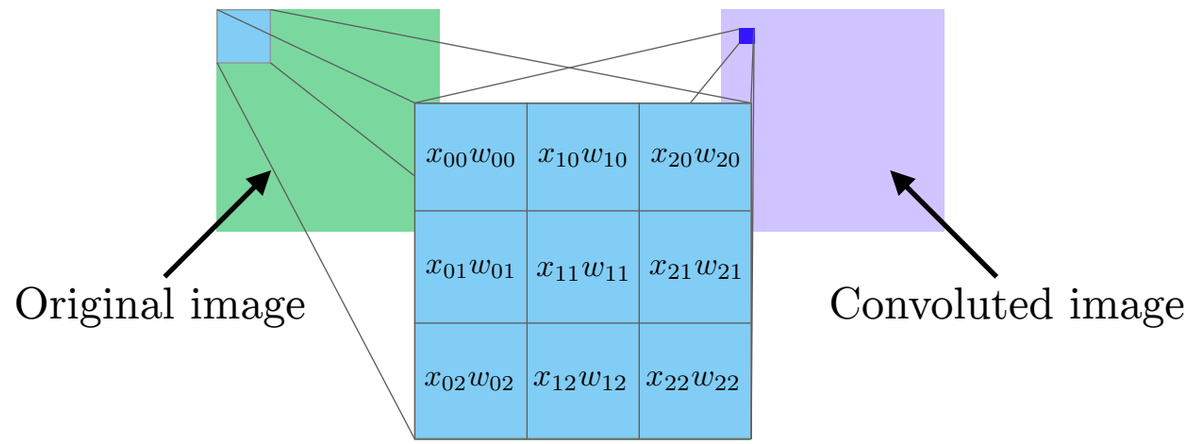
$$E = \sum_i^{N/2}\epsilon_i - \frac{1}{2}\int\int d\mathbf{r}d\mathbf{r}' \, \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\text{XC}}[n] - \int d\mathbf{r} V_{\text{XC}}(n(\mathbf{r}))n(\mathbf{r}). \tag{6}$$

What we want to predict

Convolutional neural networks



A pictorial representation of an artificial neural network



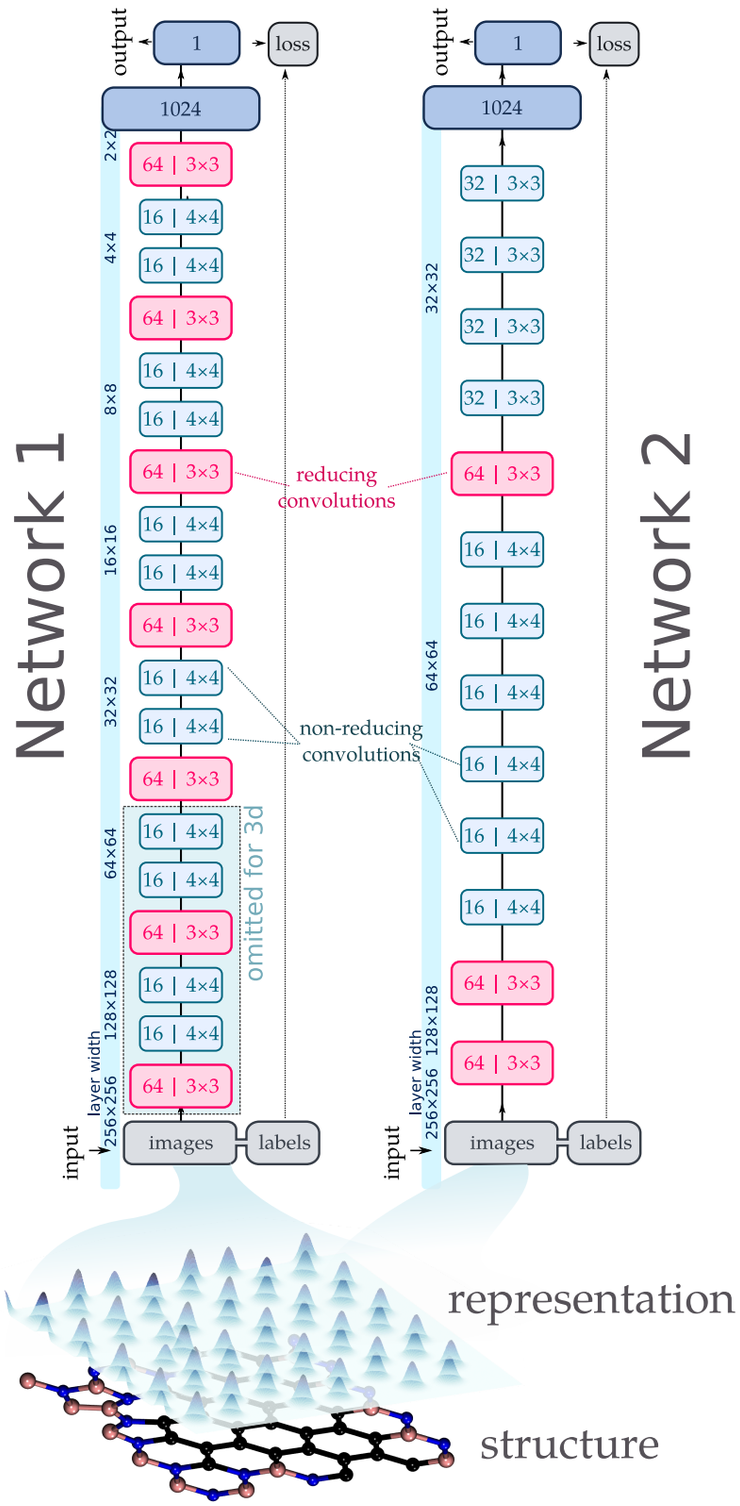
A pictorial representation of a convolutional kernel operating on a grey-scale image. The kernel is applied across the entire image (stride = 1) to produce a convoluted image. The new image pixels are given by

$$\tilde{x}_{nm} = a\left(\sum_{i=0,j=0}^{2} x_{i+n,j+m}w_{ij}\right)$$

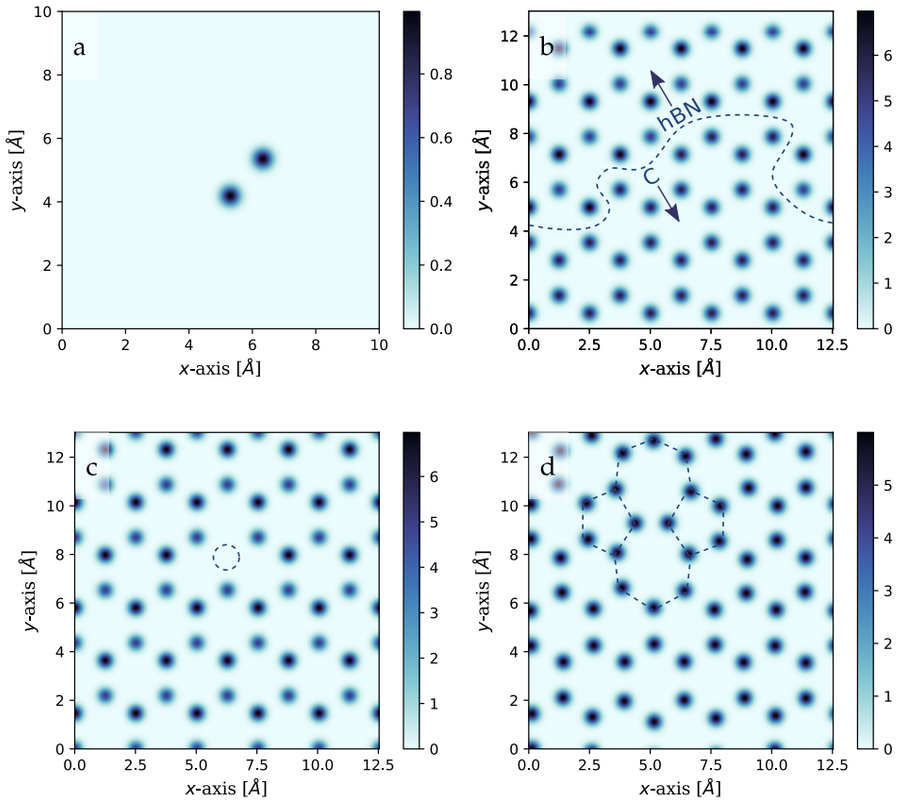
where a is an activation function. The weights are updated with the back-propagation algorithm. We essentially need to find the partial derivative of the loss function with respect to each weight.

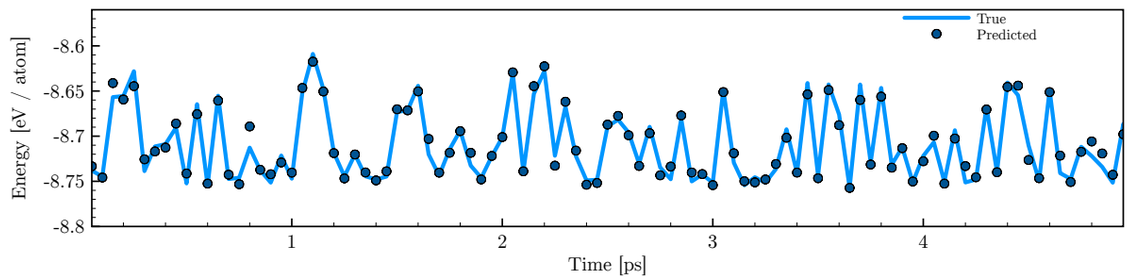
Predicting total energies with convolutional neural networks

The networks used in the study.

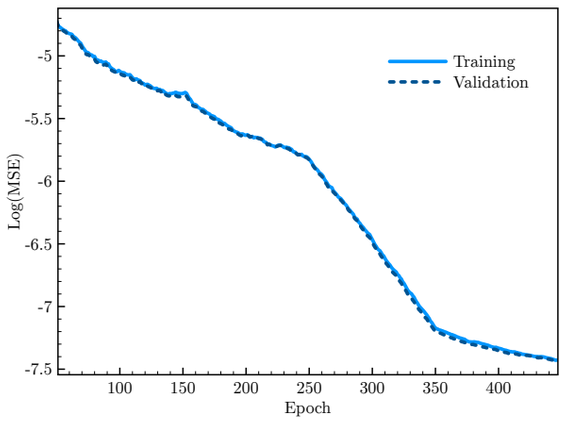


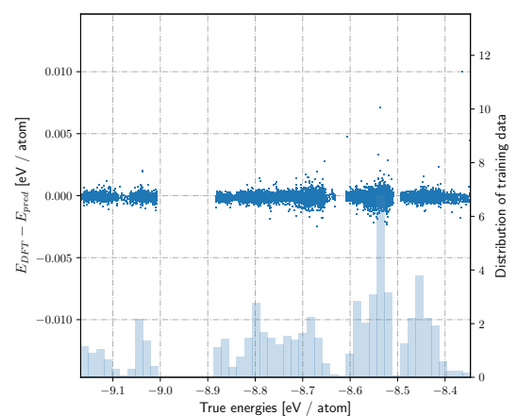
Examples of input images fed into the neural network. a) Dimer molecule interacting through a Lennard-Jones potential. b) Graphene-hexagonal Boron-Nitride (hBN) heterostructure. c) hBN with a single point defect. d) Graphene with a Stone-Wales defect.





6.85 meV / atom mean absolute error on independent molecular dynamics.





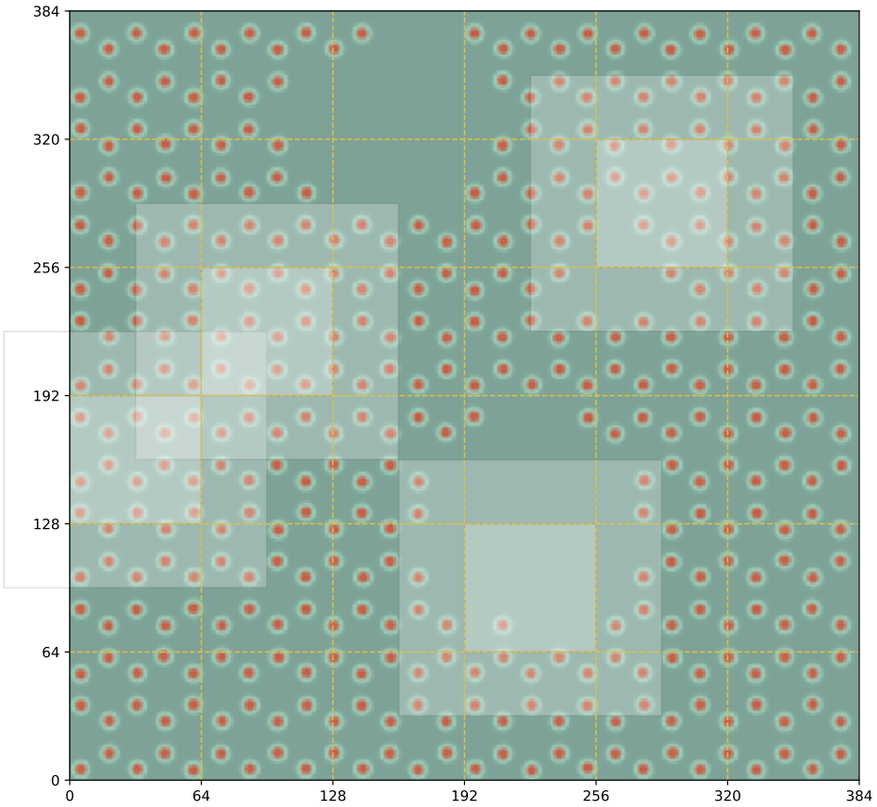
0.198 meV / atom mean absolute error on test set. This data set is correlated with the training set.

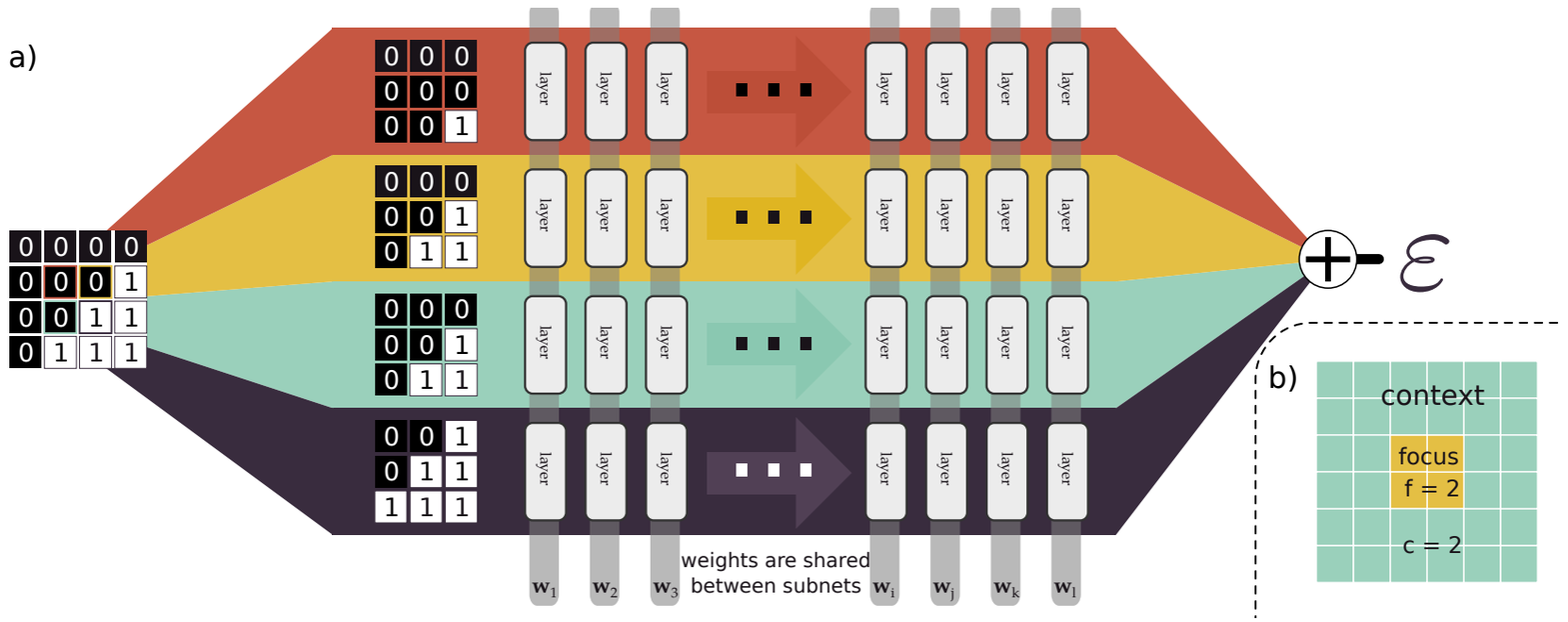
Chemical accuracy: 43.36 meV

Including extensivity in deep neural networks

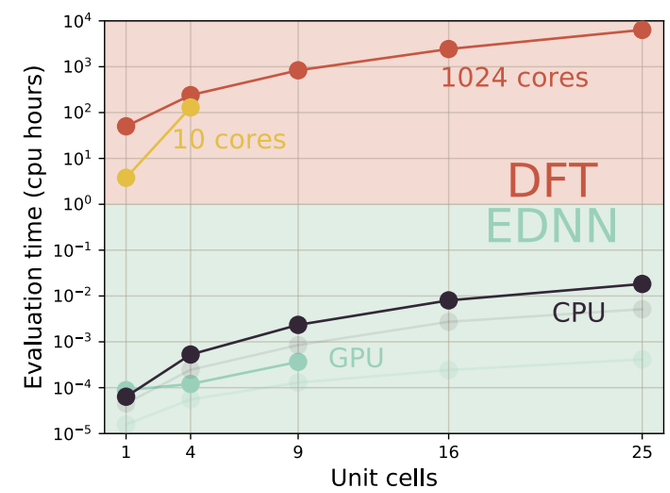
Below: The process of breaking down images into smaller ones (tiles) based on focus and context sizes. The larger image is generated using Gaussian functions to represent atomic positions. This was done for a graphene sheet with nano-pores.

Right: The tiles are (simultaneously) fed into parallel networks, which are summed over to produce the extensive quantity.





Scaling performance of density functional theory and extensive deep neural networks versus system size. Each unit cell consists of 60 atoms.



Energy predictions from extensive deep neural networks and true energy calculations done with density functional theory. All of the predictions are within chemical accuracy.