Supplementary Information (SI)

Neural-Network-Biased Genetic Algorithms for Materials Design: Evolutionary Algorithms that Learn

Tarak K. Patra, Venkatesh Meenakshisundaram, Jui-Hsiang Hung and David S. Simmons*
Department of Polymer Engineering, The University of Akron, 250 South Forge Street, Akron Ohio 44325, United States

Simulation methods

Polymer molecular dynamics simulations employ a bead-spring polymer model[1] that has been widely employed in the simulation of polymeric systems.[2–4] Within this model, the van der Waals interaction between two non-bonded monomers is represented by a 12-6 Lennard-Jones potential:

$$V_{LJ}\left(r\right)=4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12}-\left(\frac{\sigma}{r}\right)^{6}\right], \quad (1)$$

where, σ and ε correspond to a bead size and interaction strength, respectively. The interaction is truncated and shifted at $r = r_c = 2.5\sigma$ such that $V_{LJ}(r_c) = 0$. Covalently bonded beads additionally interact via a finitely extensible nonlinear elastic (FENE) potential of the form

$$V_{FENE}\left(r\right)=-\frac{1}{2}k_0R_0^2\ln\left[1-\left(\frac{r}{R_0}\right)^2\right]+4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12}-\left(\frac{\sigma}{r}\right)^{6}\right], \quad (2)$$

where, the second term is truncated at a distance of $2^{1/6}\,\sigma$, $k_0 = 30$ sets the bond energy and $R_0 = 1.5\sigma$ is the maximum bond length. This model is employed for the polymer density optimization and compatibilizer sequence optimization problems as described in the following sections.

Density optimization

We simulate a polymer melt of 40-bead copolymer chains, each consisting of a sequence of beads of type 0 and type 1, having interaction strength $\varepsilon_{00} = 1$ and $\varepsilon_{11} = 0.7$, with a cross-interaction of $\varepsilon_{01} = 0.83$. The sequence of beads of type 0 and 1 in the chain is specified within the genetic algorithm by that material's 40-bit genome, with a 0 indicating a bead of type 0 and a 1 indicating a bead of type 1. Because a higher cohesive energy favors higher density at a fixed temperature, an optimization for the sequence yielding the highest density should therefore yield all type 0 repeat units. The density of each model polymer is determined via molecular dynamics (MD) simulations within the LAMMPS MD environment.[5] Simulations are conducted in the isothermal-isobaric ensemble where the number of particles (N), pressure (P) and Temperature (T) are constant. A reduced temperature $T = 1$ and reduced pressure $P = 0$ are employed in these simulations. The Equation of motion is integrated using the Verlet algorithm with the time step $\Delta t = 0.005\tau$, where τ is the unit of time. The temperature and pressure are controlled using the Nose-Hoover thermostat and barostat with damping parameters of 2 τ in LJ units for both.[6] The system is periodic in all three directions. Each system is equilibrated for 10^5 steps, sufficient to reach the equilibrium density, followed by a data-collection run of 10^5 steps. The density of the system is averaged over data points during the production run.

Copolymer compatabilizers

Within simulations of interfacial compatibilization, non-bonded beads interact with LJ parameters $\varepsilon = 1$ and $\sigma = 1$; however, whereas interactions between like beads include attractions by employing a cutoff distance of 2.5σ, interactions between unlike beads (0-1) are made to be fully repulsive by employing a cutoff distance of $2^{1/6}$σ. As a consequence, homopolymers of type 0 and type 1 are highly immiscible, with an interface of order 1σ thick. The number of particles (N), temperature (T) and pressure (P) normal to the interface are constant in the simulation. The equation of motion are integrated using the Verlet algorithm with the integration time step of 0.005τ, and temperature (T=1) and pressure (P=0) are controlled by the Nose-Hoover thermostat and barostat.[6] The simulation box dimensions along two directions (*x* and *y*) are fixed, as the two polymer domains form the interface along the corresponding plane (*xy*). The box dimension along the third direction (z) is adjusted by the barostat in order to maintain system's pressure at zero. The system is periodic in all three directions. The system is equilibrated for $2\text{x}10^6$ steps, and data are collected for the following $2\text{x}10^6$ steps. The genetic algorithm seeks to minimize the interfacial energy γ, which is computed from the data collection period from the system pressure tensor as[7]

$$\gamma = \left\langle \frac{L_z}{2}\left(P_{zz} - \frac{1}{2}\left(P_{xx} + P_{yy}\right)\right)\right\rangle, \qquad (3)$$

where P_{xx}, P_{yy} and P_{zz} are the normal diagonal components of the pressure tensor. The normal to the interface is in the z-direction of the simulation box, and L_z corresponds to the length of the box in this direction.

Neural network methodology

Ground state of Ising model:

The topologies of the ANN for 2D and 1D Ising Models are {36 36 36 1} (36 input nodes, two 36-node hidden layers, and one output node), and {40 40 40 1}, respectively. The training period for the ANN is 5000. Figure 1 illustrates the reduction in prediction error during the training for the largest training data set tested. In order to ensure that the ANN is not overfitting to the data, we also performed a check in which we tested the network performance on a randomly chosen set of test data withheld from training. Figure 2 shows the error in ANN's prediction on test data, which are not incorporated in the training set, as a function of the training period. These results indicate that the ANN is not overfitting to the data and is nearing convergence at the time of training termination.

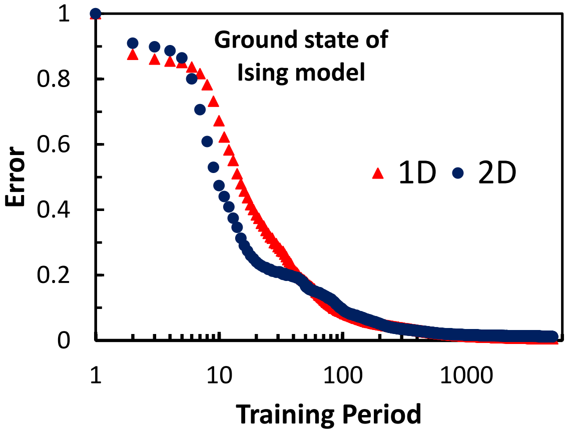

Figure 1: Error in the ANN's prediction during the training. The error is calculated on the training data set. Total data points used for training the ANN for 1D and 2D Ising models are 4992 and 2240, respectively.

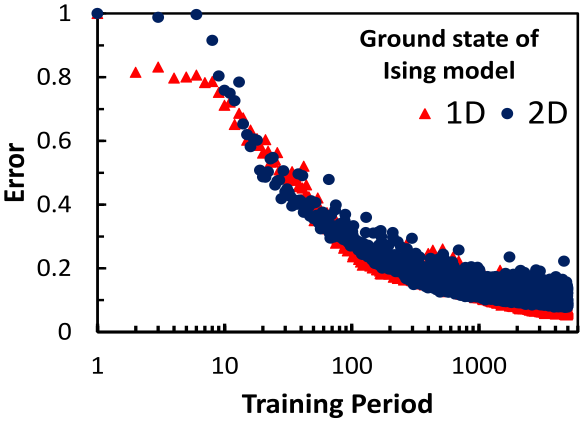

Figure 2: Prediction error during training, measurement on an unknown data set. The training and testing data points for 1D Ising model system are 4493 and 499 respectively. Similarly, the training and testing data points for 2D Ising model are 2016 and 224 respectively.

Gene maximization

The ANN topology for gene maximization is {100, 100, 1}. Figure 3 represents the error during the training for the largest training data set used in this study. The error is seen to converge with in the training period. Similar to the previous case study, in Figure 4, we report the results of employing the neural network to predict test data, which are not included in the training set. This indicates that the ANN is not being over fitted and is near convergence at the end of training.

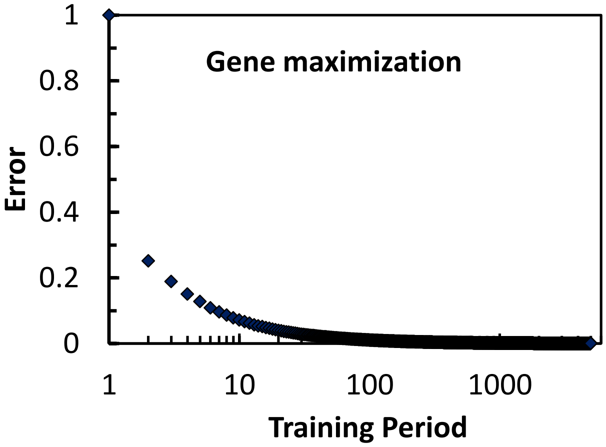

Figure 3: Error on the training data set as a function of training period. Total data points used for training the ANN are 2272.

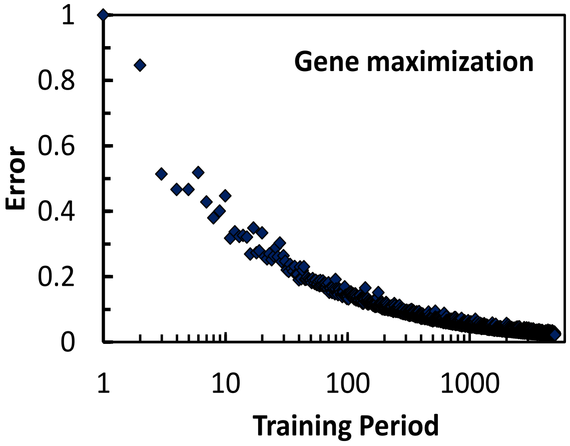

Figure 4: Error calculated on the unknown data set. The training and test data points are 2045 and 227 respectively.

Polymer density optimization

The topology of the ANN employed for this problem is {40 40 20 1}. Figure 5 shows the reduction of error during the training of the network. Error is also calculated for unknown data set during the training of the ANN, which is shown in Figure 6. Figure 5 and 6 suggest that the prediction error during the training is decreasing systematically and reaching a convergence at the end of the training.

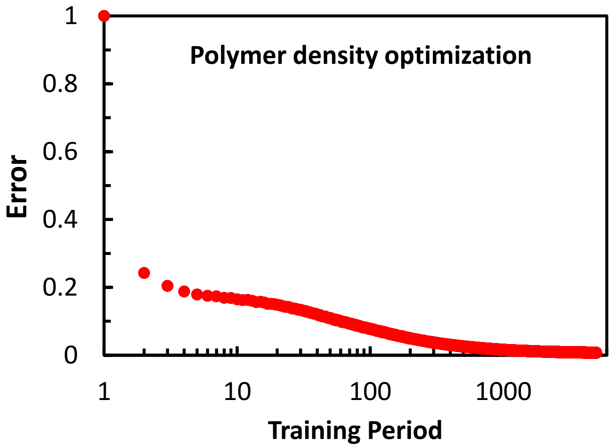

Figure 5: Error reduction during the training of the ANN. The ANN was trained with 2528 data points.

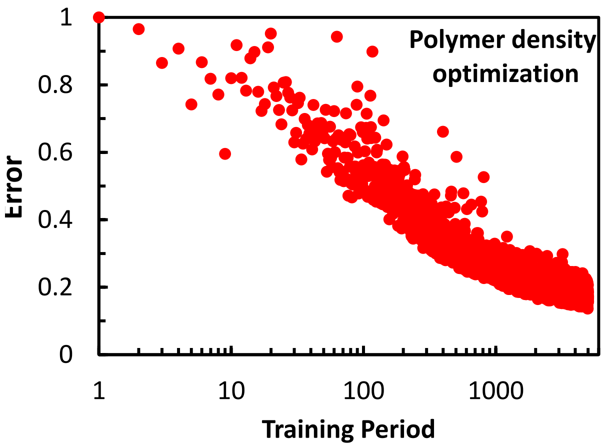

Figure 6: Prediction error calculated using unknown data set. The training and test data set consist of 2276 and 252 data points, respectively.

Designing high-performance copolymer compatabilizers

The topology of the ANN used for compatabilizer system is {20 30 30 1}. Figure 7 represents the error during the training of the network. The error in predicting the unknown data is shown as a function of the training period in the Figure 8. These results suggest that there is no overfitting during the training of the network.

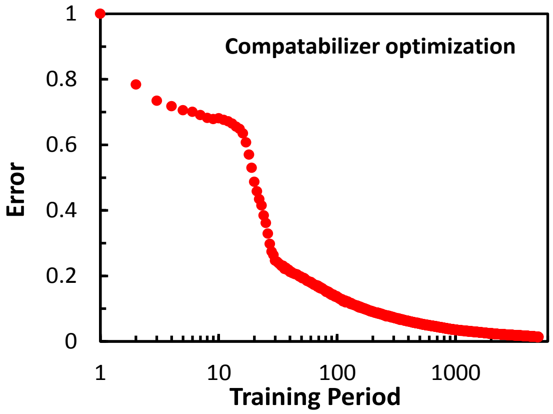

Figure 7: Error in training data set as a function of training period. Total training data points are 3232.

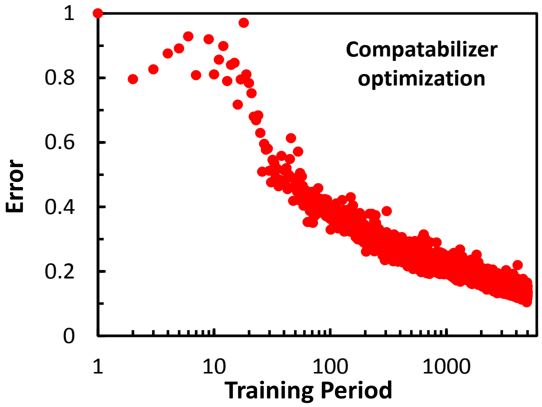

Figure 8: Prediction error calculated based on unknown data set. The total training and test data points are 2909 and 323, respectively.

References:

(1) Kremer, K.; Grest, G. S. Molecular Dynamics (MD) Simulations for Polymers. *J. Phys. Condens. Matter* **1990**, *2* (S), SA295-SA298.

(2) Simmons, D. S.; Douglas, J. F. Nature and Interrelations of Fast Dynamic Properties in a Coarse-Grained Glass-Forming Polymer Melt. *Soft Matter* **2011**, *7*, 11010–11020.

(3) Auhl, R.; Everaers, R.; Grest, G. S.; Kremer, K.; Plimpton, S. J. Equilibration of Long Chain Polymer Melts in Computer Simulations. *J. Chem. Phys.* **2003**, *119* (24), 12718–12728.

(4) Kalathi, J. T.; Kumar, S. K.; Rubinstein, M.; Grest, G. S. Rouse Mode Analysis of Chain Relaxation in Homopolymer Melts. *Macromolecules* **2014**, *47* (19), 6925–6931.

(5) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1–19.

(6) Tuckerman, M. E.; Alejandre, J.; López-Rendón, R.; Jochim, A. L.; Martyna, G. J. A Liouville-Operator Derived Measure-Preserving Integrator for Molecular Dynamics Simulations in the Isothermal–isobaric Ensemble. *J. Phys. Math. Gen.* **2006**, *39* (19), 5629.

(7) Harris, J. G. Liquid-Vapor Interfaces of Alkane Oligomers: Structure and Thermodynamics from Molecular Dynamics Simulations of Chemically Realistic Models. *J. Phys. Chem.* **1992**, *96* (12), 5077–5086.