Quantum Material Synthesis by Reinforcement Learning

Pankaj Rajak
Leadership Computing Facility
Argonne National Laboratory
Lemont, IL 60439
prajak@anl.gov

Aravind Krishnamoorthy
Department of Chemical Engineering
Materials Science
University of Southern California
Los Angeles, California 90089-0242
kris658@usc.edu

Aiichiro Nakano
Department of Computer Science
University of Southern California
Los Angeles, California 90089-0242
anakano@usc.edu

Rajiv Kalia
Department of Physics
Astronomy
University of Southern California
Los Angeles, California 90089-0242
rkalia@usc.edu

Priya Vashistha
Department of Chemical Engineering
Materials Science
University of Southern California
Los Angeles, California 90089-0242
priyav@usc.edu

Abstract

Designing conditions for experimental synthesis is the primary bottleneck for the realization of new functional quantum materials. Current strategies to synthesize new promising materials with desired properties are based upon a trial and error approach, which is a time consuming process and does not generalize to different materials. Here, we use deep reinforcement learning to learn synthesis schedules, which are time-dependent synthesis conditions of temperatures and reactant concentrations for a prototypical quantum material, monolayer MoS_2 via chemical vapor deposition (CVD). The reinforcement learning (RL) agent is coupled to a deep generative model that captures the probability density function of MoS_2-CVD dynamics and is trained on 10,000 computational synthesis simulations. After training, the RL agent successfully learns the optimal policy in terms of threshold temperatures and chemical potentials for the onset of chemical reactions and provides mechanistic insight to predict new synthesis schedules that produce well-sulfidized crystalline and phase-pure MoS_2 in minimum time, which is validated by reactive molecular dynamics.

1 Introduction

Advancement of technology based on promising new materials requires significantly shortening the current materials development timeline of $\sim$ 20 years [1]. This long timeline occurs due to the combination time needed to discover new candidate materials with desired set of properties from a vast search space and then identify scalable synthesis route for these materials. In recent years,

Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), Vancouver, Canada.

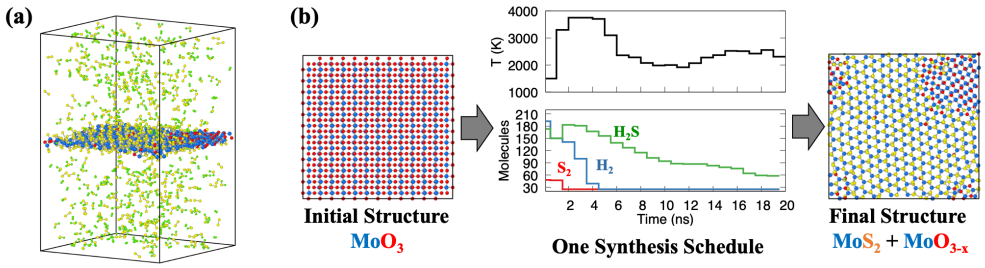

Figure 1: (a) Snapshot of RMD simulation cell consisting of MoO_xS_y slab in the middle, which is surrounded by sulfidizing atmosphere containing S_2, H_2 and H_2S gases. (b) Shows a $20ns$ synthesis schedule, where an initial MoO_3 slab at $t = 0$ reacts with time varying sulfidizing environment to generate MoS_2+MoO_{3-x}

significant development has been made to address the first challenge using data-driven material science, which combines machine learning (ML) methods with first-principal based simulations for accelerated discovery of new materials. The exponential growth in computational power combined with high-throughput simulation helped us to build rich material database, which is mined by ML models to discover new material [2, 3]. Successful example of this strategies has created new ultrahard materials, battery materials, polymers, organic solar cells, OLEDs, thermoelectrics etc [4, 5, 6].

Another important component in rapid materials development is the identification of experimental synthesis route of these promising materials and composition, which has not kept pace with the computational materials screening [7, 8]. Current strategies for material synthesis are based on trial and error approach and are largely based on empirical insight and materials intuition. There have been limited attempts at predicting the outcome of chemical reactions for the solution synthesis of small molecules using chemical insights and machine learning [9, 10]. However, synthesis planning for bulk inorganic materials and non-solution based quantum material is still in its infancy as they involve complicated time-correlation between synthesis parameters [11, 12]. This requires considerably more refined models than previous efforts which only considered the combination of reactants to learn the outcome of chemical reaction of molecular and organic systems [13, 14]. Efforts based on text-mining of published literature on synthesis schedule has been made to understand the effect of solvent concentrations, heating temperatures, processing times, and precursors on synthesis schedules of materials [15, 16]. However, even these upcoming ML techniques are limited both by the scarcity and sparsity of data in terms of existing schedules and synthesized materials and therefore their extension to new, potentially unknown materials [15].

In this work, we describe a reinforcement learning (RL) [17] scheme to optimize synthesis routes for a prototypical member of the family of 2D quantum material, MoS_2, via Chemical Vapor Deposition (CVD). CVD, a popular scalable technique for the synthesis of 2D materials, has numerous time-dependent parameters such as temperature, flow rates, concentration of gaseous reactants, and type of reaction precursors, dopants and substrates (together referred to as the synthesis profile) that need to be optimized [18]. Specifically, we use RL to (1) identify synthesis profiles that create material structure with desired properties, which in our case is maximum phase fraction of semiconducting crystalline phase of MoS_2 in shortest possible time, (2) effect of different RL policies on the quality of generated MoS_2 structure in terms of time dependent synthesis parameters and mechanistic insight of the synthesis process. Experimental synthesis is time-consuming and not suitable for high-throughput screening. Therefore, we have used reactive molecular dynamics (RMD) [19] to simulate CVD process that has previously shown to reproduce the potential energy surface of reacting system as well as capture important mechanisms of the CVD synthesis reaction [20, 21].

2 Method

2.1 Neural Autoregressive Density Estimation of CVD Dynamics

We perform RMD simulations to simulate synthesis of MoS_2 by CVD with a multi-step reaction of MoO_3 crystal in a sulfidizing atmosphere containing H_2S, S_2 and H_2 molecules. Each MD simulation

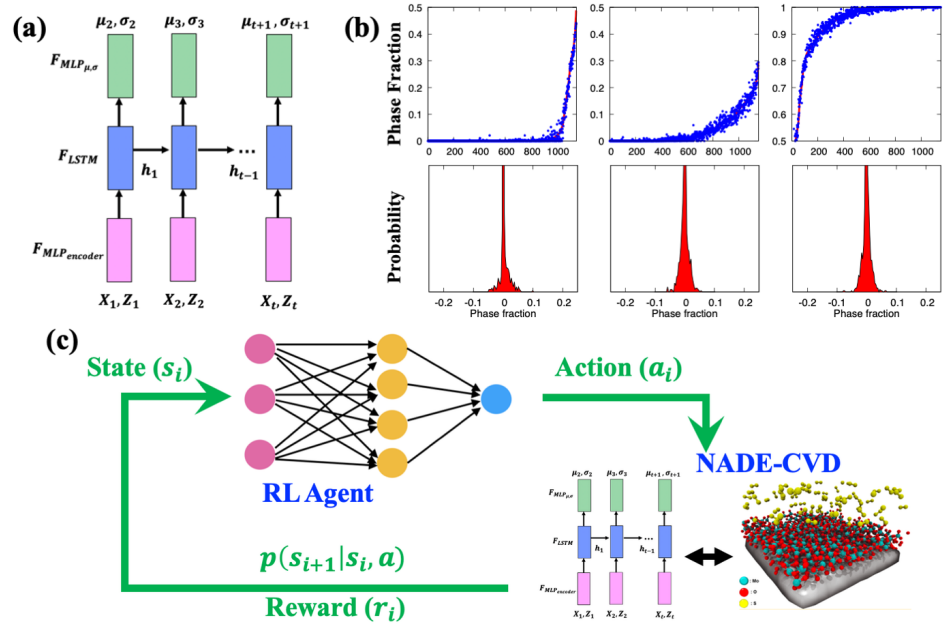

Figure 2: (a) Schematic of the NADE-CVD architecture consisting of fully connected networks as encoder ($F_{MLP_{encoder}}$) and decoder ($F_{MLP_{decoder}}$) and an intermediate lstm block (F_{LSTM}). (b) Shows accuracy of NADE-CVD on 1000 RMD simulation. (c) Shows schematics of the RL framework for the synthesis schedule for MoS_2 synthesis.

models a 20-ns long synthesis schedule, divided into 20 steps, each 1 ns long and characterized by 4 variables: the system temperature, and the number of S_2, H_2S and H_2 molecules in the reacting environment as $(T, n^{H_2}, n^{S_2}, n^{H_2S})$.The final output structure (MoS_2 + MoO_{3-x}) generated at the end is a non-trivial function of its synthesis schedule as shown in Figure 1. Here, each RMD simulation takes $\sim$ 3days, and thus it won't be possible to directly use them to train the RL agent. For this purpose, we have built a neural autoregressive density estimator (NADE-CVD) to approximate the probability density function of MoS_2-CVD dynamics in RMD simulation [22, 23, 24]. The probability density function of CVD dynamics, $P(X, Z)$, is characterized by two sets of random variables, which are (1) the observed variables, $X = X_{1:t_{max}}$, given by the user defined synthesis condition and (2) the unobserved variable $Z = Z_{1:t_{max}}$ given by the time dependent phase fraction of 2H, 1T and defect phases in the MoO_xS_y surface. Using conditional independence between variables and chain rule, $P(X, Z)$ is written as following autoregressive function:

$$P(Z|X, Z_1) = P(Z_2|Z_1, X_1) \dots P(Z_{(t+1)}|Z_{1:t}, X_{1:t}) \dots P(Z_T|Z_{1:t_{max}-1}, X_{1:t_{max}-1}) \quad (1)$$

where $X_t = (T_t, n_t^{H_2}, n_t^{S_2}, n_t^{H_2S})$ and $Z_t = (n_t^{2H}, n_t^{1T}, n_t^{defect})$. Here, NADE-CVD models each of the conditional probability, $P(Z_{t+1}|Z_{1:t}, X_{1:t})$, as a Gaussian distribution and output its parameters: mean $\mu_{t+1} = (\mu_{t+1}^{2H}, \mu_{t+1}^{1T}, \mu_{t+1}^{defect})$ and variance $\sigma_{t+1} = (\sigma_{t+1}^{2H}, \sigma_{t+1}^{1T}, \sigma_{t+1}^{defect})$. The architecture of NADE-CVD is given in Figure 2a, which consists of an encoder, LSTM cell and a decoder. The NADE-CVD is trained using 10,000 RMD simulation profile of 20 ns each using maximum likelihood estimate. Figure 2b shows the prediction error by NADE-CVD on test dataset of 1000 RMD simulations.

2.2 Reinforcement Learning for MoS_2 synthesis

Our RL workflow consists of an RL agent coupled with the NADE-CVD that serves as an environment of CVD synthesis, Figure 2c. The RL agent (π_θ) is represented using a fully connected neural network, where each episode of RL is of length 20 and difference between consecutive timestep is 1 ns that is equivalent to a 20 ns synthesis schedule. Each episode of RL starts from an arbitrary synthesis condition, $(T^0, S_2^0, H_2^0, H_2S^0)$, and a MoO_3 crystal, where the goal of the RL agent is to maximize

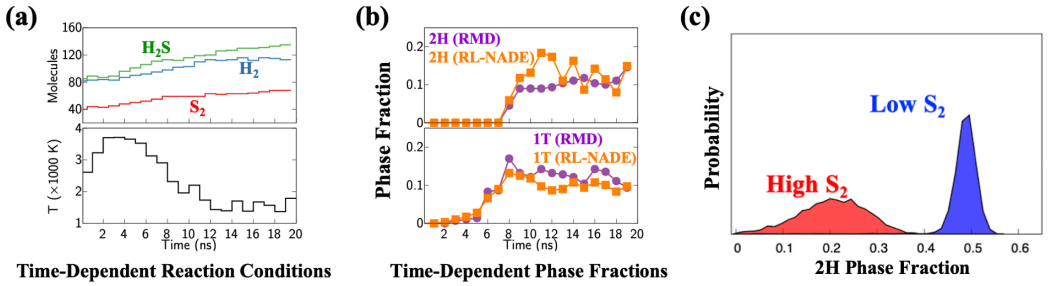

Figure 3: a) Shows a RL-agent generated 20 ns synthesis profile and (b) corresponding phase fraction of 2H and 1T phases in the MoS_2+MoO_{3-x} predicted by NADE and validated by RMD simulations using this synthesis profile. (c) Probability density function of 2H phase under different synthesis schedule proposed by RL agent where each schedule has either low initial S_2 conc. or high.

the following objective function:

$$\arg\max_{\theta} \mathbb{E}_{\tau\sim\pi_\theta}[\sum_{t=1}^{t=t_{max}} r(s_t,a_t)], r(s_t,a_t) = \begin{cases} 0, & \text{if } Z = Z_{t+1}[n_{2H}] < 0.40 \\ 0.2Z_{t+1}, & Z_{t+1}[n_{2H}] >= 0.40 \end{cases} \quad (2)$$

The above objective function is equivalent to proposing actions that convert the initial MoO_3 crystal into 2H-MoS_2 structure with maximum 2H phase fraction in minimum time.

During each timestep t in an episode , the input (s_t) to (π_θ) is a 128-dimension embedding vector of the entire simulation history till t, $(Z_{1:t}, X_{1:t})$, and the agent proposes an action $a_t = \Delta X = (\Delta T, \Delta S_2, \Delta H_2, \Delta H_2 S)$, which is the change in synthesis condition (i.e. reaction temperature and gas concentrations). Here, the action a_t to take at s_t is modeled using a Gaussian distribution $((a_t \sim \mathcal{N}(\mu(s_T), \sigma^2))$, whose parameters $\mu(s_t)$ – state dependent mean – is the output of the RL agent, $\mu(s_t) = \pi_\theta(s_t)$. The variance, σ^2 is assumed to be constant and is tuned as a hyper-parameter of the RL scheme. After that, the synthesis condition for the next timestep is defined as $X_{t+1} = X_t + a_t$. Using this, NADE-CVD predicts the distribution of various phases in the synthesized product Z_{t+1} for next timestep and provides a new state s_{t+1} and reward $r_t(s_t, a_t)$ to the RL agent. For training the RL agent, we use policy gradient along with value function $V(s_t)$ as baseline [25, 26, 27].

3 Result

After training for 15,000 episodes, the RL agent learns synthesis policies composed of time-dependent temperatures, and concentrations of H_2S, S_2 and H_2 molecules optimized to synthesize 2H-rich MoS_2 structures in least time. Closer inspection of these polices provides mechanistic insights into the CVD process and the effect of variations of synthesis condition on the quality on final structure. Figure 3a shows one such policy proposed by the RL agent, which consists of an early high-temperature (>3000 K) phase spanning the first 7-10 ns followed by annealing to an intermediate temperature (∼2000 K) for the reminder of the synthesis profile. This strategy is consistent with atomistic simulation-based synthesis of MoS_2, where high temperature (>3000 K) is necessary for both the reduction of MoO_3 surface and its sulfidation, and the subsequent lower temperature (∼2000 K) is necessary for enabling crystallization in the 2H structure, while continuing to promote residual sulfidation. Figure 3b shows the validation of this RL-generated profile using NADE-CVD and RMD simulation, which shows both the NADE-CVD prediction and the actual RMD simulation results are in close agreement. We further note that this synthesis schedule captures a non-trivial mechanistic detail about phase evolution during synthesis – the nucleation of the 1T phase precedes the nucleation of the 2H crystal structure, which was also previously observed in MoS_2 synthesis simulations [20].

Another important phenomenon identified by RL agent is the effect of initial gas concentration on the quality of the final synthesized material. In Figure 3c, we compute the phase fraction of 2H phase in the synthesized MoS_2 product over the last 10 ns of the simulation for 3200 synthesis conditions proposed by the RL agent under different initial gas concentrations. Here, higher mean of the probability distribution function provides an indication of the extent of sulfidation as well as the

time required to generate 2H phases. Figure 3c shows that RL agent recommends low concentration of S_2 at early stages (0-3 ns) of the synthesis, when the temperature is high. This partially evacuated synthesis atmosphere with low S_2 concentration promotes the evolution of oxygen from and self-reduction of the MoO_3 surface that helps to generate in a significantly higher 2H phase fraction in the synthesized product.

4 Conclusion

We have developed a reinforcement learning scheme for the predictive synthesis of two-dimensional MoS_2 monolayers using chemical vapor deposition. The RL model successfully proposed several new reaction schedules, i.e. time-dependent reaction conditions, to synthesize MoS_2 with maximum crystallinity and phase fraction of the 2H semiconducting structure. More importantly, the RL model also provides mechanistic insight into the material synthesis process and and an understanding of the role of synthesis conditions (temperature, chemical environment) on the quality of the synthesized crystal. This RL scheme provides the first viable high-throughput approach to screening material synthesis conditions to tackle the as-yet unsolved problem of predictive synthesis of novel nanomaterials.

Broader Impact

While the RL scheme described here is implemented on simulated chemical vapor deposition, the scheme is highly generalizable to any other synthesis technique that relies on time-dependent reaction conditions, including sputtering, pulsed laser deposition, flame synthesis, vapor transport etc. The model can also be used to predict synthesis schedules to yield more complex products, including heterostructures containing interfaces between multiple phases. By modifying the reward function used in the RL model to capture functional properties (and not just structure/phases) of the synthesized product, the RL scheme can be directly used to identify promising synthesis schedules to fabricate materials with desired properties. This provides an complementary framework to existing materials discovery paradigm based on high-throughput *ab initio* calculations.

Acknowledgements

This work was supported as part of the Computational Materials Sciences Program funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award Number DE-SC0014607. This research was partly supported by Aurora Early Science programs and used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

References

[1] M. L. Green, C. L. Choi, J. R. Hattrick-Simpers, A. M. Joshi, I. Takeuchi, S. C. Barron, E. Campo, T. Chiang, S. Empedocles, J. M. Gregoire, A. G. Kusne, J. Martin, A. Mehta, K. Persson, Z. Trautt, J. Van Duren, and A. Zakutayev. Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies. *Applied Physics Reviews*, 4(1):011105, 2017.

[2] Alex Zunger. Inverse design in search of materials with target functionalities. *Nature Reviews Chemistry*, 2:0121, 03 2018.

[3] Keith Butler, Daniel Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. Machine learning for molecular and materials science. *Nature*, 559, 07 2018.

[4] Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, Timothy Hirzel, David Duvenaud, Dougal Maclaurin, Martin Blood-Forsythe, Hyun Chae, Markus Einzinger, Dong-Gwang Ha, Tony Wu, Georgios Markopoulos, Soonok Jeon, Hosuk Kang, Hiroshi Miyazaki, Masaki Numata, Sunghan Kim, Wenliang Huang, Seong Hong, Marc Baldo, and Alán Aspuru-Guzik. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature Materials*, 15, 08 2016.

[5] Rampi Ramprasad, Rohit Batra, Ghanshyam Pilania, Arun Mannodi-Kanakkithodi, and Chiho Kim. Machine learning and materials informatics: Recent applications and prospects. *npj Computational Materials*, 3, 07 2017.

[6] Lindsay Bassman, Pankaj Rajak, Rajiv Kalia, Aiichiro Nakano, Fei Sha, Jifeng Sun, David Singh, Muratahan Aykol, Patrick Huck, Kristin Persson, and Priya Vashishta. Active learning for accelerated design of layered materials. *npj Computational Materials*, 4, 12 2018.

[7] Juan de Pablo, Nicholas Jackson, Michael Webb, Long-Qing Chen, Joel Moore, Dane Morgan, Ryan Jacobs, Tresa Pollock, Darrell Schlom, Eric Toberer, James Analytis, Ismaila Dabo, Dean DeLongchamp, Gregory Fiete, Gregory Grason, Geoffroy Hautier, Yifei Mo, Krishna Rajan, Evan Reed, and Ji-Cheng Zhao. New frontiers for the materials genome initiative. *npj Computational Materials*, 5:41, 04 2019.

[8] Qian Yang, Carlos Sing-Long, and Evan Reed. Learning reduced kinetic monte carlo models of complex chemistry from molecular dynamics. *Chem. Sci.*, 8, 06 2017.

[9] Zhenpeng Zhou, Xiaocheng Li, and Richard N. Zare. Optimizing chemical reactions with deep reinforcement learning. *ACS Central Science*, 3(12):1337–1344, 2017.

[10] Connor Coley, Dale Thomas, Justin Lummiss, Jonathan Jaworski, Christopher Breen, Victor Schultz, Travis Hart, Joshua Fishman, Luke Rogers, Hanyu Gao, Robert Hicklin, Pieter Plehiers, Joshua Byington, John Piotti, William Green, A. Hart, Timothy Jamison, and Klavs Jensen. A robotic platform for flow synthesis of organic compounds informed by ai planning. *Science*, 365:eaax1566, 08 2019.

[11] Daniel Tabor, Loïc Roch, Semion Saikin, Christoph Kreisbeck, Dennis Sheberla, Joseph Montoya, Shyam Dwaraknath, Muratahan Aykol, Carlos Ortiz, Hermann Tribukait, Carlos Amador-Bedolla, Christoph Brabec, Benji Maruyama, Kristin Persson, and Alán Aspuru-Guzik. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nature Reviews Materials*, page 1, 04 2018.

[12] Paul Raccuglia, Katherine Elbert, Philip Adler, Casey Falk, Malia Wenny, Aurelio Mollo, Matthias Zeller, Sorelle Friedler, Joshua Schrier, and Alexander Norquist. Machine-learning-assisted materials discovery using failed experiments. *Nature*, 533:73–76, 05 2016.

[13] Connor W. Coley, Regina Barzilay, Tommi S. Jaakkola, William H. Green, and Klavs F. Jensen. Prediction of organic reaction outcomes using machine learning. *ACS Central Science*, 3(5):434–443, 2017.

[14] Jennifer Wei, David Duvenaud, and Alán Aspuru-Guzik. Neural networks for the prediction organic chemistry reactions. *ACS Central Science*, 2, 08 2016.

[15] Edward Kim, Kevin Huang, Stefanie Jegelka, and Elsa Olivetti. Virtual screening of inorganic materials synthesis parameters with deep learning. *npj Computational Materials*, 3, 12 2017.

[16] Olga Kononova, Haoyan Huo, Tanjin He, Ziqin Rong, Tiago Botari, Wenhao Sun, Vahe Tshitoyan, and Gerbrand Ceder. Text-mined dataset of inorganic materials synthesis recipes. *Scientific Data*, 6, 12 2019.

[17] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei Rusu, Joel Veness, Marc Bellemare, Alex Graves, Martin Riedmiller, Andreas Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518:529–33, 02 2015.

[18] Gangtae Jin, Chang-Soo Lee, Xing Liao, Juho Kim, Zhen Wang, Odongo Okello, Bumsu Park, Jaehyun Park, Cheolhee Han, Hoseok Heo, Jonghwan Kim, Sang Oh, Si-Young Choi, Hongkun Park, and Moon-Ho Jo. Atomically thin three-dimensional membranes of van der waals semiconductors by wafer-scale growth. *Science Advances*, 5:eaaw3180, 07 2019.

[19] Ken-ichi Nomura, Rajiv Kalia, Aiichiro Nakano, and Priya Vashishta. A scalable parallel algorithm for large-scale reactive force-field molecular dynamics simulations. *Computer Physics Communications*, 178:73–87, 01 2008.

[20] Sungwook Hong, Ken-ichi Nomura, Aravind Krishnamoorthy, Pankaj Rajak, Chunyang Sheng, Rajiv Kalia, Aiichiro Nakano, and Priya Vashishta. Defect healing in layered materials: A machine learning-assisted characterization of MoS_2 crystal-phases. *Journal of Physical Chemistry Letters*, 05 2019.

[21] Sungwook Hong, Aravind Krishnamoorthy, Pankaj Rajak, Subodh Tiwari, Masaaki Misawa, Fuyuki Shimojo, Rajiv Kalia, Aiichiro Nakano, and Priya Vashishta. Computational synthesis of MoS_2 layers by reactive molecular dynamics simulations: Initial sulfidation of MoO_3 surfaces. *Nano Letters*, 17, 07 2017.

[22] Benigno Uria, Marc-Alexandre Côté, Karol Gregor, Iain Murray, and Hugo Larochelle. Neural autoregressive distribution estimation, 2016. arXiv:1605.02226.

[23] Karol Gregor, Ivo Danihelka, Andriy Mnih, Charles Blundell, and Daan Wierstra. Deep autoregressive networks, 2014. arXiv:1310.8499.

[24] Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks, 2016. arXiv:1601.06759.

[25] John Schulman, Philipp Moritz, Sergey Levine, Michael Jordan, and Pieter Abbeel. High-dimensional continuous control using generalized advantage estimation, 2018. arXiv:1506.02438.

[26] Richard Sutton, David Mcallester, Satinder Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. *Adv. Neural Inf. Process. Syst*, 12, 02 2000.

[27] Yan Duan, Xi Chen, Rein Houthooft, John Schulman, and Pieter Abbeel. Benchmarking deep reinforcement learning for continuous control, 2016. arXiv:1604.06778.