

Applications of Data Science to Molecules and Materials 2021

Topical Conference at the 2021 AIChE Annual Meeting

Boston, Massachusetts, USA and Online
7 - 11 November and 15 - 19 November 2021

ISBN: 978-1-7138-5689-4

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2021) by AIChE
All rights reserved.

Printed with permission by Curran Associates, Inc. (2022)

For permission requests, please contact AIChE
at the address below.

AIChE
120 Wall Street, FL 23
New York, NY 10005-4020

Phone: (800) 242-4363
Fax: (203) 775-5177

www.aiche.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

Understanding Barrierless Adsorption in Electro-Catalytic Systems Using Accelerated Transition Path Sampling.....	1
<i>Mayank Agrawal, Adam P. Willard, Andrew A. Peterson</i>	
Methods and Applications of Ex-Machina Molecular Dynamics to Catalytic Processes.....	2
<i>Boris Kozinsky</i>	
Statistical Models for Predicting Oil Composition from Hydrothermal Liquefaction of Biomass.....	3
<i>Seshasayee Mahadevan Subramanya, Nicholas Rios, Abbey J. Kollar, Rachel Stofanak, Katherine Maloney, Kayley. E. Waltz, Chinmayee Rane, Sandeep Endluri, Phillip E. Savage</i>	
A Big Data Approach to Predict the Kinetics of Organic Chemistry Reactions.....	5
<i>Stephanie Valteau</i>	
Uncertainty Quantification of Catalyst Structure Effects on Kinetics	6
<i>Xue Zong, Dionisios Vlachos</i>	
Open Catalyst Project: Advancements and Challenges in Building a Flexible Machine Learning Potential.....	7
<i>Muhammed Shuaibi, Lowik Chanussot, Siddharth Goyal, Thibaut Lavril, Abhishek Das, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, Aini Palizhati, Anuroop Sriram, Brandon Wood, Junwoong Yoon, Devi Parikh, C. Lawrence Zitnick, Zachary Ulissi</i>	
Machine Learning Accelerated Scale-Up for Microporous Materials	8
<i>Di Du, Anna Ivashko, Preeti Kamakoti, Stuart L. Soled</i>	
Physics-Informed Neural Networks for Kinetic Parameter Estimation and Uncertainty Quantification.....	9
<i>Gabriel Gusmão, Adhika Retnanto, Shashwati da Cunha, Matthew Ross Kunz, Andrew Medford</i>	
Deep Learning Quantum Reaction Rate Constants	10
<i>Stephanie Valteau</i>	
Building Blocks for Autonomous Computing Materials: Dimers, Trimers and Tetramers.....	11
<i>Xingfei Wei</i>	
A Novel Machine Learning Model for the Accurate Design of Highly Selective Zeolitic-Imidazolate Frameworks	12
<i>Ioannis Economou, Panagiotis Krokidas, Stelios Karozis, George Giannakopoulos, Michael Kainourgiakis, Theodore Steriotis</i>	
Designing and Synthesizing Novel Dye Molecules Using Generative Modeling and Data-Driven Synthesis Planning	14
<i>Camille Bilodeau, Brent Koscher, Kevin P. Greenman, Rafael Gómez-Bombarelli, Klavs F. Jensen</i>	
Predicting Critical Micelle Concentrations for Surfactants Using Graph Convolutional Neural Networks	15
<i>Shiyi Qin, Tianyi Jin, Reid Van Lehn, Victor M. Zavala</i>	

Cript: Establishing and Harnessing a Polymer Database.....	17
<i>Dylan Walsh, Tzyy-Shyang Lin, Chris Borg, Eric Muckley, Vinay Hegde, Ken Kroenlein, Klavs F. Jensen, Bradley Olsen</i>	
Predicting Adhesive Free Energies of Polymer-Surface Interactions with Machine Learning	18
<i>Jiale Shi, Michael Quevillon, Jonathan K. Whitmer</i>	
Virtual High-Throughput Screening of Vapor-Deposited Amphiphilic Polymers for Biofilm Reduction with Machine/Deep Learning.....	19
<i>Zhihao Feng, Yifan Cheng, Andrew L. Hook, Rong Yang, Jeffrey D. Varner</i>	
High-Throughput and Data-Driven Strategies for the Design of Deep Eutectic Solvent Electrolytes.	21
<i>Jaime Rodriguez Jr., Maria Politi, Lilo Pozzo</i>	
AI-Guided Autonomous Flow Reactor Platform for Accelerated Nanomaterial Synthesis Screening and Parameter Space Mapping	22
<i>Ajit Vikram, Arwa Zahid, Paul Kenis</i>	
Machine Learning Models for Rapid Compositional Quantification of Complex Multicomponent Mixtures Using Vibrational Spectroscopy.....	23
<i>Andrea Angulo Figueira, Lankun Yang, Eray Aydil, Miguel Modestino</i>	
Are Structure-Function Relationships Governing Polymeric Gene Delivery Payload-Specific?	24
<i>Ramya Kumar, Ngoc Le, Mary Brown, Theresa M. Reineke</i>	
Enhancing Insight to Individual and Population-Based Microglial Reactivity with Image Analysis.....	25
<i>Hawley Helmbrecht, Elizabeth Nance</i>	
High-Affinity Non-Immunoglobulin Scaffolds Promoted by Sequence Mining Impactful in High-Value Therapeutics.....	28
<i>Mehrsa Mardikoraem, Joelle Eaves, Daniel Woldring</i>	
The Euler Characteristic: A General Topological Descriptor for Complex Data.....	29
<i>Alexander Smith, Victor M. Zavala</i>	
Machine Learning + Automated Reasoning for Theory Discovery.....	32
<i>Tyler Josephson, Vernon Austel, Kenneth A. Clarkson, Cristina Cornelio, Sanjeeb Dash, Bachir El Khadir, Joao Goncalves, Lior Horesh, Sharon Liu, Nimrod Megiddo, Francesca Nacion, Neil Tran, Catherine M. Wraback</i>	
Characterizing Uncertainty and Error in Machine Learning Chemical Property Prediction	33
<i>Esther Heid, Charles J. McGill, Florence Vermeire, William Green</i>	
Giving Attention to Generative Models for De Novo Molecular Design.....	34
<i>Orion Dollar, Nisarg Joshi, David Beck, Jim Pfaendner</i>	
Improving Data Sub-Selection for Supervised Tasks with Principal Covariates Regression.....	35
<i>Rose Cersonsky, Benjamin Helfrecht, Sergei Kliavinek, Edgar A. Engel, Michele Ceriotti</i>	
Chemical Identification in Multicomponent Electrolytes Using Voltammetry, Physics-Based Modeling, and Bayesian Inference	36
<i>Alexis Fenton Jr., Fikile R. Brushett</i>	
Deep Learning-Assisted Analysis of Anomalous Nanoparticle Surface Diffusion in Liquid Phase Transmission Electron Microscopy	38
<i>Vida Jamali, Cory Hargus, Assaf Ben-Moshe, Hyun Dong Ha, Kranthi K. Mandadapu, A. Paul Alivisatos</i>	

A Principal Component Normalizing Flow for Modeling Renewable Electricity Generation	39
<i>Eike Cramer, Alexander Mitsos, Manuel Dahmen</i>	
Lessons and Opportunities in Data-Driven High Throughput Experimentation	41
<i>Lilo Pozzo, Jaime Rodriguez Jr., Sage Scheiwiller, Kacper Lachowski, Maria Politi</i>	
Automating Discovery: Accelerating the Search for Novel Perovskites with Robots and Data	42
<i>Joshua Schrier</i>	
Streamlining Small Molecule Design and Synthesis	44
<i>Connor Coley</i>	
Building Catalytic Descriptors with Iterative Bayesian Additive Regression Trees (iBART)	45
<i>Chun-Yen Liu, Shengbin Ye, Meng Li, Thomas Senftle</i>	
Unsupervised Machine Learning to Extract the Electronic and Chemical Properties of Alloy and Metal Oxide Surfaces	46
<i>Jacques Esterhuizen, Bryan Goldsmith, Suljo Linic</i>	
Implementing and Accelerating Autonomous Closed-Loop Control	47
<i>Bryan Pope</i>	
State of Art in Modeling and Modeling Deployment for the Process and Pharma Industries; From Advanced Simulations to Web Apps	48
<i>Hossam Metwally</i>	
Analytical Methods to Improve Diagnostic Protocols Using Infrared Spectroscopic Imaging	49
<i>Shachi Mittal, Jonathan Kim, Rohit Bhargava</i>	
Message Passing Neural Networks for Prediction of IR Spectra	51
<i>Charles J. McGill, Michael Forsuelo, Yanfei Guan, William Green</i>	
Transfer Learning for Prediction of Absorption and Emission Spectra from Multi-Fidelity Data	52
<i>Kevin P. Greenman, William Green, Rafael Gomez-Bombarelli</i>	
Multi-Task Property Prediction: Importance of the “Chemist-In-The-Loop” in Model Building	53
<i>Adem Rosenkvist Nielsen Aouichaoui, Seyed Soheil Mansouri, Jens Abildskov, Gürkan Sin</i>	
Specific Ion Effects in Aqueous Electrolyte Solutions from First Principles-Derived Machine- Learning Potentials	55
<i>Shuwen Yue, Athanassios Panagiotopoulos</i>	
Combining Molecular Dynamics Simulations and Active Learning to Study the Hydrophobicity of Chemically Heterogeneous Surfaces	56
<i>Atharva Kelkar, Bradley C. Dallin, Reid Van Lehn</i>	
Machine Learning the COSMO Model for Predicting Thermodynamics of Electrolyte Mixtures	57
<i>Eric Fonseca, Ashwin Ravichandran, Richard G. Hennig, John W. Lawson</i>	
Generating 3D Conformer Ensembles of Molecular Graphs with End-To-End Differentiable Networks	58
<i>Lagnajit Pattanaik, Octavian Ganea, Connor Coley, Regina Barzilay, Tommi S. Jaakkola, Klavs F. Jensen, William Green</i>	
Graph Invariants as Rapid Features for Classification and Regression of Bulk Hydrogen Intercalation Energies in Reducible Metal Oxides	59
<i>Evan Miu, James R. McKone, Giannis Mpourmpakis</i>	

Something Old, Something New: Generative Adversarial Approaches to Conditional Sampling in Complex System Simulations.....	60
<i>Juan Bello-Rivas, Ellis Crabtree, Andrew Ferguson, Ioannis G. Kevrekidis</i>	
Teaching Process Data Analytics and Machine Learning	62
<i>Richard D. Braatz, Weike Sun, Brian W. Anthony</i>	
Teaching Big Data Science to Undergraduate Students in the University of Iowa	63
<i>Jun Wang, Joseph S. Gomes, Charles Stanier</i>	
Textile: Tutorials in Experimentalist Interactive Learning.....	64
<i>Hawley Helmbrecht, Elizabeth Nance</i>	
Teaching Artificial Intelligence to Chemical Engineers: Experiences from a 35-Year-Old Course	67
<i>Venkat Venkatasubramanian</i>	
Deriving Descriptors and Physical Understanding from Data in Catalysis Research.....	68
<i>Sergey Levchenko</i>	
Inverse Design of Nanoparticles Using Charge Transfer Properties and Multi-Target Machine Learning	69
<i>Sichao Li, Amanda Barnard</i>	
Design of Optoelectronic Self-Assembling π -Conjugated Peptides Via Combined Experimental-Computational Multi-Fidelity Active Learning	77
<i>Kirill Shmilovich, Sayak S. Panda, John D. Tovar, Andrew Ferguson</i>	
A Data-Driven Approach to Understanding and Predicting the Early Formation of the Solid-Liquid Electrolyte Interphase.....	78
<i>Kristin Persson</i>	
Building a Universal Data Science Workflow for Monodentate Organophosphorus Ligands	79
<i>Gabriel dos Passos Gomes, Tobias Gensch, Pascal Friederich, Ellyn Peters, Theophile Gaudin, Robert Pollice, Kjell Jorner, AkshatKumar Nigam, Michael Lindner-D'Addario, Matthew Sigman, Alan Aspuru-Guzik</i>	
From Brownian Dynamics Simulations and Experimental Observations of Colloidal Suspensions to Data-Driven Observables and Effective Sdes with Manifold Learning	81
<i>Nikolaos Evangelou, Noah Wichrowski, Ellis Crabtree, Felix Dietrich, Juan Bello-Rivas, Alexander Yeh, Jianli Zhang, Michael A. Bevan, Ioannis G. Kevrekidis</i>	
A Data-Driven Approach to Disentangled Parametrization of Relations, with an Application to Multisite Phosphorylation	83
<i>Nikolaos Evangelou, Noah J. Wichrowski, George Kevrekidis, Felix Dietrich, Mahdi Kooshkbaghi, Sarah McFann, Stanislav Y. Shvartsman, Ioannis G. Kevrekidis</i>	
Predicting Chemical Adsorption onto Montmorillonite Clays Through a Combination of Experimental Isotherms, Minimalistic Simulations and Data-Driven Models	85
<i>Asuka Orr, Meichen Wang, Burcu Beykal, Hari S. Ganesh, Sara E. Hearon, Efstratios N. Pistikopoulos, Timothy D. Phillips, Phanourios Tamamis</i>	
Machine Learning Electronic Structure for Heterogeneous Catalysis	87
<i>Victor Fung, Bobby G. Sumpter, Panchapakesan Ganesh</i>	

Machine Learning the Quantum-Chemical Properties of Metal-Organic Frameworks for Accelerated Materials Discovery with a New Electronic Structure Database	88
<i>Andrew Rosen, Shaelyn Iyer, Debmalya Ray, Zhenpeng Yao, Alan Aspuru-Guzik, Laura Gagliardi, Justin Notestein, Randall Snurr</i>	
Bayesian Parameter Estimation for Connecting Theory & Predictions to Experiments.....	90
<i>Aditya Savara</i>	
Towards Development of Novel Heterogeneous Catalysts Using Extrapolative Machine Learning Methods	91
<i>Takashi Toyao, Ken-ichi Shimizu</i>	

Author Index