Computational Molecular Science and Engineering Forum
for the combined community of engineers and scientists developing and applying molecularly based theories, modeling, and simulation
http://comsef.org/

Newsletter

AIChE

June 2022

In This Issue

- Election Results
- Grad Student Awards
- Research Highlight
- A Broader View
- Spotlight on the Web
- i-CoMSE
- Conferences
- Where Are They Now?
- Why CoMSEF?

CoMSEF Executive Committee Elections

Congratulations to Yamil Colón (Notre Dame) and Reid Van Lehn (Wisconsin) who were elected as Liaison Directors to the CoMSEF Executive Committee in the fall of 2021! Thanks to Heather Kulik (MIT) and Jindal Shah (Oklahoma St.) who completed their terms on the executive committee.

Two Liaison Directors are elected each year and serve two-year terms. Their responsibilities include identifying opportunities for co-sponsorship and communicating and advocating CoMSEF activities with other organizations. Liaison Directors also aid the other officers in developing and carrying out CoMSEF activities and preparing the CoMSEF newsletter.

Call for Nominations

****Graduate Student Awards in Computational Molecular Science and Engineering****

AIChE's Computational Molecular Science and Engineering Forum (CoMSEF) graduate student awards recognize excellence in research by graduate students. The intent of the awards is to reward significant contributions to research in computational molecular science and engineering by students. The award consists of a certificate and an honorarium.

Nominations should consist of a **nominating letter** from the student's research advisor and the **curriculum vitae** of the nominee. These should be sent by the advisor via e-mail in pdf format to the CoMSEF co-Chair (co-chair@comsef.org) by **October 1.** The student should have already submitted an abstract to the CoMSEF poster session at the AIChE annual meeting.

In addition, nominees must **present a poster** at the CoMSEF Poster session at the AIChE annual meeting. The nominee must be a **graduate student** at the time of the poster presentation, and **the faculty nominator must be a member of CoMSEF**. The winners will be selected by a committee composed of CoMSEF officers based on the student's CV, the nomination letter from the advisor (who must be a **member of CoMSEF**), and the quality of the poster presentation.

2021 CoMSEF Graduate Student Awards

The CoMSEF Graduate Student Awards in Computational Molecular Science and Engineering were awarded following the annual AIChE meeting. The awards recognize excellence in research by graduate students in the field of computational molecular science and engineering. Two awardees were selected based on the nomination letters received from each student's advisor, their CV, and a poster presented at the CoMSEF poster session:

- **Michael Cowan** (University of Pittsburgh, Advisor: Giannis Mpourmpakis)
- **Chenru Duan** (MIT, Advisor: Heather Kulik)

Scenes from the AIChE Annual Meeting 2021

Note: Opinions presented in the newsletter are the authors' and do not necessarily represent the view of the CoMSEF organization.

Research Highlight: Molecular Music

Yamil J. Colón, Department of Chemical and Biomolecular Engineering, University of Notre Dame

Generative models are an exciting class of machine learning (ML) algorithms that have greatly impacted the computational molecular design field. Variational autoencoders (VAEs), for example, can transform string-based molecular representations like SMILES[1] and SELFIES[2] into a latent variable space and back.[3-6] The advantage of this type of approach is that it converts discrete variables from the molecular description (atom identity, bonding information, etc.) into a continuous variable. This conversion enables interpolations and gradient-based optimizations along that continuous space. They have been used with great success in the context of drug discovery, energy applications, and others.[7-9] Other approaches augment the dimensions of the string-based representations through the development of fingerprints, which have also been used in various ML algorithms with great success.[10] The drawback of these approaches is that chemical intuition can be lost in these transformations and recovering the chemical structure from the latent space or fingerprint is not always straightforward and can sometimes lead to unrealistic chemical structures.[11, 12]

A recent preprint by Mahjour and coworkers entitled "Molecular Sonification for Molecule to Music Information Transfer" demonstrates an alternative molecular encoding that is literal music to one's ears.[13] In their work, the authors show how music can be used to store molecular information. There have been previous examples of chemical sonification looking at amino acid sequence and IR spectra.[14, 15] This work opens the door to using music generation artificial intelligence (AI) algorithms to explore chemical spaces. In broad strokes, the Sonic Architecture for Molecule Production and Live-Input Encoding Software (SAMPLES) produces melody from a molecular structure where the key encodes a physicochemical property and the sequence of notes is determined by SELFIES. It can turn chemical structures into music, and it also performs the reverse operation; in their paper, they show the molecule corresponding to "Twinkle, Twinkle, Little Star." The authors show how combinations of different molecules yield musical scores resembling the addition of the notes from the individual molecules. Likewise, molecules that bear resemblance in terms of their chemical structure also have similar melodies. While it is clear the changes in the chemical domain affect the music, the authors also demonstrate how changes in the music domain affect the chemistry. It is remarkable how a single note change alters the resulting chemical structure while retaining a clear chemical relationship between the molecules. In addition to storing the information in the musical dimension space, the framework allows for molecular generation in a latent space. They use MusicVAE[16] to interpolate between two melodies corresponding to different molecules. The resulting musical interpolation was then successfully converted into a new chemical structure. This type of blending procedure is well-established for music generation and was successfully demonstrated when the music is directly tied to the chemistry. Importantly, this type of framework can potentially be used to increase accessibility by allowing blind scientists to audibly interact with the chemistry.

References

1. Weininger, D., SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* **1988,** *28* (1), 31-36.
2. Krenn, M.; Häse, F.; Nigam, A.; Friederich, P.; Aspuru-Guzik, A., Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation. *Machine Learning: Science and Technology* **2020,** *1* (4), 045024.
3. Nigam, A.; Pollice, R.; Krenn, M.; Gomes, G. d. P.; Aspuru-Guzik, A., Beyond generative models: superfast traversal, optimization, novelty, exploration and discovery (STONED) algorithm for molecules using SELFIES. *Chemical Science* **2021,** *12* (20), 7079-7090.
4. Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A., Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science* **2018,** *4* (2), 268-276.
5. Kadurin, A.; Nikolenko, S.; Khrabrov, K.; Aliper, A.; Zhavoronkov, A., druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Molecular pharmaceutics* **2017,** *14* (9), 3098-3104.
6. Zhumagambetov, R.; Kazbek, D.; Shakipov, M.; Maksut, D.; Peshkov, V. A.; Fazli, S., cheML. io: an online database of ML-generated molecules. *RSC Advances* **2020,** *10* (73), 45189-45198.
7. Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W., Deep learning for molecular design—a review of the state of the art. *Molecular Systems Design & Engineering* **2019,** *4* (4), 828-849.
8. Sanchez-Lengeling, B.; Aspuru-Guzik, A., Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018,** *361* (6400), 360-365.
9. Blaschke, T.; Olivecrona, M.; Engkvist, O.; Bajorath, J.; Chen, H., Application of Generative Autoencoder in De Novo Molecular Design. *Molecular Informatics* **2018,** *37* (1-2), 1700123.
10. Rogers, D.; Hahn, M., Extended-connectivity fingerprints. *Journal of chemical information and modeling* **2010,** *50* (5), 742-754.
11. Arús-Pous, J.; Johansson, S. V.; Prykhodko, O.; Bjerrum, E. J.; Tyrchan, C.; Reymond, J.-L.; Chen, H.; Engkvist, O., Randomized SMILES strings improve the quality of molecular generative models. *Journal of cheminformatics* **2019,** *11* (1), 1-13.
12. Gao, W.; Coley, C. W., The Synthesizability of Molecules Proposed by Generative Models. *Journal of Chemical Information and Modeling* **2020,** *60* (12), 5714-5723.
13.vMahjour B, B. J., Zhang R, Frazier J, Cernak T, Molecular Sonification for Molecule to Music Information Transfer. *ChemRxiv Cambridge: Cambridge Open Engage; 2022* **2022**.
14. Yu, C.-H.; Qin, Z.; Martin-Martinez, F. J.; Buehler, M. J., A Self-Consistent Sonification Method to Translate Amino Acid Sequences into Musical Compositions and Application in Protein Design Using Artificial Intelligence. *ACS Nano* **2019,** *13* (7), 7471-7482.
15. Garrido, N.; Pitto-Barry, A.; Soldevila-Barreda, J. J.; Lupan, A.; Boyes, L. C.; Martin, W. H. C.; Barry, N. P. E., The Sound of Chemistry: Translating Infrared Wavenumbers into Musical Notes. *Journal of Chemical Education* **2020,** *97* (3), 703-709.
16. Roberts, A.; Engel, J.; Raffel, C.; Hawthorne, C.; Eck, D., A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music. In *Proceedings of the 35th International Conference on Machine Learning*, Jennifer, D.; Andreas, K., Eds. PMLR: Proceedings of Machine Learning Research, 2018; Vol. 80, pp 4364--4373.

A Broader View: On the Hidden Curriculum

Tyler R. Josephson, Assistant Professor in Chemical, Biochemical, and Environmental Engineering at the University of Maryland, Baltimore County

The unwritten lessons and unspoken expectations in academic settings, what is known as "hidden curriculum," often give learning advantages to those coming from the predominant culture [1]. Historically excluded students, including students of color, of low socioeconomic class, or first-generation, must learn to adopt a new set of social norms to excel in college. Because these lessons are not explicitly taught in the core curriculum, students will either pick them up through their varied experiences outside class, or they won't learn them at all.

A classic example is office hours [1]. Though office hours may be listed in the syllabus, many first-generation college students don't know their purpose and fail to take advantage of these. After all, "office hours" is not particularly self-descriptive (What happens in the office? Why should I come?). One of my grad student peers, a first-generation college student, didn't learn what office hours were until *after* college, when she was assigned to host them as a TA. She never took advantage because they had not been explained to her. One way to do better is to make mandatory the first office hours of the semester, so all students see for themselves what they involve. Another way is to rebrand these as "help hours," a more self-descriptive term.

Another aspect of hidden curriculum is how to manage professional references and letters of recommendation. These are critical to students' grad school and fellowship applications, yet I've heard from surprised seniors, "Wait, I need three? Who should I ask?" In contrast, I've seen a junior reach out early in the semester: "I love your class! I'm applying to X next year – can you write me a strong letter? See you in office hours!" To address this knowledge gap, faculty need to broach the subject first. I tell all students in my elective course on computing in chemical engineering that they can reach out to me if they need a reference for any coding-oriented jobs or graduate programs. My Google Form prompts students to give me information; these prompts give students a glimpse into what reference letters are all about.

I also give a lecture on "Networking: How to Make Professional Friends" (lessons I learned through my friend from industry), and I assign students to set up an informational interview and write a reflection on the experience. The diversity of responses is stark: one student learns about networking for the first time and finds contacts on her own, while another has a family member introduce him to the CEO of a startup. While this is beneficial to the more advantaged student, it can be transformational to the student for whom the concept was unknown.

One instructional model for addressing the hidden curriculum is sharing 5-minute lessons throughout the semester at the beginning of class, to regularly convey advice in an approachable and structured setting. This class design equips students *and* faculty to recognize the hidden curricula in STEM [2, 3]. These formal activities can form a starting point to establishing mentoring relationships with individual students, to help them navigate the particular challenges they face.

Hidden curricula continue through grad school, faculty application season, and the tenure process – in each stage, we should look for opportunities to better educate underrepresented scientists and engineers. However, we should also be open to challenging these norms [4]. Rather than equipping outsiders to "fit into" the ivory tower, why not reform our own culture, so they can increasingly come as they are?

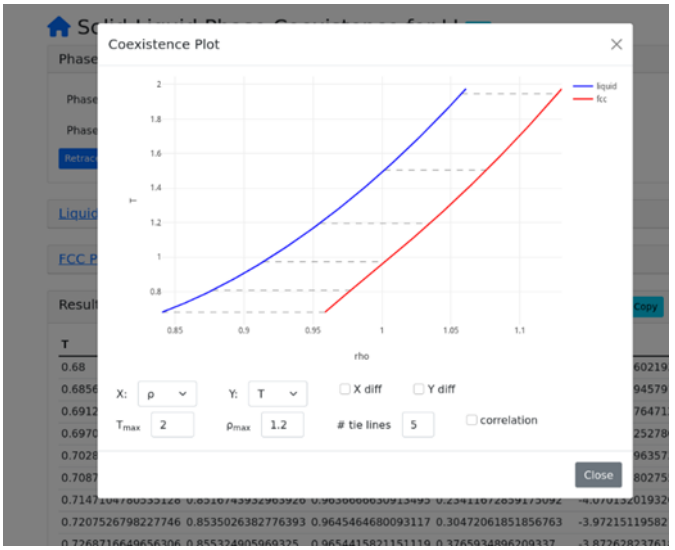
[1] Chatelain, Marcia, "We Must Help First-Generation Students Master Academe's 'Hidden Curriculum'", The Chronicle of Higher Education (2018).
[2] I. Villanueva, T. Carothers, M. Di Stefano, and M. T. H. Khan, "'There is never a break': The hidden curriculum of professionalization for engineering faculty", Education Sciences 8 (2018).
[3] M. Polmear, A. R. Bielefeldt, D. Knight, C. Swan, and N. E. Canney, "Hidden Curriculum Perspective on the Importance of Ethics and Societal Impacts in Engineering Education", in ASEE annual conference and exposition, conference proceedings (2019).
[4] A. M. Vaillancourt and O. H. Ersin. "Is It Time to Eliminate Recommendation Letters? (Hint: Yes)," The Chronicle of Higher Education (2021).

Spotlight on the Web: Lennard-Jones Equation of State Web App

Andrew Shultz, Department of Chemical and Biological Engineering School of Engineering and Applied Sciences, University at Buffalo

In 2018, Dave Kofke and I published a paper (DOI: 10.1063/1.5053714) to illustrate the power of our recently-developed harmonically mapped averaging – applying the technique to the Lennard-Jones system, computing highly precise and accurate thermodynamic properties of the solid phases (FCC and HCP) from molecular simulation and presenting equations of state (EOS) for the solids as well as (parts of) the liquid phase, each consistent with the simulation data within their uncertainties. The equations of state (augmented with existing virial coefficients to describe the vapor) were used to produce a complete phase diagram (for low temperature and/or high density); the resulting solid-liquid equilibrium data are two orders of magnitude more accurate and precise than previous literature calculations. This newsletter column describes a web app (www.etomica.org/apps/lj_eos/) I recently developed to make these equations of state and analysis of coexistence more broadly available.

The web pages for evaluating properties use the EOS from the paper or, alternatively, an EOS from the literature. Using the paper's EOS, the user can set various parameters including system size, potential truncation, and inclusion of anharmonic or vacancy contributions, allowing the user to see how these impact thermodynamic properties. The user can choose to evaluate a single state point (temperature and either density or pressure) or can vary them parametrically to see how properties vary as the state point changes. There are additional pages for coexistence, coexistence curves and triple points. Parametric EOS data and coexistence curves can be saved for comparison with other data within the app, plotted, or exported for external analysis. Generated coexistence data can also be compared against correlations from the paper.



The site is perhaps most useful for researchers who need to generate data for the Lennard-Jones model without running molecular simulations or programming an existing EOS. Beyond that, it provides an open-ended playground where any user can explore the thermodynamics of Lennard-Jones and its sensitivities to the equation of state.

The site reuses the Python scripts I originally used for the paper as a library via Brython, which translates Python to JavaScript for execution in a browser. Some of the expensive scripts had to be rewritten to use an event loop to maintain browser responsiveness while the third virial coefficient computation had to be replaced with a WebAssembly implementation due to the expense of the FFT code.

I welcome bug reports or feature requests that can make the app more useful for a particular application you might have in mind.

Institute for Computational Molecular Science Education Established to meet Training Needs of the CoMSEF Community

We are pleased to announce the establishment of the Institute for Computational Molecular Science Education (i-CoMSE) through support from the National Science Foundation (NSF). This institute is a result of the need for a training institute widely recognized by the CoMSEF community. In the 2019 CoMSEF Executive meeting, we began to explore the possibility of such a training institute, which led to pursuing funding opportunities within NSF to jumpstart the complex task of establishing this institute. The proposal had the support of nearly 30 CoMSEF community members to either serve on the advisory board and instructors or committed to providing instructional materials for the Institute. With the NSF award in October, i-CoMSE was established in Fall 2021.

The Institute's primary goal is to provide training and educational modules for guided training in advanced computational tools and bolster the fundamental understanding of the underlying theoretical concepts. Specifically, the Institute will offer (1) summer/winter schools for hands-on training on advanced computational techniques and enhancing peer networking for early-stage researchers, (2) web-based content to support training at a larger scale, and (3) curriculum and instructional materials for senior undergraduate (UG) and graduate courses to support course development across the community. The course materials will include modeling approaches spanning different length and time scales (quantum mechanics, atomistic modeling, mesoscale simulations) and advanced data analysis tools, including machine learning. Through the summer/winter schools and online instructional material, this community-led effort will serve a multidisciplinary community of chemical engineers, material scientists, chemists, and biophysicists.

Currently, two summer schools are planned: DFT for Catalysis at Mississippi State University (June 12-18, 2022, https://www.che.msstate.edu/dft/), and Molecular Dynamics/Monte Carlo at Oklahoma State University (July 10-15, 2022, https://www.che.msstate.edu/mdmc_osu/). Two additional winter schools are planned, and we will share the details with the community in late Summer 2022.

Efforts are currently underway way to set up the website for the Institute. We hope to have the website live by June 2022. If you want to be informed about future events and activities or have any comments/suggestions, please email info@i-comse.org.

i-CoMSE core committee members:
Eric Jankowski, Boise State University
Neeraj Rai, Mississippi State University
Sapna Sarupria, University of Minnesota Twin Cities
Jindal Shah, Oklahoma State University
Michael Shirts, University of Colorado, Boulder

Upcoming Conferences of Interest to CoMSEF Members

MolSim2022: Present, Past and Future
Erice, Italy
June 25-29, 2022
https://bricabrac.fisica.unimo.it/ErcMlk80/

Congress of the World Association of Theoretical and Computational Chemists
Vancouver, Canada
July 3-8, 2022
https://www.cheminst.ca/conference/watoc-2020/

European Symposium on Applied Thermodynamics
Graz, Austria
July 17-20, 2022
https://www.tugraz.at/events/esat2022/home/

FOMMS
Delavan, WI
July 17-21, 2022
http://fomms.org

CCP5 Summer School
Durham, England
July 17-28, 2022
https://summer2022.ccp5.ac.uk/

SC22
Dallas, TX
November 13-18, 2022
https://sc22.supercomputing.org/

PPEPPD
Tarragona, Spain
May 21-25, 2023
https://ppeppd.org/

International Congress on Quantum Chemistry
Bratislava, Slovakia
June 26-July 1, 2023
https://icqc2023.org/

STATPHYS28
Tokyo, Japan
August 7-11, 2023
https://www.statphys28.org/

22nd Symposium on Thermophysical Properties
Boulder, CO
June 23-28, 2024
https://thermosymposium.org/

Where are They Now?

Now that CoMSEF has been giving the graduate student awards for more than 10 years, we've started including a "where are they now?" section in the newsletter, catching up with the winners from ~ 10 years ago.

Christopher Wilmer
2012 Grad Student Award winner (Northwestern University, Advisor: Randy Snurr)
Poster Title: New Methods for High Throughput Screening of Porous Materials

Chris Wilmer is now an Associate Professor in the Department of Chemical & Petroleum Engineering at the University of Pittsburgh, where he directs the "hypothetical materials" lab. His research focuses on advanced applications of metal-organic frameworks, for example, by considering how they can be used to make electronic noses. Additionally, his group is making significant contributions to the understanding of thermal transport in MOFs, and also in efficient methods for large-scale computational screening of porous materials for gas adsorption applications. After co-founding NuMat Technologies as a grad student at Northwestern, Chris brought his entrepreneurial enthusiasm to Pitt, where in addition to co-founding another company, Aeronics, he has taken on the service role of helping spur student entrepreneurship. Outside of the lab, Chris enjoys long distance running, editing his wife's (Emily) papers on Renaissance dance history, proselytizing about cryptocurrencies, and making animations in Blender with his 7 year old son William (see "Ninja on The Train 3").

Pictured: Chris [left], William (son) [middle], Emily (wife) [right]

Why CoMSEF?

Occasionally it is worthwhile to remind everyone what CoMSEF does for our community and why your membership support is important. CoMSEF was founded in 2000, and since that time it has worked to advance molecular science and engineering in diverse ways:

* We provide a forum for communication and networking within the community. The document you're reading now is a prime example, but there is more. The annual membership meeting provides a venue for communication and interaction among members. The CoMSEF web site http://comsef.org is another useful resource for this purpose. It often hosts notices about upcoming workshops, available post-doc positions, etc.

* We provide a vehicle for communication and advocacy for molecular science and engineering in relation to other research communities. For example, our four Liaison Directors identify opportunities for co-sponsorship of sessions at the AIChE Annual Meeting, facilitate programming with other organizations, and communicate and advocate CoMSEF activities with other organizations.

* We help to recognize and promote outstanding researchers and promising graduate students by funding and administering several awards. Our awards help the contributions of some of our best researchers to be recognized by a broad audience, extending into the larger chemical engineering community. Your dues make these awards possible.

* We provide technical programming support, ensuring we have sessions of interest to you at the AIChE meeting. These include the many sessions we sponsor or co-sponsor, as well as the CoMSEF plenary, CoMSEF poster, and Industrial Fluid Properties Simulation Challenge sessions. We also work externally to AIChE, providing technical sponsorship to conferences in our discipline (e.g., FOMMS), where we help to ensure that these events have molecular science and engineering content of the highest quality.

Your support of CoMSEF through your membership is very important in enabling us to fulfill our mission. The financial element is valuable of course, but we also gain strength in demonstrating the size of the community we represent. So please make sure to check the box to include renewal of your CoMSEF membership whenever you pay your annual dues to AIChE. When the opportunity arises, encourage your non-member colleagues in the molecular science and engineering community to join too!