The Search for Novel Mesoscale Materials

Rose K. Cersonsky

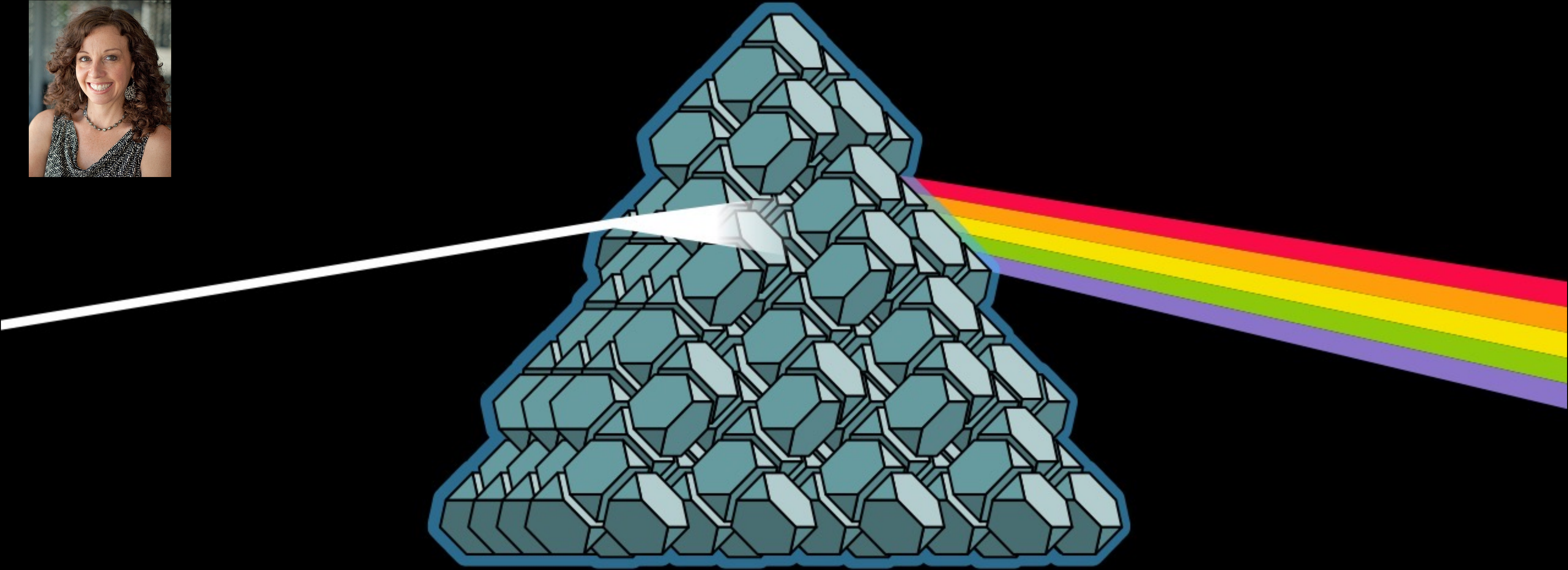
Laboratory of Computational Science and Modeling (COSMO)

École Polytechnique Fédérale de Lausanne (EPFL)

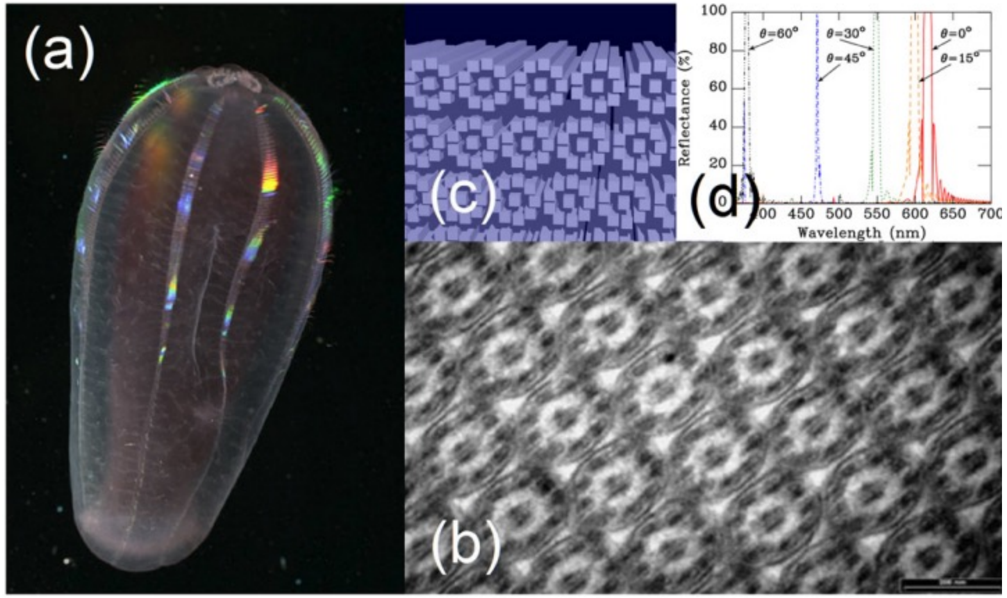
Lausanne, Switzerland

MICHIGAN ENGINEERING

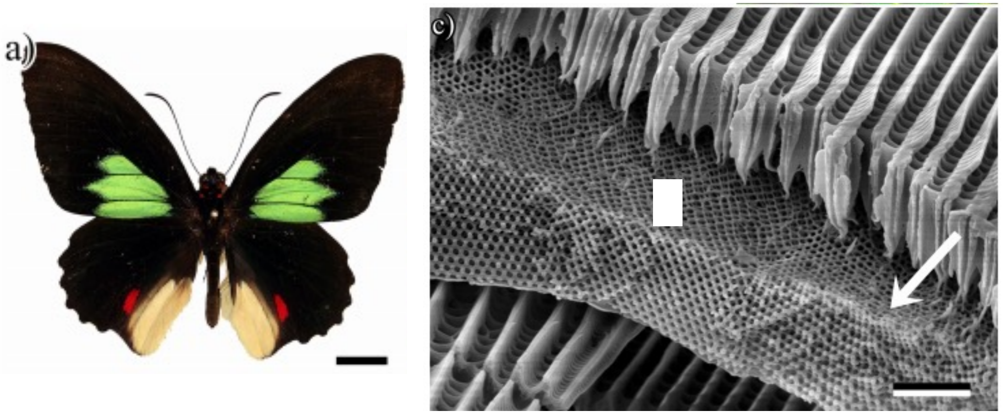
MACROMOLECULAR SCIENCE & ENGINEERING PROGRAM



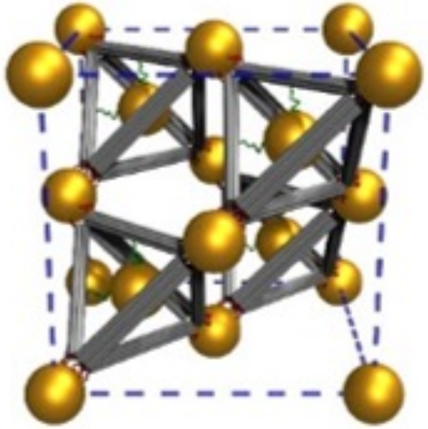
Designing Nanoparticles for the Self-Assembly of Novel (Photonic) Materials

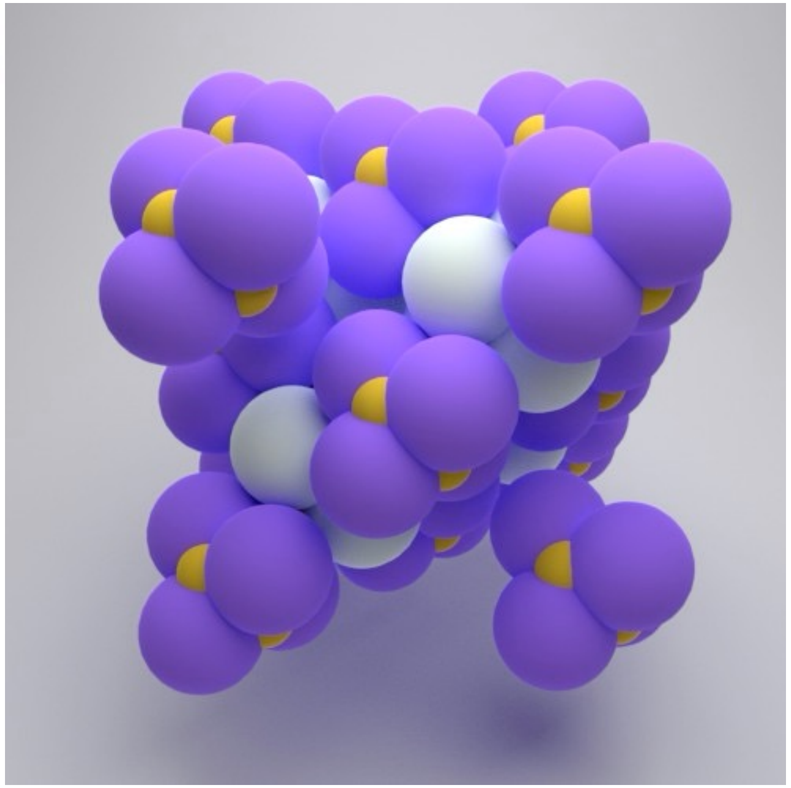


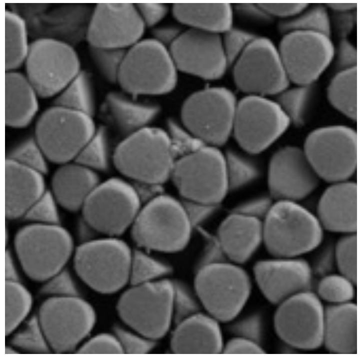
Optical properties of the iridescent organ of the comb-jellyfish Beroë cucumis (Ctenophora)
Victoria Welch, *et al.*
Phys. Rev. E 73, 041916 2006



Optical properties of gyroid structured materials: from photonic crystals to metamaterials
James A. Dolan , *et al.*
Advanced Optical Materials 3 (1), 12-32





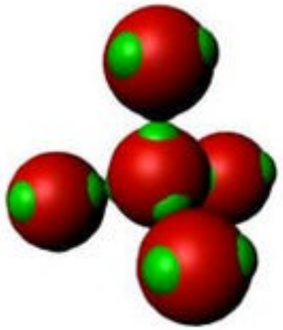


Colloidal crystals with diamond symmetry at optical lengthscales
Yifan Wang, et al.
Nature Comm. 8, 14173 (2017)

Entropy driven assembly of truncated colloidal tetrahedra into diamond structure
Zhe Gong, et al.

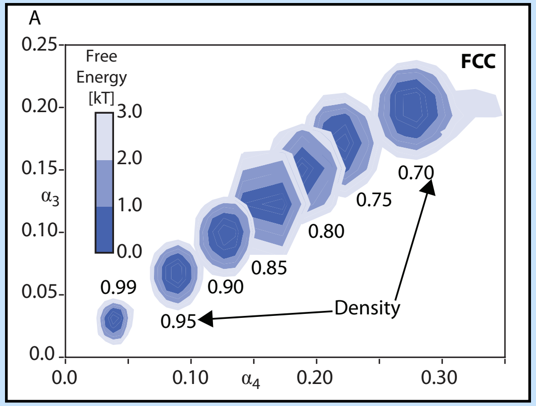
Colloidal Diamond
He, M., et al.
Nature 585, 524-529 (2020).

Diamond family of nanoparticle superlattices
W. Liu, et. al,
Science 351, 582-586 (2016).



Relevance of packing to colloidal self-assembly.

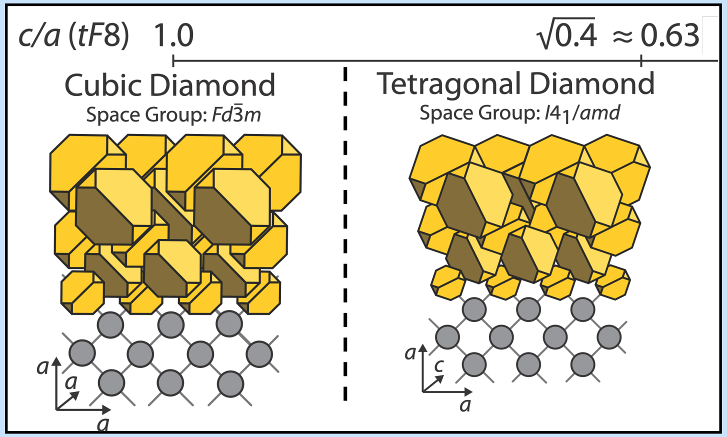
Cersonsky, R. K., van Anders, G., Dodd, P. M., & Glotzer, S. C. (2018). *Proceedings of the National Academy of Sciences*, *115*(7), 1439-1444.



- Pauling's packing rules are **not** a causal mechanism for nanoparticle self-assembly
- Using the Digital Alchemy framework, I showed that **adding small imperfections** to nanoparticle shapes would better stabilize nanocrystals

Pressure-Tunable Photonic Band Gaps in an Entropic Colloidal Crystal

Cersonsky, R. K., Dshemuchadse, J., Antonaglia, J., van Anders, G., & Glotzer, S. C. (2018). *Physical Review Materials*, *2*(12), 125201.



- Nanoparticles that stabilized diamond in self-assembly can **transition to lower-symmetry derivatives** at high pressure
- Small distortions in diamond **did not destroy the photonic band gap**

At time of presentation, this manuscript was not yet published, please see rosecersonsky.com for recent publications.

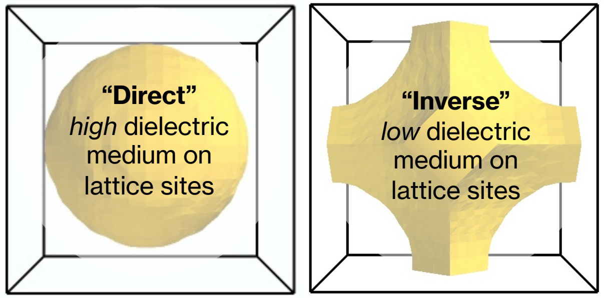
…Small distortions in diamond **did not destroy the photonic band gap…**
…minimal effect on the photonic band structure…

what is the span of crystallographic structures capable of supporting a photonic band gap?

COD

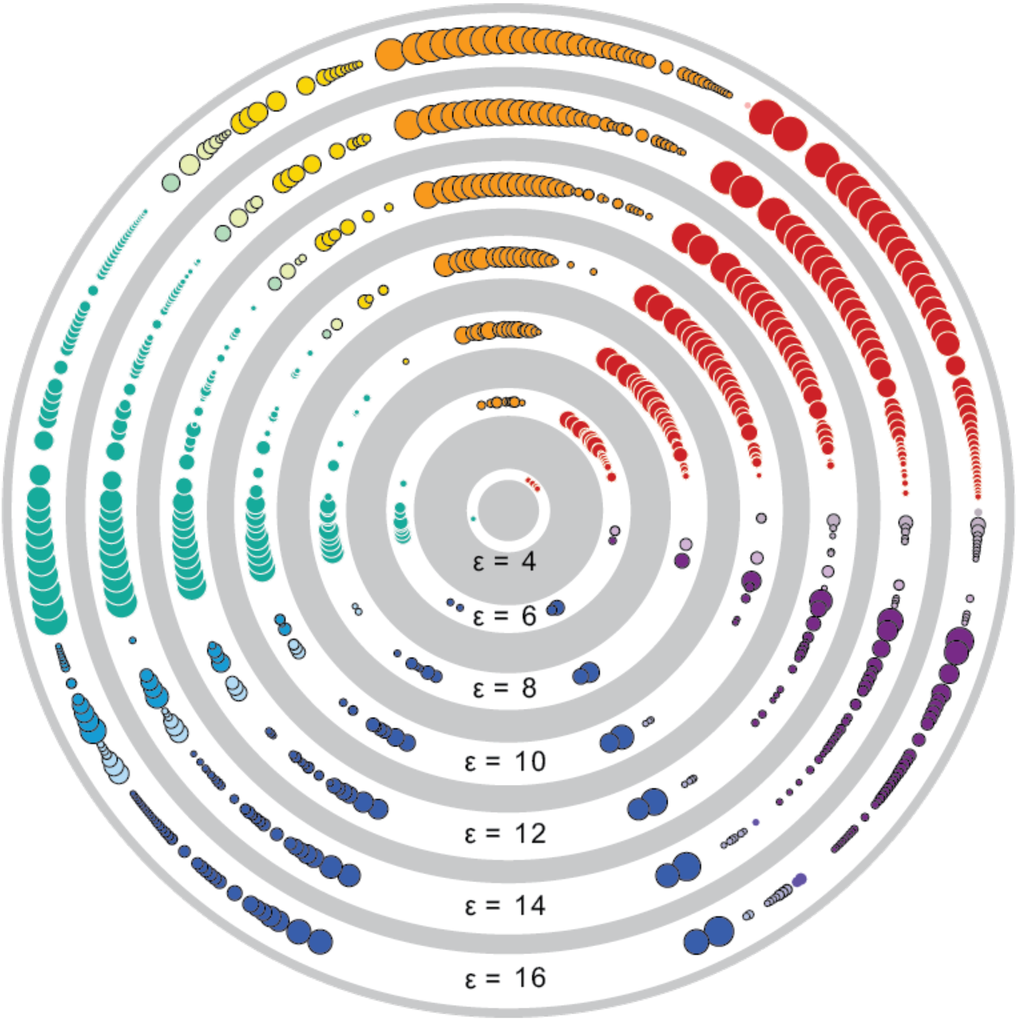
ICSD
FIZ Karlsruhe

Scraped 1300 unique crystal structures from crystal repositories



Ignoring atomic species, turned each structure into 2 templates with which to sample multiple parameters

Ran >150,000 band structure to determine which “templates” supported PBGs



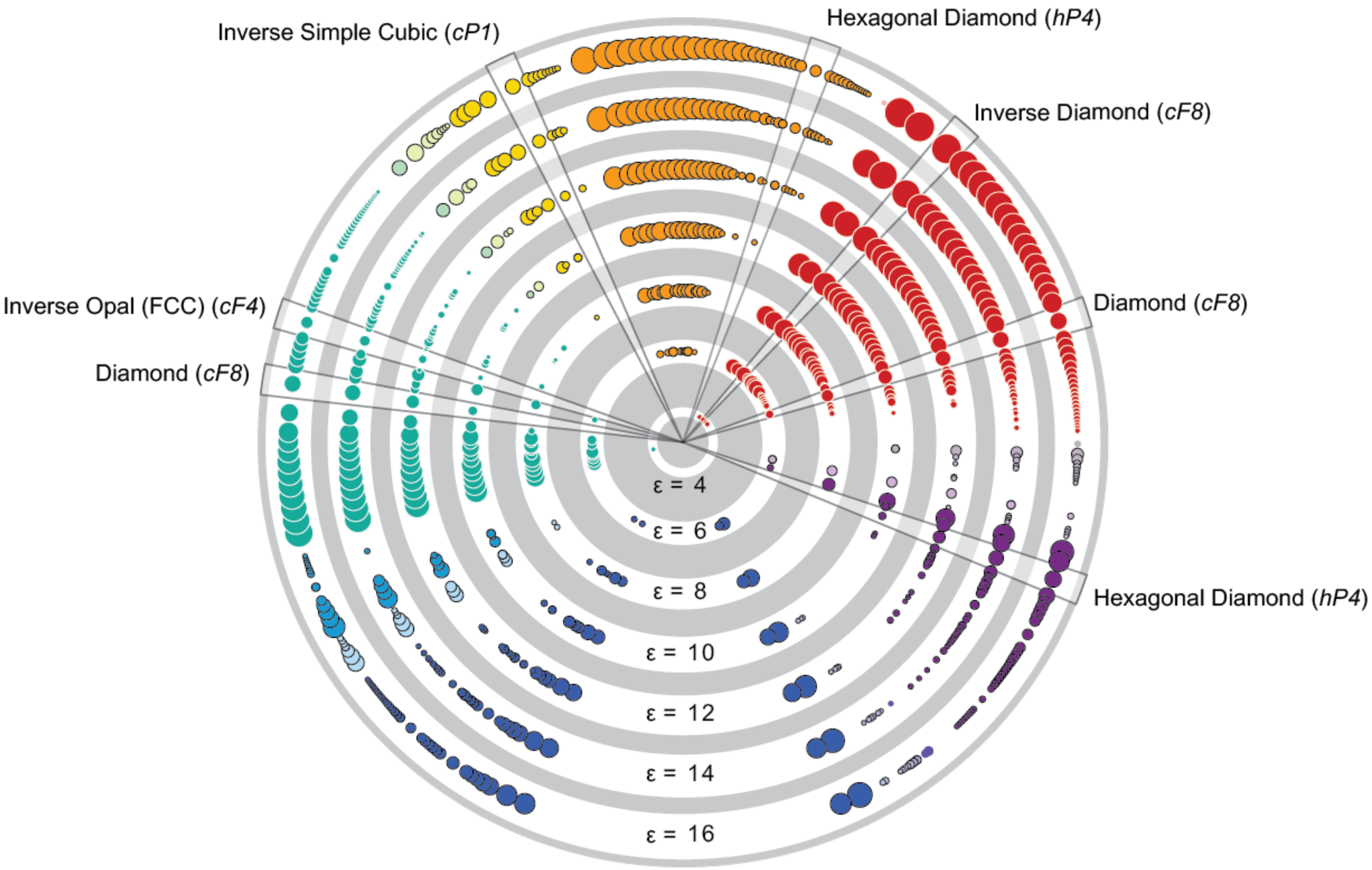
351 Photonic “Templates”

474 Unique Gaps

Database of Photonic Crystals:
https://glotzerlab.engin.umich.edu/photonics/index.html

Appendix of Band Structures:
https://deepblue.lib.umich.edu/handle/2027.42/153520

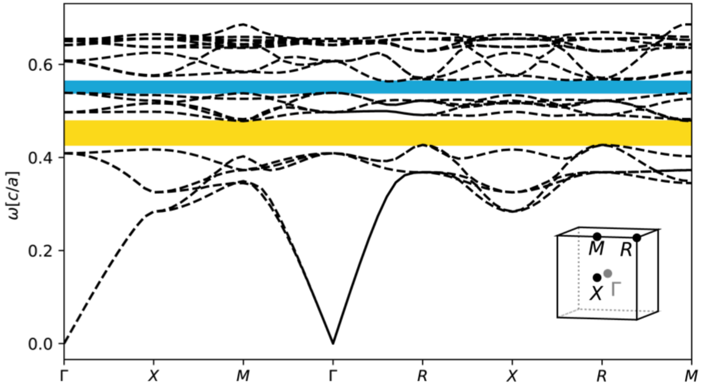
Each circle represents the maximum gap (circle size) found for a given template (radius), dielectric contrast (ring), and band location (color).

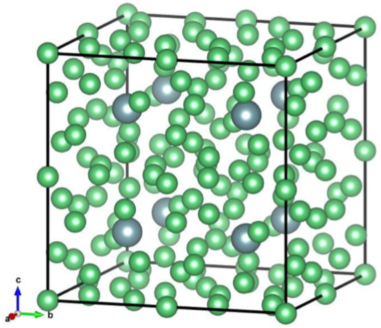


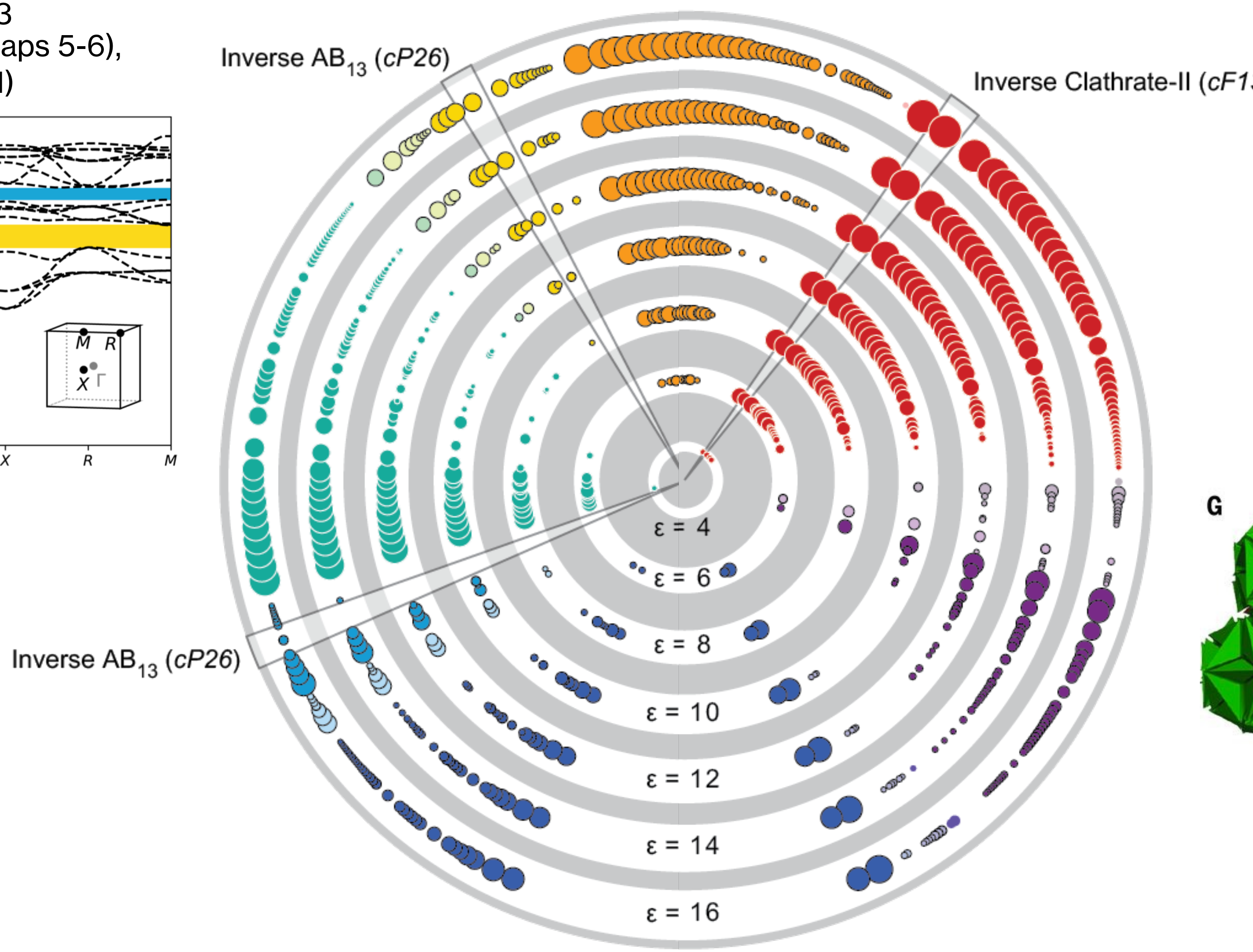
Each circle represents the maximum gap (circle size) found for a given template (radius), dielectric contrast (ring), and band location (color).

Inverse AB_{13}

Maximum Gaps: 13.3% (Gaps 5-6), 4.78% (Gaps 10-11)

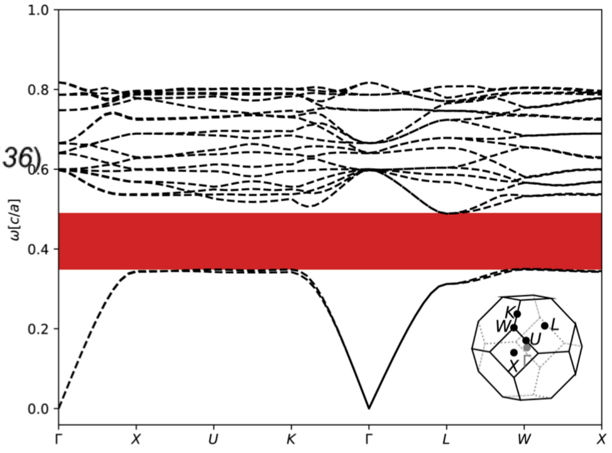


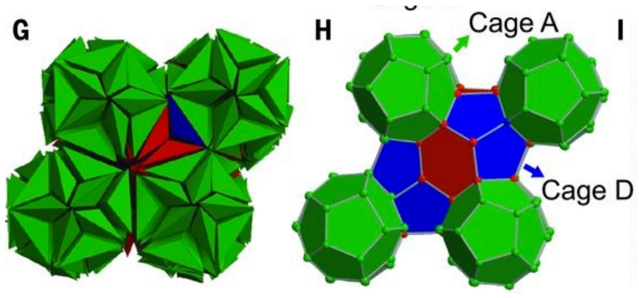




Inverse Clathrate-II

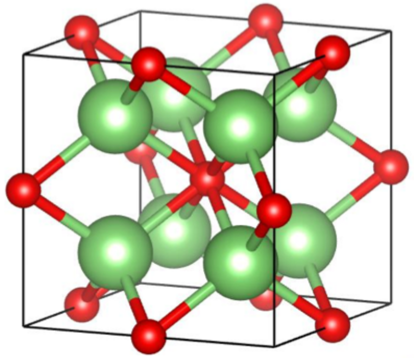
Maximum Gap: 33.9%

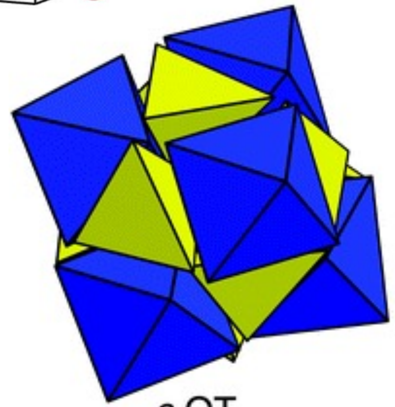




Clathrate colloidal crystals.
Lin, H., Lee, S., Sun, L., Spellings, M., Engel, M., Glotzer, S. C., & Mirkin, C. A. Science, 355(6328), 931-935.

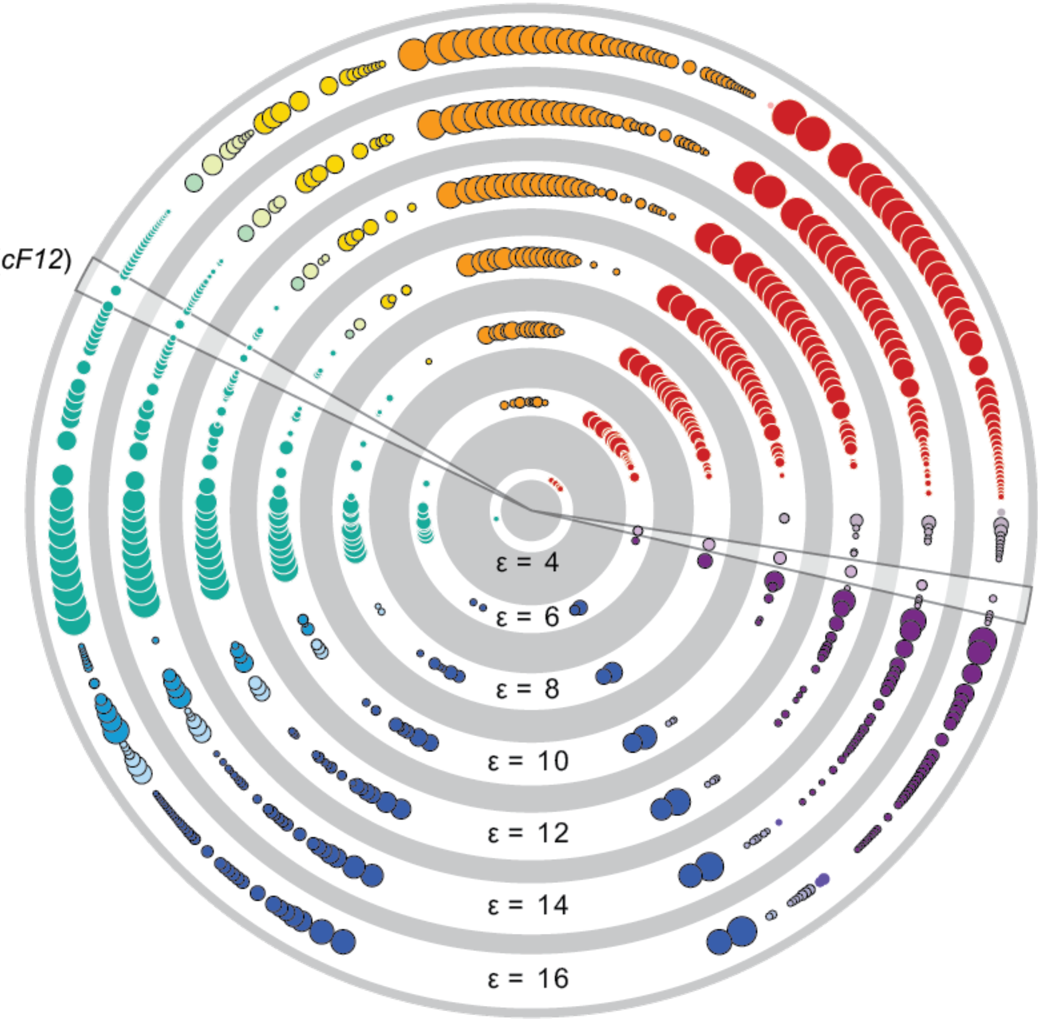
Each circle represents the maximum gap (circle size) found for a given template (radius), dielectric contrast (ring), and band location (color).





c-OT$_2$

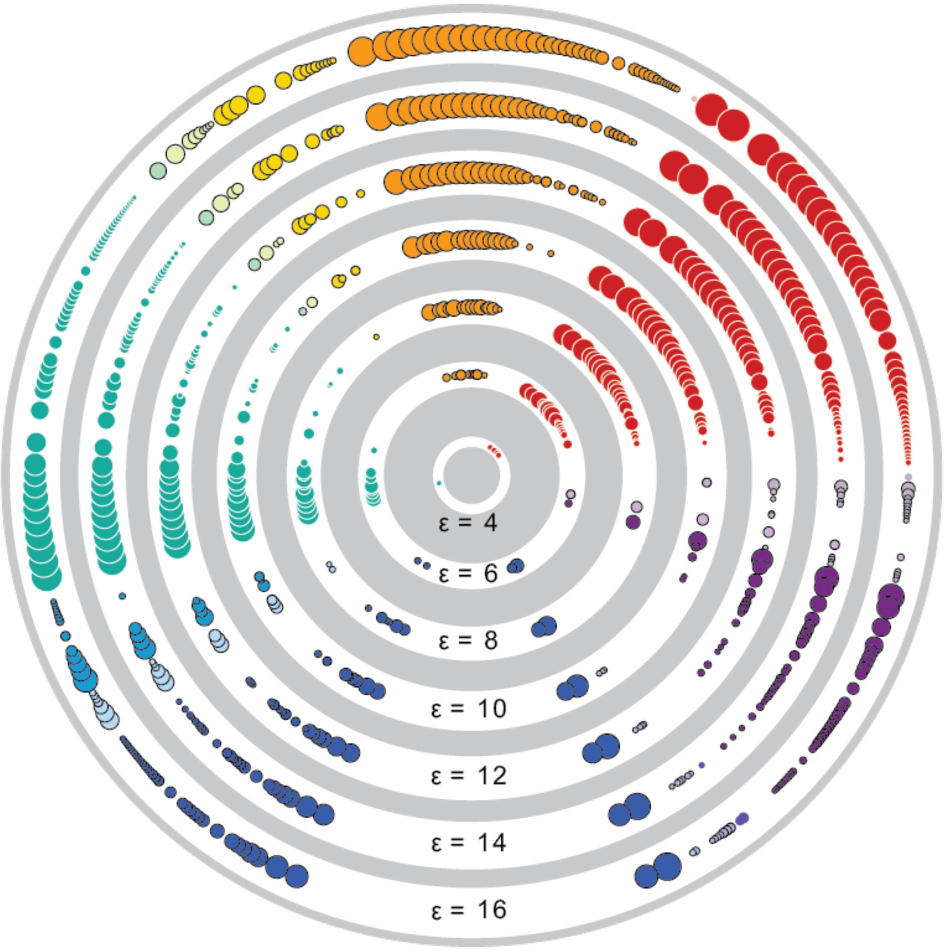
Self-assembly of a space-tessellating structure in the binary system of hard tetrahedra and octahedra.
Cadotte, Andrew T., et al.
Soft matter 12.34 (2016): 7073-7078.

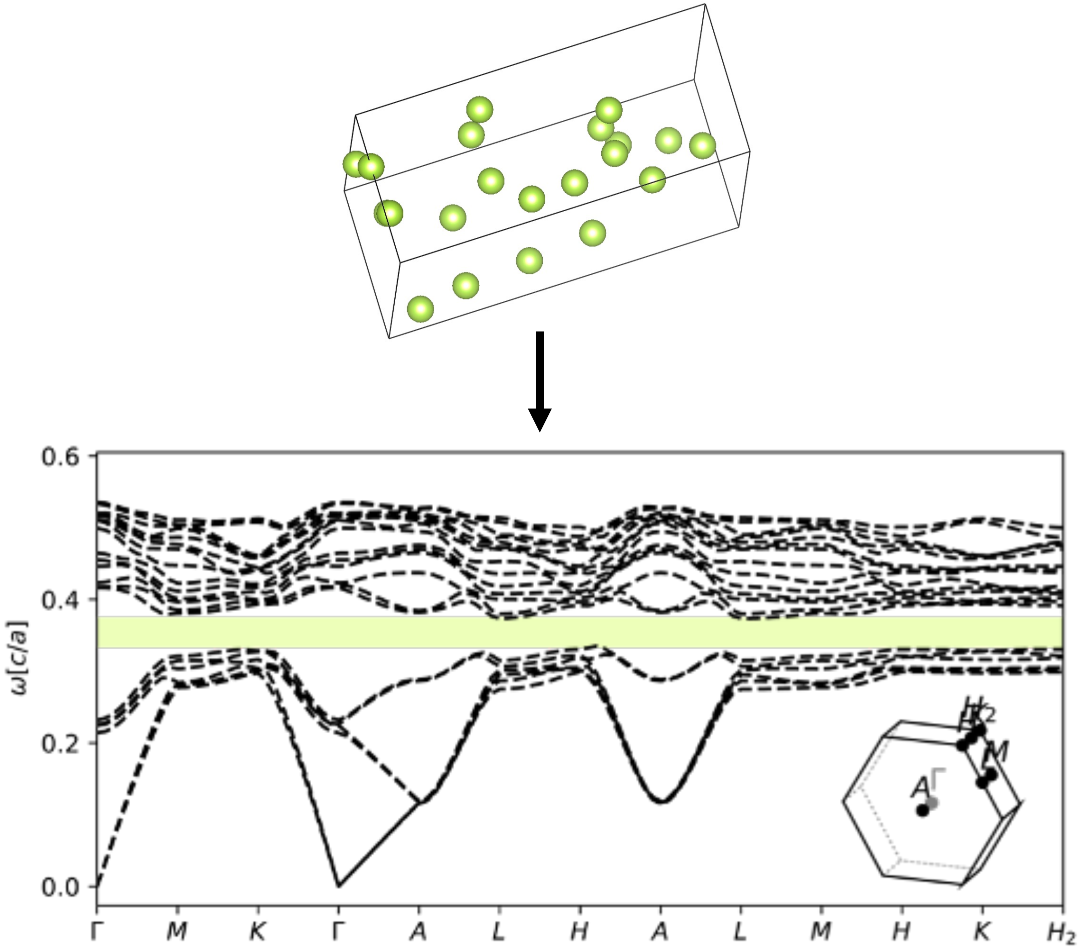


The lithium-oxide structure (a.k.a. Fluorite, c-OT$_2$, and F-RD) exhibits photonic anomalies, including a band gap that is largest at lower dielectric contrast.

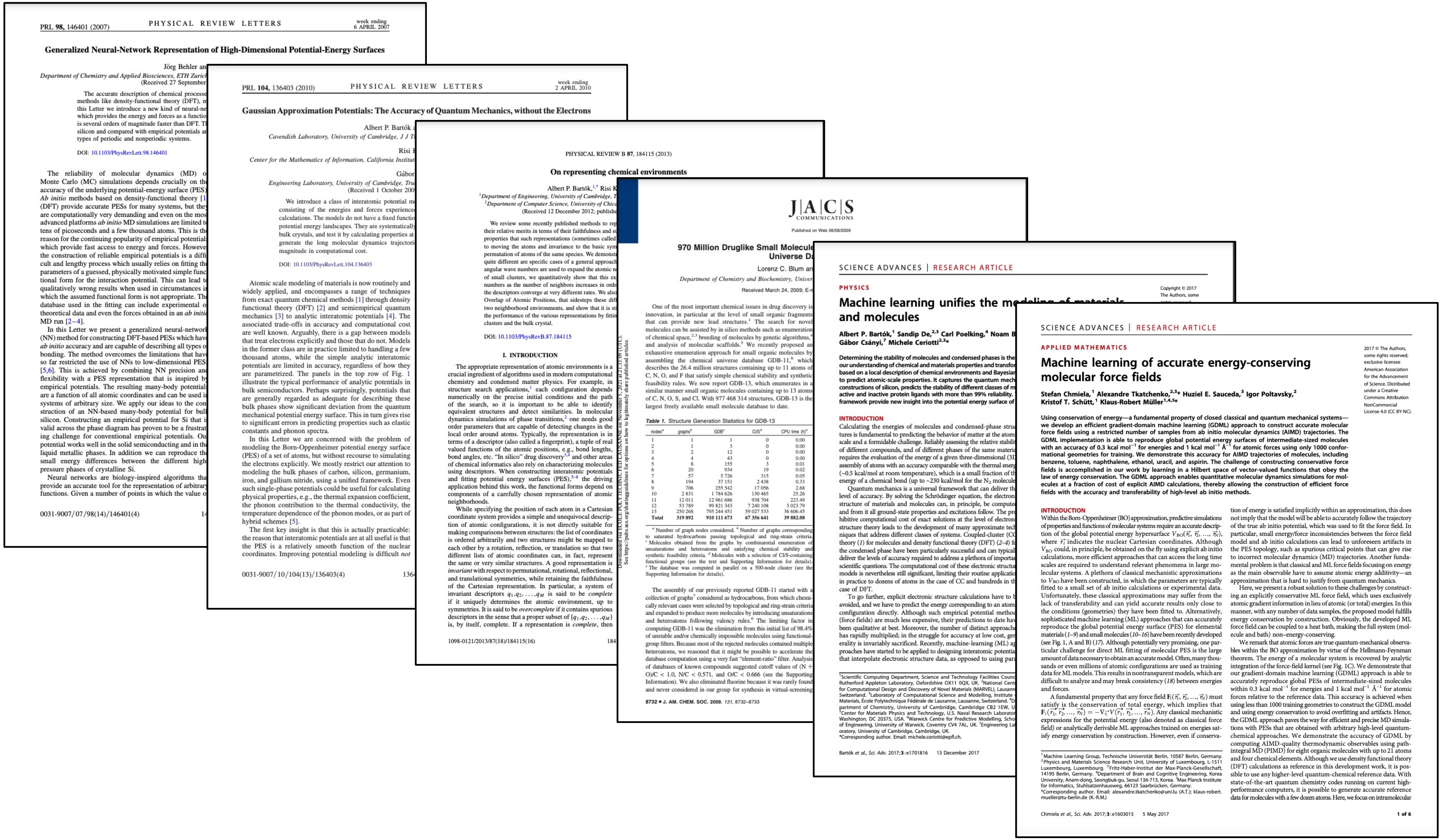
Each circle represents the maximum gap (circle size) found for a given template (radius), dielectric contrast (ring), and band location (color).

How else can we use this large dataset?





Database of Photonic Crystals:
https://glotzerlab.engin.umich.edu/photonics/index.html



PRL 98, 146401 (2007) PHYSICAL REVIEW LETTERS week ending 6 APRIL 2007

Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces

Jörg Behler an

DOI: 10.1103/PhysRevLett.98.146401

PRL 104, 136403 (2010) PHYSICAL REVIEW LETTERS week ending 2 APRIL 2010

Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons

Albert P. Bartók

DOI: 10.1103/PhysRevLett.104.136403

PHYSICAL REVIEW B 87, 184115 (2013)

On representing chemical environments

Albert P. Bartók,[1,*] Risi K

DOI: 10.1103/PhysRevB.87.184115

J|A|C|S COMMUNICATIONS

Published on Web 06/08/2009

970 Million Druglike Small Molecule Universe D

Lorenz C. Blum an

8732 ■ J. AM. CHEM. SOC. 2009, 131, 8732–8733

SCIENCE ADVANCES | RESEARCH ARTICLE

PHYSICS

Machine learning unifies the modeling of materials and molecules

Albert P. Bartók,[1] Sandip De,[2,3] Carl Poelking,[4] Noam B
Gábor Csányi,[7] Michele Ceriotti[2,3*]

Bartók et al., Sci. Adv. 2017;3:e1701816 13 December 2017

SCIENCE ADVANCES | RESEARCH ARTICLE

APPLIED MATHEMATICS

Machine learning of accurate energy-conserving molecular force fields

Stefan Chmiela,[1] Alexandre Tkatchenko,[2,3*] Huziel E. Sauceda,[3] Igor Poltavsky,[2]
Kristof T. Schütt,[1] Klaus-Robert Müller[1,4,5*]

2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Using conservation of energy—a fundamental property of closed classical and quantum mechanical systems—we develop an efficient gradient-domain machine learning (GDML) approach to construct accurate molecular force fields using a restricted number of samples from ab initio molecular dynamics (AIMD) trajectories. The GDML implementation is able to reproduce global potential energy surfaces of intermediate-sized molecules with an accuracy of 0.3 kcal mol^{-1} for energies and 1 kcal mol^{-1} Å^{-1} for atomic forces using only 1000 conformational geometries for training. We demonstrate this accuracy for AIMD trajectories of molecules, including benzene, toluene, naphthalene, ethanol, uracil, and aspirin. The challenge of constructing conservative force fields is accomplished in our work by learning in a Hilbert space of vector-valued functions that obey the law of energy conservation. The GDML approach enables quantitative molecular dynamics simulations for molecules at a fraction of cost of explicit AIMD calculations, thereby allowing the construction of efficient force fields with the accuracy and transferability of high-level ab initio methods.

Chmiela et al., Sci. Adv. 2017;3:e1603015 5 May 2017 1 of 6

EPFL

COSMO
COMPUTATIONAL SCIENCE AND MODELLING

IOP Publishing *Mach. Learn.: Sci. Technol.* 2 (2021) 035038 https://doi.org/10.1088/2632-2153/abfe7c

MACHINE LEARNING
Science and Technology

PAPER

Improving sample and feature selection with principal covariates regression

Rose K Cersonsky[1,*], Benjamin A Helfrecht[1], Edgar A Engel[2], Sergei Kliavinek[1] and Michele Ceriotti[1,*]

[1] Laboratory of Computational Science and Modeling, IMX, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
[2] TCM Group, Cavendish Laboratory, University of Cambridge, J.J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom
[*] Authors to whom any correspondence should be addressed.

E-mail: rose.cersonsky@epfl.ch and michele.ceriotti@epfl.ch

OPEN ACCESS

RECEIVED
18 December 2020

REVISED
23 April 2021

ACCEPTED FOR PUBLICATION
6 May 2021

PUBLISHED
14 July 2021

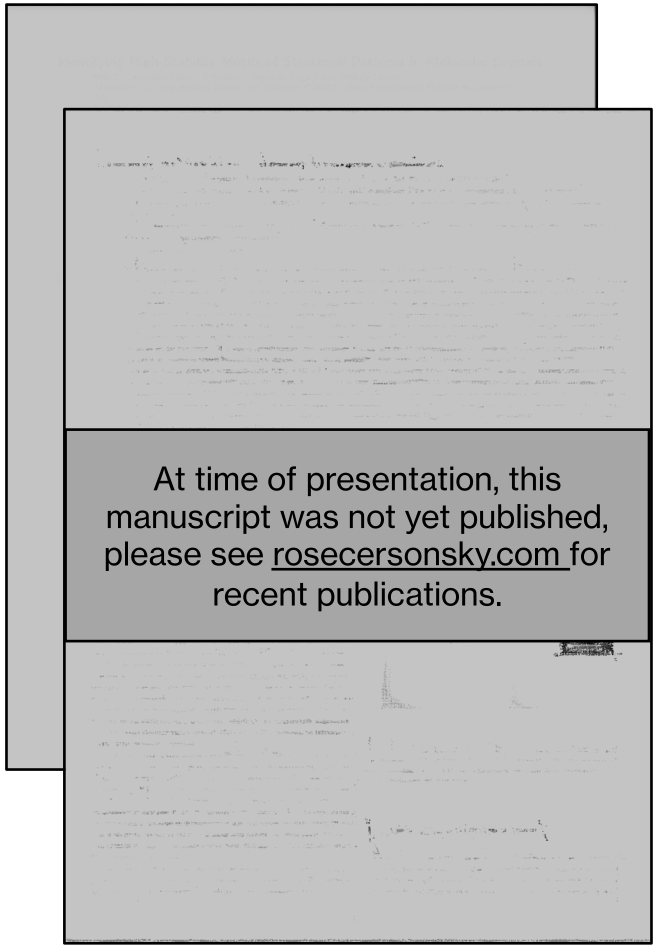
Keywords: machine learning, feature selection, sample selection, farthest point sampling, materials science, physical chemistry, semi-supervised learning

Supplementary material for this article is available online

Abstract

Selecting the most relevant features and samples out of a large set of candidates is a task that occurs very often in the context of automated data analysis, where it improves the computational performance and often the transferability of a model. Here we focus on two popular subselection schemes applied to this end: CUR decomposition, derived from a low-rank approximation of the feature matrix, and farthest point sampling (FPS), which relies on the iterative identification of the most diverse samples and discriminating features. We modify these unsupervised approaches, incorporating a supervised component following the same spirit as the principal covariates (PCov) regression method. We show how this results in selections that perform better in supervised tasks, demonstrating with models of increasing complexity, from ridge regression to kernel ridge regression and finally feed-forward neural networks. We also present adjustments to minimise the impact of any subselection when performing unsupervised tasks. We demonstrate the significant improvements associated with PCov-CUR and PCov-FPS selections for applications to chemistry and materials science, typically reducing by a factor of two the number of features and samples required to achieve a given level of regression accuracy.

203: Innovations in Methods of Data Science
Monday, November 8, 2021, 4:30 PM - 4:45 PM
Marriott Copley Place - Salon H/I



At time of presentation, this manuscript was not yet published, please see rosecersonsky.com for recent publications.

At time of presentation, this work was not yet published, please see rosecersonsky.com for recent publications.

The Search for Novel Mesoscale Materials

Rose K. Cersonsky

Laboratory of Computational Science and Modeling (COSMO)
École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland

At time of presentation, this work was not yet published, please see rosecersonsky.com for recent publications.

Come see my last talk!
203e - Improving Data Sub-Selection for Supervised Tasks with Principal Covariates Regression
Monday, November 8, 2021
4:30 PM - 4:45 PM EDT
Marriott Copley Place - Salon H/I

rosecersonsky.com