

泛化逆向设计的可逆无机晶体机器学习表征方法

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/progress/17107.html>

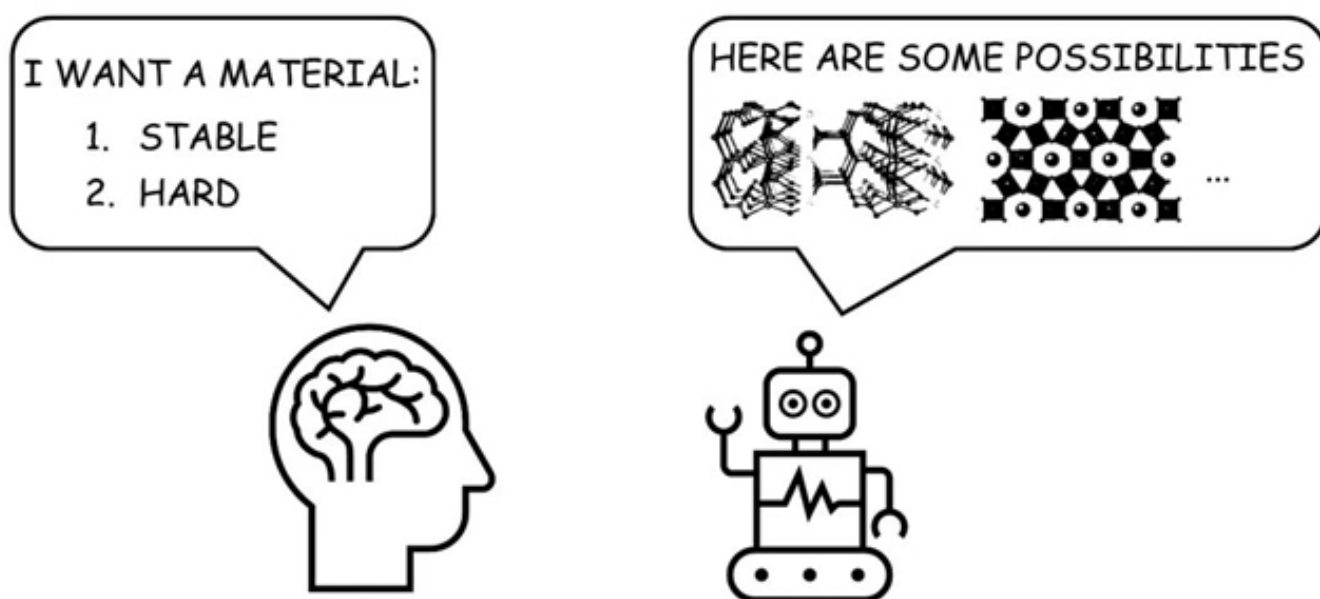
本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

泛化逆向设计的可逆无机晶体机器学习表征方法。

从人为设定的性质出发去预测材料的逆向设计（inverse design）是材料科学的一个终极目标，对材料学的发展颇为重要。而得益于近年来材料性质的数据库和机器学习的发展，一些基于生成模型的前作被陆续研究出来。不过这些前作仍然具有它们的局限性——生成的晶体材料局限于特定的组成元素，或者特定的晶体结构，即目前还没有一个泛化（可以生成多种组成元素多种晶体结构的材料）的逆向设计模型。

2021年12月20日，新加坡-麻省理工学院科研中心Tonio Buonassisi教授团队在Matter上发表了题为An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties的论文。

该研究提出了一个可逆的无机晶体机器学习的表征方法（傅里叶变换的晶体性质表征，Fourier-Transformed Crystal Properties representation），与变分自动编码器（一种生成模型）的模型结合，可以达到泛化的逆向设计的目标。



这个无机晶体的表征方法 (FTCP representation) 结合了两部分的特征：

1 正空间特征 (real-space features)，类似于晶体学信息文件里面的信息。

2 倒易空间特征 (reciprocal-space features)，通过用元素性质替换X光衍射计算中结构因子里面的散射因子来计算得到。

其中，正空间特征保证了晶体表征的可逆性（从晶体到表征，以及从表征回到晶体），倒易空间特征嵌入了一些晶体的对称性，帮助变分自动编码器输出端的重建。

变分自动编码器的模型除了常规的编码分支 (encoder) 和解码分支 (decoder)，增加了一个额外的从潜空间回归晶体性质的监督学习，即一个目标学习分支 (target-learning branch)。增加了目标学习分支的变分自动编码器在训练时可以在常规的潜空间基础上将潜变量的分布与晶体性质相关联（性质分布潜空间），从而实现由晶体性质出发的逆向设计。这个方法可以大量地预测可能具有指定性质的新材料，而具体的数字取决于如何对性质分布潜空间进行抽样。

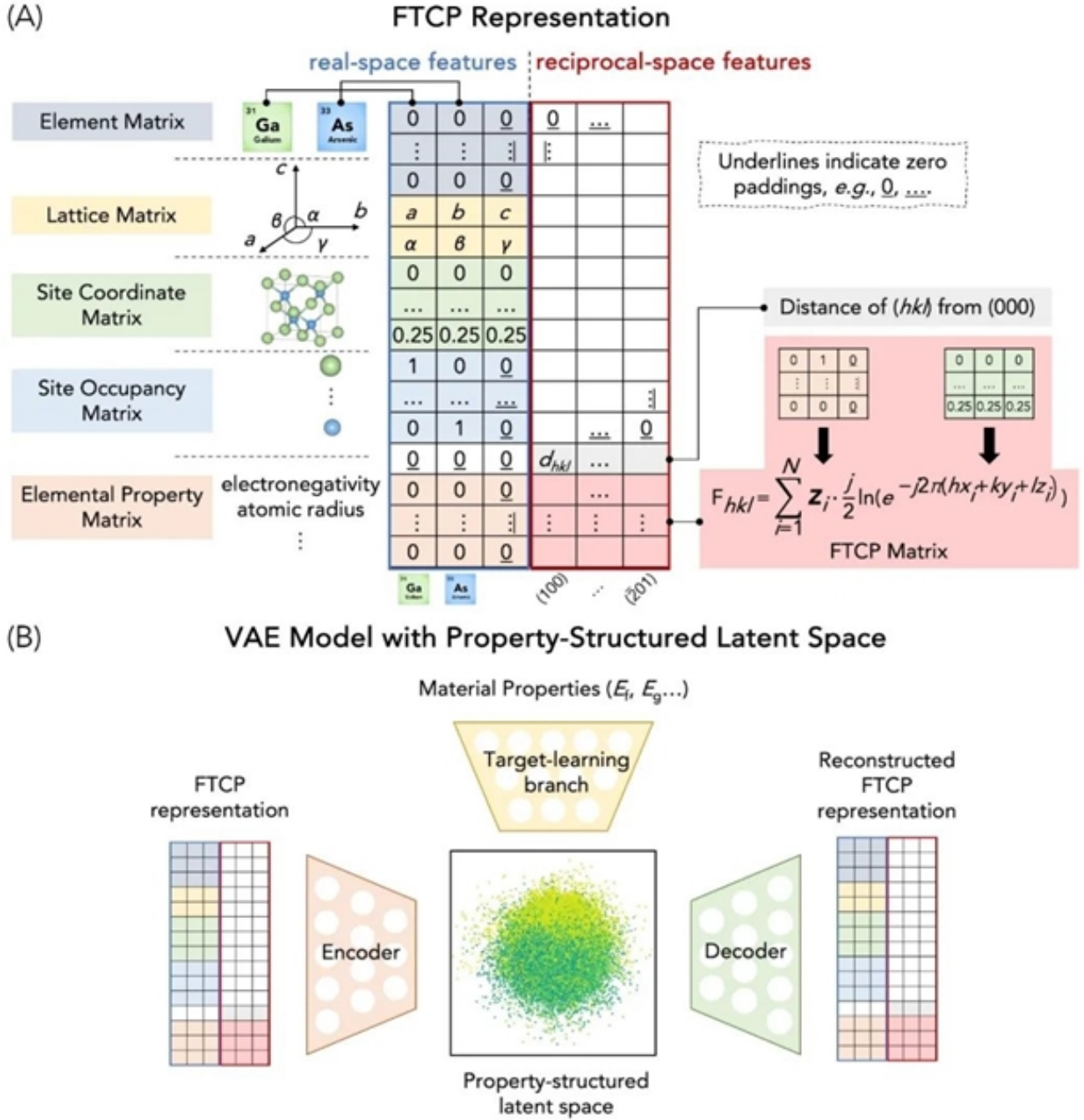


图1：(A) 傅里叶变换的晶体性质表征, FTCP representation, (B) 变分自动编码器以及性质分布潜空间, property-structured latent space。

许多FTCP设计/预测的材料通过第一性原理计算出的性质符合逆向设计里用户定义的性质—其中包括基态性质（ground-state properties），如形成能（formation energy），以及激发态性质（excited-state properties），如热电功率因子（thermoelectric power factor）。

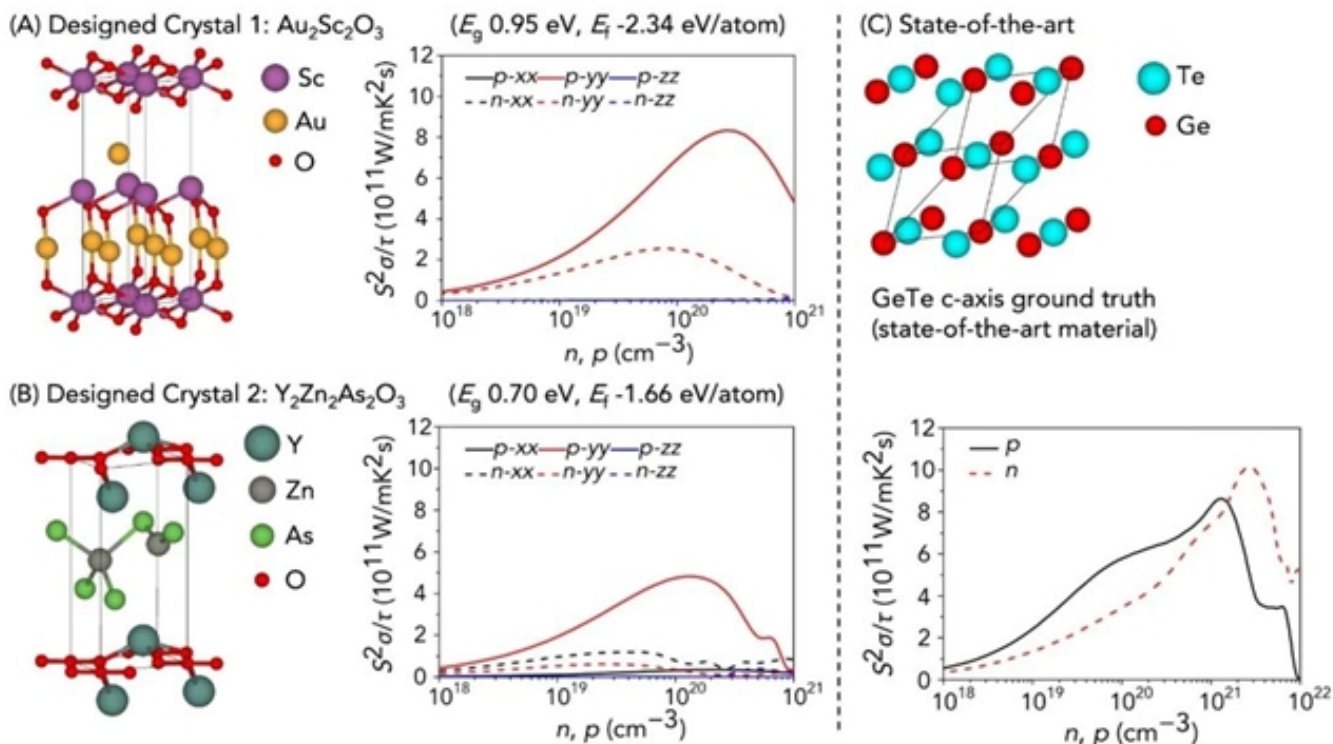


图2：(A和B)

FTCP预测的新热电材料及其第一性原理计算出的热电功率因子、带隙和形成能，(C) 最先进的热电材料GeTe及其热电功率因子。

用FTCP来预测新材料的时候，因为训练好的模型仍然有非零的各种损失函数，得到的晶胞参数，原子坐标等等会有偏差。因此，在本研究中我们在FTCP预测晶体之后用密度泛函理论进行了几何优化（structural relaxation with DFT）。

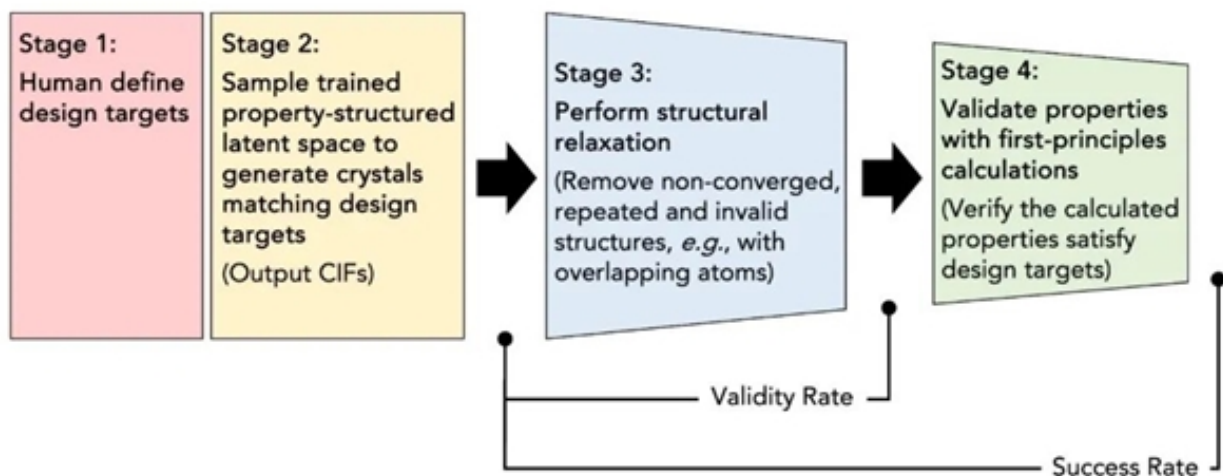


图3：实验流程

如果要把FTCP加入到材料的实验合成流程中，后续工作还需要实现几何优化和晶体稳定性预测的快速进行以达到筛选目标晶体的目的。目前使用密度泛函理论的几何优化的过流能力还不足以跟上基于生成模型的FTCP预测的速度。在本研究中，我们提出了一个很初级的筛选办法，并且引用了这个领域里面的一些前作。

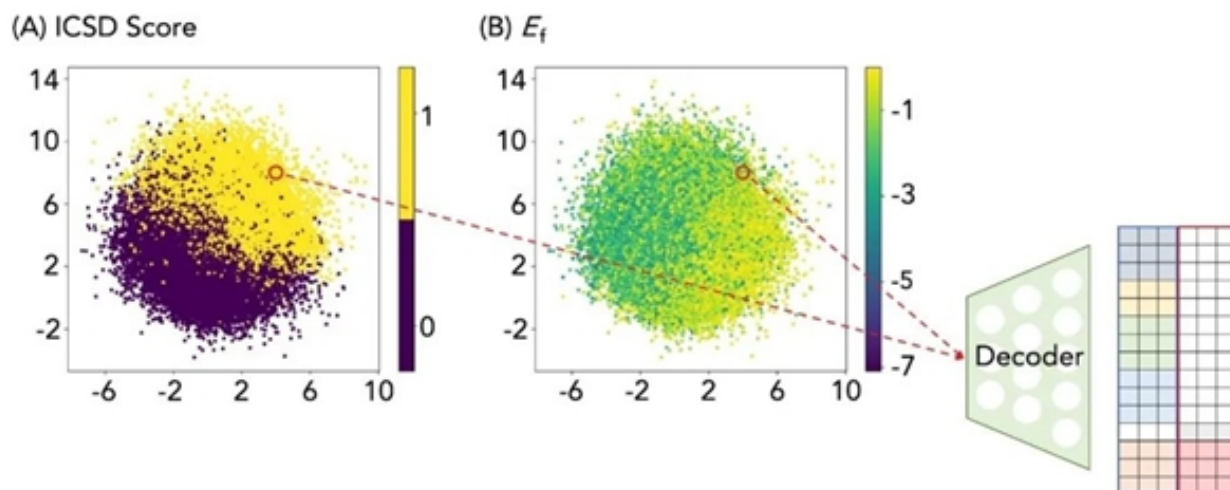


图4：用晶体是否收录在无机晶体结构数据库中（ICSD score）作为晶体稳定性的代替标准构建的筛选机制

使用者留心：如果您只需要从晶体结构到性质的预测模型，其他专门的晶体性质预测模型会有更好的表现（见<http://matbench.materialsproject.org>）。但如果您想找能够做泛化逆行设计的模型，可以考虑FTCP。就作者自己而言，我们希望看到FTCP能够启发可逆的基于图（graph-based）的晶体表征，使FTCP和图各自的优点结合起来。（注：基于图的神经网络在性质预测上是目前表现最好的模型之一，比如CGCNN，MEGNet等。）（来源：科学网）

相关论文信息：<https://doi.org/10.1016/j.matt.2021.11.032>

作者：Tonio Buonassisi 来源：《物质》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发