声明：本手稿由 UT-Battelle, LLC 根据与美国能源部签订的合同编号 DE-AC0500OR22725 撰写。美国政府保留相关权利；出版商接受本文发表，即承认美国政府拥有非独占、已付清、不可撤销、全球范围的许可，可为美国政府之目的出版或复制本手稿的已发表版本，或允许他人这样做。能源部将根据能源部公共访问计划 (http://energy.gov/downloads/doe-public-access-plan) 向公众开放这些由联邦资助的研究成果。

用于显微镜数据工作流的处理与分析系统：pycroscopy生态系统软件包

R. Vasudevan*[,1], S. M. Valleti[2], M. Ziatdinov[1,3], G. Duscher[4,**], S. Somnath[5,61]纳米相材料科学中心，橡树岭国家实验室[2]布雷德森跨学科研究中心，田纳西大学，诺克斯维尔[3]计算科学与工程部，橡树岭国家实验室[4]材料科学与工程系，田纳西大学，诺克斯维尔[5]国家计算科学中心，橡树岭国家实验室

[6]现任职于：麦肯锡公司，马萨诸塞州波士顿市国会街280号1100室，邮编02210

摘要

生物学和量子计算等领域的重大进展，都依赖于多种显微技术。随着新型探测器技术的出现，以及光谱学、成像和衍射的集成，光学显微术、电子显微术和扫描探针显微术均取得了进步。尽管这类仪器已大量涌现，但在采集数据集的处理、分析、存储和检索方面仍存在重大瓶颈。除了缺乏文件标准之外，各个领域专用分析软件包往往与底层数据集相互脱节，因此，最终用户仍需费力追踪分析和处理步骤，妨碍了可复现性。在此，我们介绍 pycroscopy 软件包生态系统，这是一个以通用数据模型为基础的开源 Python 生态系统。我们的数据模型称为 N 维光谱成像数据格式，在 pycroscopy 的 sidpy 软件包中得以实现。该软件包构建于 dask 数组之上，利用其数组特性，同时扩展这些特性，以加速显微术相关的分析和可视化。文中展示了若干使用 pycroscopy 生态系统创建数据摄取和分析工作流的示例。采用此类标准化例程对于推动下一代自主仪器的发展至关重要；在这类仪器中，处理、计算和元数据存储将对整体实验运行至关重要。

通讯作者

*vasudevanrk@ornl.gov

**gduscher@utk.edu

引言

过去三十年来，显微术的结构、功能和化学成像一直走在众多领域创新与发现的前沿，并成为推动纳米技术革命的关键工具。（Bian, et al., 2021; Fernandez-Leiro & Scheres, 2016; Pennycook & Nellist, 2011）事实上，显微术能够在多个长度和时间尺度上观察材料及其相关动态过程。这包括对从与 SARS-CoV-2病毒相关的刺突蛋白（Zhu, et al., 2021）到在锰酸盐表面跳跃的单个氧吸附原子（Bryant, et al., 2011; Vasudevan, et al., 2015b）等各种对象进行原子尺度研究。在许多显微镜中，进行局部光谱学测量是一项关键优势，它能够测定特定位置的局部化学、电子态或功能性质等。（Balke, et al., 2015; Newbury & Ritchie, 2013）随后，可将这些数据与从样品图像中获取的微观结构特征相关联。这一过程对于深入了解所测样品中的结构-性质关系至关重要。（Gázquez, et al., 2017）

尽管显微镜种类不断增加，且已广泛应用于众多科学领域，但许多现代显微镜在数据采集、存储、处理和分析方面的需求，如今已成为显著瓶颈。(Kalinin, et al., 2016; Spurgeon, et al., 2021) 即便是相对而言属于“小数据”的显微镜平台，例如原子力显微术（原子力显微镜），新近开发的多模态光谱学技术不断进步，加上能够直接从传感器采集未压缩数据流的能力，也大幅提高了可能的数据规模和数据速率，如今单个数据集达到 >100GB 已完全可行。(Collins, et al., 2016; Somnath, et al., 2016) 对于电子显微术而言，问题更加突出：快速探测器和诸如四维扫描透射电子显微术（4D-STEM）等模式不断涌现，单次实验即可采集数 TB 数据。(Ophus, 2019; Ophus, et al., 2014) 数据速率和数据规模的骤增意味着，依赖台式计算机且通常使用供应商编写的分析软件包的传统分析流程已不再可行。此外，现代显微术的许多进展得益于深度学习的应用 (Ziatdinov, et al., 2017) 及其他机器学习方法 (Schmidt, et al., 2019)，这些方法从数据清理 (Lee, et al., 2020)、改进函数拟合 (Borodinov, et al., 2019) 和图像分割 (Bermúdez- Chacón, et al., 2018)，到近期的自主显微术操作 (Kalinin, et al., 2021b;

Ziatdinov 等，2022)因此，数据集应能够在分布式系统或超级计算中心中使用。因此，随着自动化平台近期不断发展，为了最大限度地发挥显微术的效用，并减少传统实验范式中分析、存储、处理和检索方面的显著瓶颈，标准化数据模型和处理流程已势在必行。

截至目前，已有多个开源项目致力于解决这一问题：有的提供用于光谱数据和成像数据可视化与处理的工具，有的用于存储和文件处理，也有的兼具两者。相关项目数量众多，无法在此一一列举，并且因学科而异；但作者所知的部分示例包括 HyperSpy（de la Peña 等，2017年），其重点处理传统上在电子显微术平台上获取的高光谱数据；Hystorian（Musy 等，2021年），这是一个用于存储原子力显微术成像数据的软件包；以及 LiberTEM（Clausen 等，2020年），这是一个用于执行扫描透射电子显微术领域常见分布式计算的软件包。

与上述均为单独软件包的工作相比，本文介绍 pycroscopy生态系统，其以模块化方式处理从采集到处理的所有环节。此外，该软件包还包括将供应商指定格式转换为开放数据模型、可视化、处理以及文件写入。我们的核心理念包括六项原则：(1) 模块化，即我们提供一组可互操作、可扩展的模块化软件包，具有较低的使用门槛和较高的灵活性；(2) 一致性，即所有核心软件包均可接受并返回 sidpy.Dataset 对象（后文讨论）；(3) 易用性，即面向用户和开发者，提供用于可视化、创建和保存数据的一行式方法；(4) 可扩展性，即数据写入分层数据格式(Folk, et al., 2011) (HDF5)，且数据集对象构建于 Dask(Rocklin, 2015) 数组之上，因此能够利用分块，且不受内存限制；(5) 标准化，即采用开放的数据模型，能够表示不受来源、维度和大小限制的数据，且分析不绑定于特定数据类型；以及 (6) 可追溯性，即原始数据、中间数据和最终数据可写入同一数据文件并相互链接，以记录数据溯源。

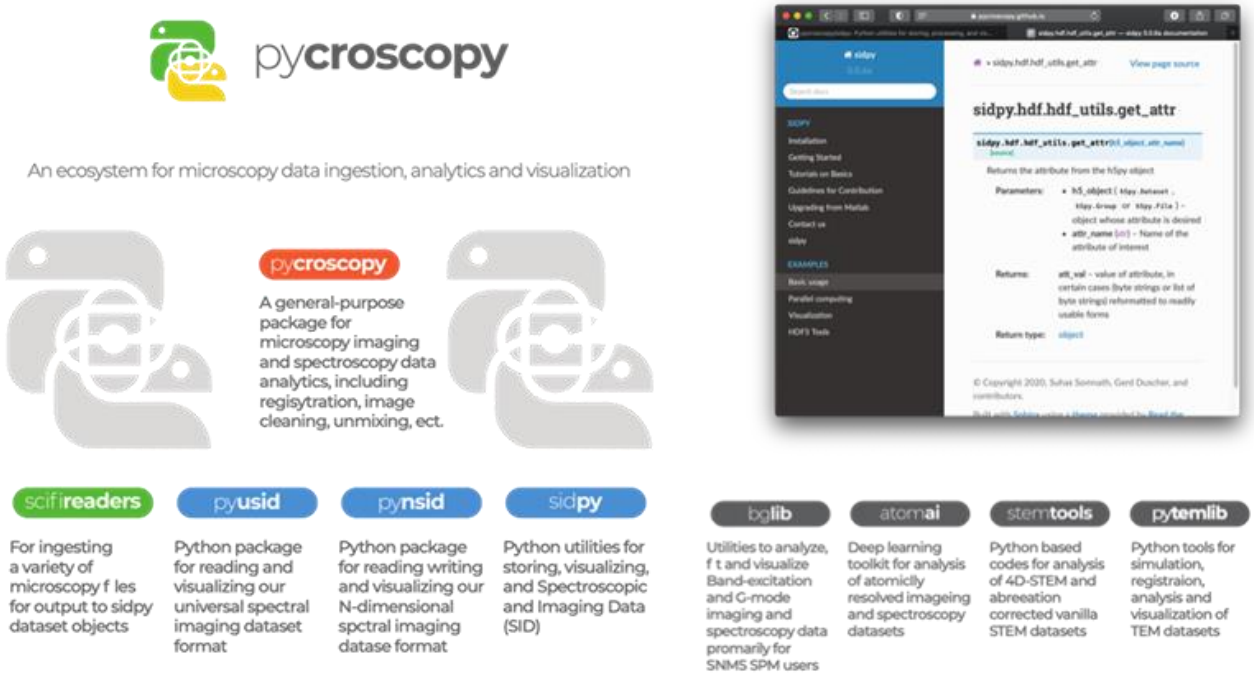

图 1：pycroscopy生态系统。绿色/蓝色软件包（SciFiReaders、Sidpy、pyNSID）用于数据输入、输出和可视化。AtomAI 和 BGLib 等领域专用工具（灰色）则用于更专业的分析或领域任务。

为实现这一理念，我们构建了可互操作的 Python 包，它们属于三个不同领域之一，即：(1) 数据与文件输入/输出及相关工具，(2) 通用成像与光谱工具，以及 (3) 领域专用包，如图1所示。数据输入/输出工具用于描述数据模型，还包括 SciFiReaders 包，用于从多种供应商格式导入数据并输出为 sidpy.Dataset对象。我们还提供了一个通用成像与光谱分析包，该软件包也命名为 pycroscopy。最后，诸如 BGLib 和 pyTEMlib 等领域专用包被纳入该生态系统，用于执行通常仅在各自领域内使用的特定定制分析流程。

数据模型：Sidpy 和 NSID

该生态系统的核心元素是 sidpy.Dataset 对象。这是一种 Python 对象，扩展了 Dask 包中的数组对象，并添加了特定功能，以便写入元数据、处理带有物理量和单位的维度，以及提供用于绘图和智能折叠（重塑）的扩展方法。sidpy 数据集对象通过 N维

维度光谱成像数据集（NSID）模型。这是一种通用模型，可接受任意 N维数值数据矩阵；与之不同的是，我们此前发表过通用光谱成像数据集（USID）模型。尽管 NSID 模型无法表示与原始 USID 模型同样多的用例（Somnath 等，2019）（只有在数据集存在 N 维表示形式时才能使用），但我们发现，它在易用性和通用性之间取得了良好的平衡。N维矩阵包含*sidpy* 数据集的数值数据，而每个 N 维度的附加信息由用户提供，包括该维度本身的数据、维度类型（例如“空间”或“光谱”），以及所测物理量和单位的标签。这些信息在内部以字典形式存储于数据集对象的 dataset._axes 属性中。最后，每个创建的数据集都可以指定一种数据类型，例如“光谱”、“光谱图像”或“图像堆栈”以及“四维图像” 。此数据类型属性有助于_ _ _确定调用数据集的 plot() 方法时应使用哪个绘图函数。如果数据集对象是由仪器采集的数据创建的，元数据将以字典形式存储在 dataset.original_metadata 属性中。随后，任何其他元数据（例如与数据处理相关的元数据）都可以存储在 dataset.metadata 属性中。因此，数据集对象能够表示通用成像数据和光谱数据、可视化数据、访问相关元数据，并了解每个坐标轴的参数。

sidpy.Dataset对象概述见图2a。图2b展示了pycroscopy生态系统的整体理念。为了深入了解该生态系统的功能，我们将介绍几个运用其多个组成部分的用例，突出展示其特性和优势。

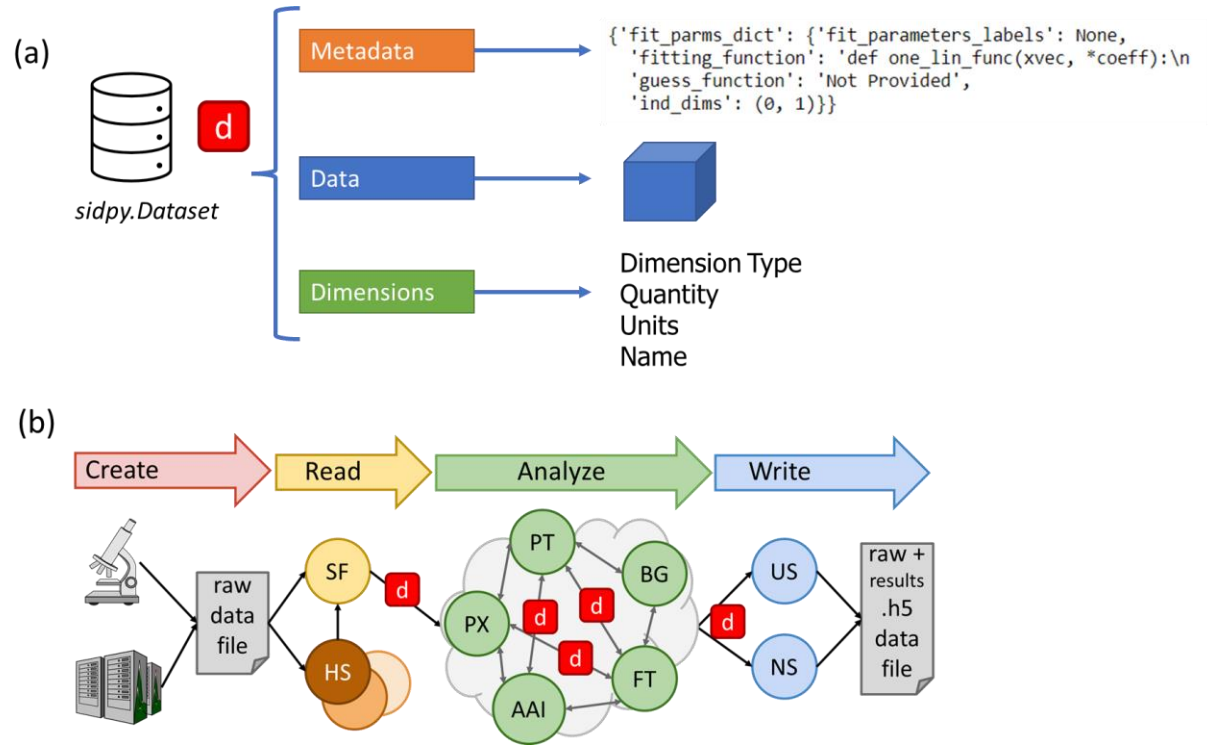

图 2：Sidpy 数据集及其生态系统。(a) Sidpy.Dataset 对象包含数据、元数据和维度信息。(b) pycrosopy 生态系统的总体理念是，在各个软件包之间，数据以 sidpy.Dataset 对象作为通用的“货币”。系统提供了专用工具，可将数据读入这些对象并将其保存到磁盘（目前我们使用 HDF5 文件来实现这一点）。

工作流程示例 1：扫描透射电子显微镜图像的纹理分析

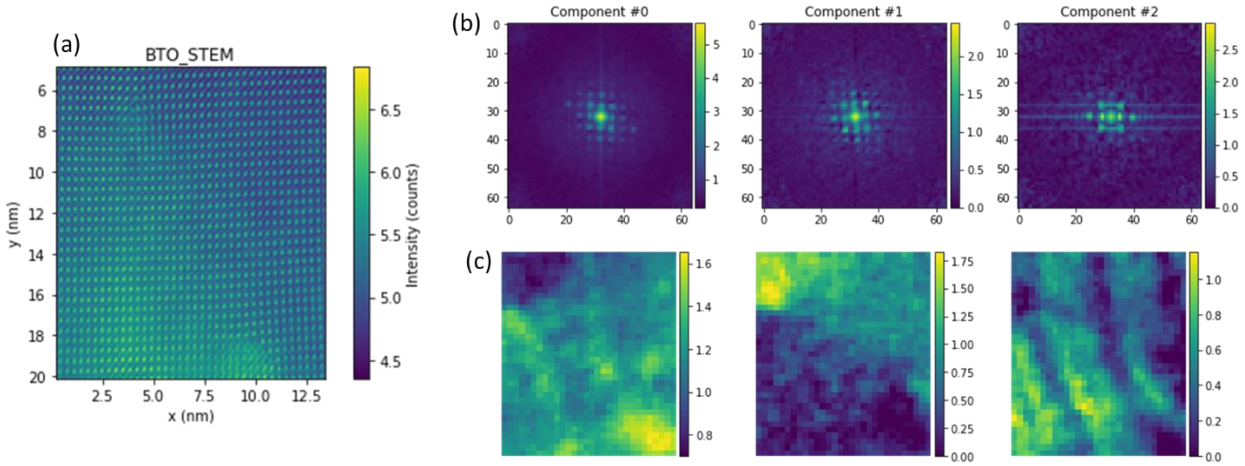

图 3：图像窗口化与矩阵分解。(a) $BaTiO_3$ 的扫描透射电子显微镜图像裁剪图。应用滑动窗口傅里叶变换后，我们可以对其进行矩阵分解

数据集。在此案例中，非负矩阵分解是所采用的方法，(b) 和 (c) 分别显示了分量和丰度。如先前报道的那样（Kelley 等，2022年），第三个分量中出现了带状结构，据推测这是由氧空位导致的。

获取成像和/或光谱数据后的首要任务是将数据转换为可用形式。对于 pycroscopy 生态系统，这意味着必须将供应商提供的数据格式转换为 sidpy.Dataset 对象。我们将此过程称为“读取”，并开发了 Python 软件包 SciFiReaders，以便从各种格式导入数据并将其读入 sidpy.Dataset 对象。随后便可在 pycroscopy 生态系统的其余部分中使用此对象，并充分利用 sidpy 数据集对象所提供的诸多优势。例如，请看图 3 所示的工作流程。这里展示的是先前发表的原子分辨数据：通过扫描透射电子显微术采集的 $BaTiO_3$ 薄膜数据。（Kelley 等，2022年）这些数据最初以 dm3 文件格式保存。我们创建 DM3reader 实例，并调用 read() 方法将文件转换为数据集对象。在交互式 Jupyter 笔记本中使用 .plot() 方法，可以立即绘制数据，并显示文件中包含的任何坐标轴信息。为了仅分析 $BaTiO_3$, 而不分析衬底，我们对数据进行了裁剪，并将数据写回 sidpy 数据集对象。裁剪后的图像如图 3(a) 所示。

接下来，我们对裁剪区域进行纹理分析，采用参考文献（Vasudevan 等，2015a）中介绍的滑动窗口变换。简而言之，通过在图像上滑动一个固定大小的窗口来生成图像块，因此在生成的数据集中，每个像素都与以该像素位置为中心的图像块相关联。此操作通过调用通用的 pycroscopy 软件包来完成，该软件包包含 ImageWindowing() 类。此过程的输出是一个四维数据集，但关键的是，它仍然是一个 sidpy.Dataset 对象。因此，只需调用该数据集的 plot()方法，即可立即使用交互式可视化器进行可视化。这对新手和资深用户都极为有用，因为它大大加快了可视化过程。在图像窗口化方法中，还可以选择执行局部二维傅里叶变换（快速傅里叶变换）。在本例中，我们对每个窗口块使用二维快速傅里叶变换，因为我们希望通过矩阵分解方法（如下文所述）研究对称性的变化。

图像窗口化有多种用途。其中一种用途是将其用于去噪：通过主成分分析或其他矩阵分解方法，使用较少的分量重构图像块，然后再将各窗口重新拼接成图像，正如我们先前所展示的那样。（Somnath 等，2018）此外，图像块还可用于训练神经网络，以关联局部结构与局部功能（即通过光谱测得的功能）。（Kalinin 等，2021a；Roccapriore 等，2021）第三，这些图像堆栈也可以通过组合分析每个图像块的二维快速傅里叶变换，再采用独立成分分析或非负矩阵分解（NMF）等技术进行分析。正如我们先前所展示的，这种方法能够确定原子分辨图像中的结构相，并绘制其空间位置及强度图。（Vasudevan 等，2016）为实现这一点，pycroscopy 软件包提供了一个 MatrixFactorization 类，它以 sidpy 数据集对象为输入，执行选定的分解，然后将结果（分量图和丰度图）分别作为独立的（sidpy）数据集提供。需要特别指出的是，该分解仅适用于二维（空间 × 光谱），因此必须按照这一预期模型对数据集进行重塑。在内部，sidpy 包含一种智能的“折叠”方法，可定位空间数据集和光谱数据集，并将任何包含 2 个空间维度和 (N-2) 个光谱维度的 N维数据集折叠为矩阵分解所需的格式。非负矩阵分解方法会生成分量图和丰度图，可在笔记本中将其可视化，如图 3(b,c) 所示。在本例中，分量 3 似乎呈现出带状特征，其四方性有所差异，这可能是由氧空位所致（请注意，这些数据和分析最近由部分作者用于一份已发表的报告中（Kelley 等，2022））。

工作流示例 2：电流-电压曲线的并行光谱拟合

作为第二个示例，我们展示一种光谱函数拟合的常见情况。在许多光谱数据集中，通常的工作流程是采集数据、对其进行可视化、执行光谱拟合，然后将拟合参数可视化。这里，我们重点关注先前在参考文献（Vasudevan 等，2017年）中报告的、在 $BiFeO_3$ 样品上获取的电流-电压曲线。数据从 HDF5 文件读取到 sidpy 数据集中，plot()方法支持通过单击和选择的方式对光谱进行可视化。如图 4(a) 所示。接下来，我们通过调用 sidpy 的 SidFitter 类来执行光谱拟合。该类以 sidpy 数据集和待拟合函数作为输入，

以及其他几个参数。还可以提供一个“猜测”函数，即一个将用于为实际函数拟合生成先验的函数。调用 fit() 方法可以利用底层 Dask 框架的并行处理能力，将拟合任务分配给指定数量的工作进程。由于这一过程在后端处理，用户无需了解任何并行处理知识即可获得显著的加速效果。这里还要指出，SidFitter 类提供了使用我们此前报告的“先验生成”策略的功能。（Creange 等，2021年）简而言之，该策略包括对光谱数据执行 k 均值聚类，然后对各簇均值执行函数拟合。随后，将拟合参数作为每个簇中各个成员的先验。这样不仅可以提高速度（因为优化过程可能需要更少的迭代次数），还可以获得更优的结果，这是因为避免了先验不佳导致优化器陷入局部最小值的情况。随后，可以使用拟合器类的 visualize_fit_results() 方法绘制结果。如图 4(b) 所示。

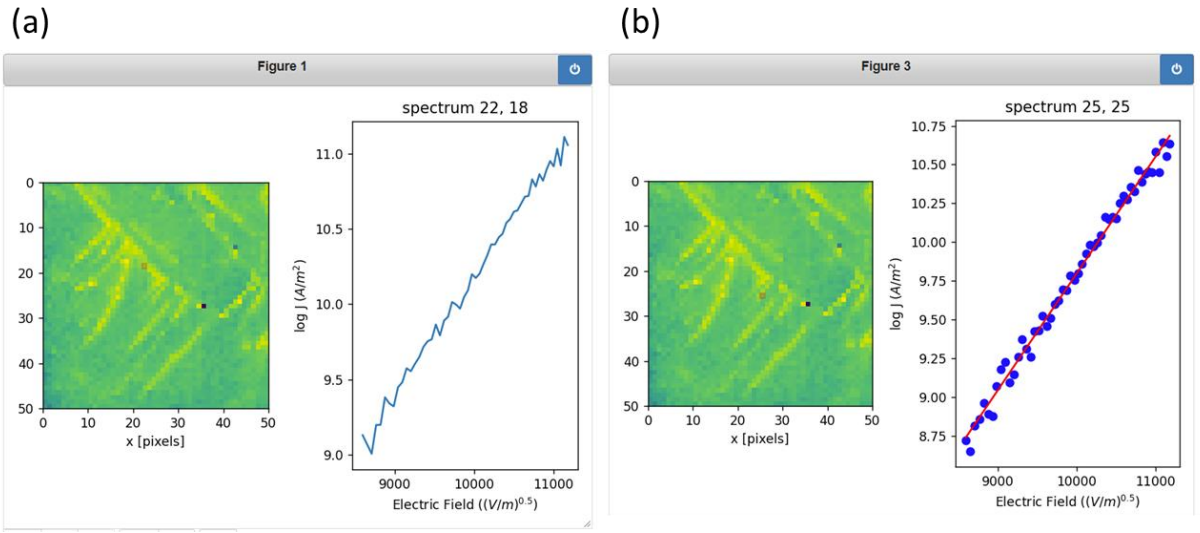

图 4：曲线拟合示例。图中展示了来自 $BiFeO_3$ 薄膜样品的电流-电压数据（数据引自参考文献 Vasudevan 等，2017年），使用 .plot() 命令可在 Jupyter 笔记本中进行交互式绘制和查看。定义待拟合函数后，调用 SidFitter 类即可启用并行拟合，并返回包含拟合结果的另一个 sidpy 数据集对象。这些结果可以与原始数据叠加显示，如 (b) 所示。

工作流 3 示例：图像配准

成像数据中一个非常常见的任务是进行图像配准。这里，我们使用 pyTEMlib 软件包进行了图像配准。首先，我们导入一个图像堆栈并显示其中的图像。内置可视化器支持使用鼠标滚轮在堆栈中的不同图像之间滚动；在本例中，这些图像是 $SrTiO_3$ 的扫描透射电子显微术图像堆栈。由于存在非零漂移，采集到的图像从一幅到下一幅会出现一些轻微偏移。随后，我们可以进行刚性配准或非刚性配准，以对齐堆栈中的所有图像。图 5(a) 展示了 $SrTiO_3$ 的高角环形暗场扫描透射电子显微镜图像。PyTEMLib 调用 simpleITK 软件包来实现微分同胚 Demon 非刚性配准。输出再次是一个 sidpy 数据集，结果表明，配准后，一幅图像与下一幅图像之间的差异已无法察觉。图 5(b) 展示了随时间变化的漂移拟合结果。应用 Lucy-Richardson 反卷积方法后，PyTEMLib 还可以通过简单的斑点查找器确定原子位置。后续拟合能够将位置精修到亚像素精度，并且文件工具可确保将数据写回 h5 文件。需要注意的是，还可以通过 pycroscopy 的 Atoms 模块，利用精修后的原子位置进行局域晶体学分析，但本文不对此展开讨论。

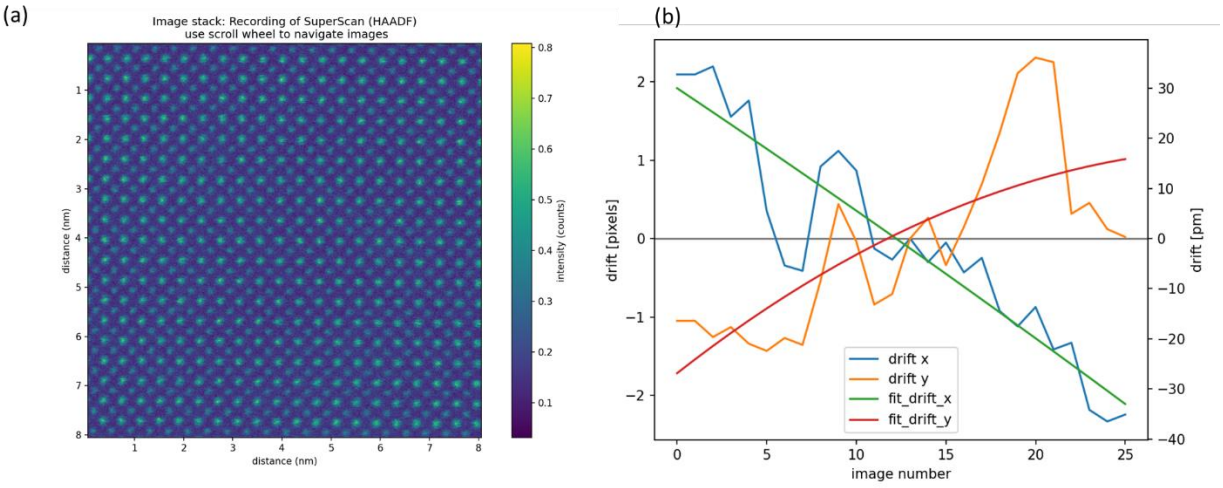

图 5：pyTEMLib 中的图像配准。(a) $SrTiO_3$ 扫描透射电子显微镜图像的图像堆栈。该堆栈包含 25 张图像。(b) 图像配准后，显示了 x 和 y 方向的漂移及其各自的拟合结果。

工作流 4 示例：EELS 光谱分析

最后，我们来看一个芯损失谱光谱分析的示例，这是 PyTEMLib 提供的一项功能。电子能量损失谱学（EELS）能够在扫描透射电子显微镜内对样品进行定量化学映射。PyTEMLib 提供了构建于 Jupyter 笔记本中的图形用户界面，从而可以完成 EELS 光谱分析。首先，可以按照图 6(a) 所示将光谱可视化，然后裁剪到相关能量标尺范围，以便进行后续处理。随后，用户可以选择样品中存在的元素，这些元素将用于拟合过程；用户还可以输入其他相关参数，如图 6(b) 所示。“拟合成分”按钮可对光谱峰进行拟合，并根据该拟合结果计算成分。所有这些操作都可以通过 PyTEMLib 中编写的弹出式图形用户界面在 pyTEMLib 内完成；随后，还可以将数据保存回 HDF5 文件，以确保可复现性。图 6(c) 展示了拟合窗口和结果作为示例，但我们鼓励用户在交互式 Jupyter 笔记本中自行探索这一功能。

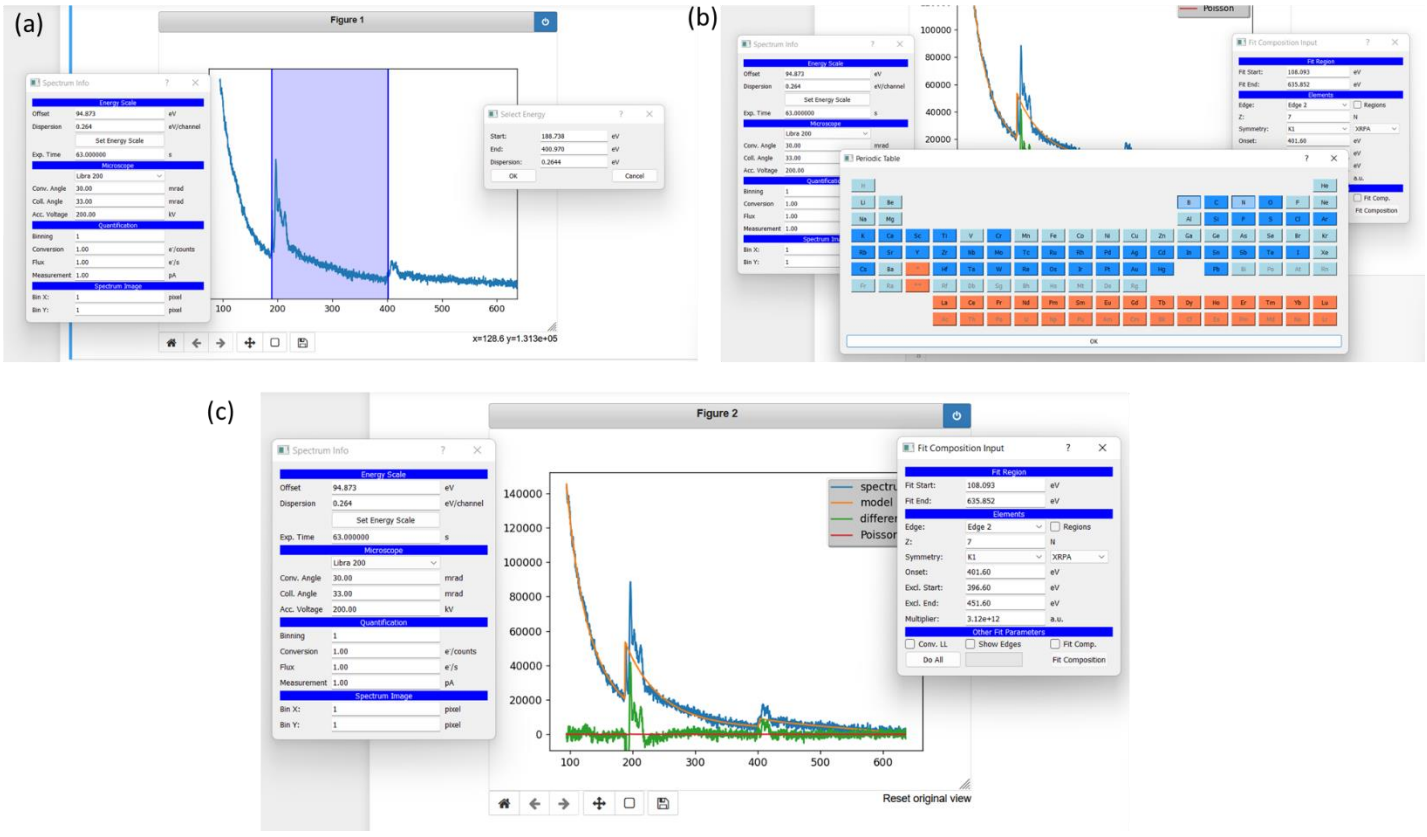

图 6：pyTEMlib 中用于分析 STEM-EELS 数据集的图形用户界面。(a) 可以选择光谱区域以进行进一步分析。(b) 用户随后可以选择样品中的元素，这些元素

将启用峰拟合。随后即可确定成分，如“拟合成分”窗口（见（c））所示。

讨论

上述示例说明，pycroscopy生态系统中的软件包如何有效用于为显微数据集生成可复现的科学数据分析工作流。该系统的一项主要优势在于，sidpy 数据集对象中内置了用于并行拟合、可视化和元数据存储的方法，并且能够快速将其写入磁盘。随着自动化与自主仪器（Hammer 等，2021；Kalinin 等，2021b）以及云实验室（Arnold，2022）的兴起，人们需要可复现的标准化工作流，这就要求具备支持这些工作的软硬件基础设施。例如，用于显微图像缺陷检测的机器学习模型训练，需要从海量数据中进行搜索，以检索并创建训练数据集。如果实验室或研究设施中的所有数据都实现标准化，并且能够便捷地访问元数据，这类搜索将会更加简单。在我们自己的研究设施中，遵循 NSID 或 USID 范式的 HDF5 文件都可以添加标签，并存储在由内部开发的软件解决方案 DataFed（Stansberry 等，2019）所管理的数据库中。仪器采集的图像和数据集可以推送至集中式存储库，并立即在云端触发处理流程，以确保数据被读取并转换为标准的 sidpy 数据集，提取所有元数据，创建 DataFed记录，并根据文件和实验类型运行相应的处理步骤。

随着合成和表征领域的实验平台日益自动化，我们预计工作流程将变得更加复杂，并涉及多台仪器。因此，未来 pycroscopy生态系统的工作不应只着眼于增加机器学习功能性，以及改进当前可用各软件包之间的集成，还应完善用于记录操作数据溯源的模式。总体而言，我们认为，与较为僵化的系统相比，当前系统具有很高的灵活性，因此降低了有意将该平台用于数据分析的用户所面临的门槛。

致谢

本研究得到纳米相材料科学中心的支持。该中心是橡树岭国家实验室内由美国能源部科学办公室设立的用户设施。

利益冲突声明

作者声明不存在任何利益冲突

数据与代码

用于复现此处结果的 pycroscopy Jupyter 笔记本可在 https://github.com/pycroscopy/pycroscopy中获取。与 pyTEMlib 相关的笔记本可在 https://github.com/pycroscopy/pyTEMlib

参考文献

ARNOLD, C.（2022年）。云实验室：让机器人开展研究。载于《自然》。BALKE, N., MAKSYMOVYCH, P., JESSE, S., HERKLOTZ, A., TSELEV, A., EOM, C.-B., KRAVCHENKO, I.I., YU, P. & KALININ, S.V.（2015）。利用扫描探针显微术区分铁电与非铁电机电效应。《ACS Nano》9(6), 6484-6492。BERMÚDEZ-CHACÓN, R., MÁRQUEZ-NEILA, P., SALZMANN, M. & FUA. 一种用于电子显微图像分割的领域自适应双流 U-Net。载于《2018 IEEE 第 15 届生物医学成像国际研讨会》（ISBI 2018），第 400-404 页。IEEE。BIAN, K., GERBER, C., HEINRICH, A.J., MÜLLER, D.J., SCHEURING, S. & JIANG, Y.（2021年）。扫描探针显微术。《自然综述：方法入门》1(1), 1-29。BORODINOV, N., NEUMAYER, S., KALININ, S.V., OVCHINNIKOVA, O.S., VASUDEVAN, R.K. & JESSE, S.（2019）。用于理解噪声数据的深度神经网络及其在扫描探针显微术物理性质提取中的应用。《npj 计算材料》5(1), 1-8。BRYANT, B., RENNER, C., TOKUNAGA, Y., TOKURA, Y. & AEPPLI, G.（2011）。对锰氧化物表面氧缺陷及其运动进行成像。《自然通讯》2(1), 1-6。CLAUSEN, A., WEBER, D., RUZAEVA, K., MIGUNOV, V., BABURAJAN, A., BAHULEYAN, A., CARON, J., CHANDRA, R., HALDER, S. & NORD, M.（2020）。LiberTEM：用于透射电子显微镜的可扩展多维数据处理软件平台。《开放源代码软件期刊》5(50), 2006。COLLINS, L., BELIANINOV, A., SOMNATH, S., RODRIGUEZ, B.J., BALKE, N., KALININ, S.V. & JESSE, S.（2016）。使用全数字 G 模式开尔文探针力显微镜进行多频谱分析。《纳米技术》27(10), 105706。CREANGE, N., KELLEY, K., SMITH, C., SANDO, D., PAULL, O., VALANOOR, N., SOMNATH, S., JESSE, S., KALININ, S. & VASUDEVAN, R.（2021）。传播先验，以实现更准确、高效的光谱功能拟合及其在铁电滞回中的应用。《机器学习：科学与技术》2(4), 045002。DE LA PEÑA, F., OSTASEVICIUS, T., TONAAS FAUSKE, V., BURDET, P., JOKUBAUSKAS, P., NORD, M., PRESTAT, E., SARAHAN, M., MACARTHUR, K.E. & JOHNSTONE, D.N.（2017）。hyperspy/hyperspy：HyperSpy 1.3。Zenodo。FERNANDEZ-LEIRO, R. & SCHERES, S.H.（2016）。利用冷冻电子显微术解析生物大分子。《自然》537(7620), 339-346。FOLK, M., HEBER, G., KOZIOL, Q., POURMAL, E. & ROBINSON, D. HDF5 技术套件及其应用概述。载于《EDBT/ICDT 2011 阵列数据库研讨会论文集》，第 36-47 页。GÁZQUEZ, J., SÁNCHEZ-SANTOLINO, G., BIŠKUP, N., ROLDÁN, M.A., CABERO, M., PENNYCOOK, S.J. & VARELA, M.（2017年）。扫描透射电子显微镜-电子能量损失谱在复杂氧化物中的应用。《半导体加工材料科学》65, 49-63。HAMMER, A.J., LEONOV, A.I., BELL, N.L. & CRONIN, L.（2021年）。化学计算与化学信息学的标准化。《JACS Au》1(10), 1572-1587。KALININ, S.V., KELLEY, K., VASUDEVAN, R.K. & ZIATDINOV, M.（2021a）。借助隐藏潜变量，探索铁电体中畴结构与功能性的关系。《ACS 应用材料与界面》13(1), 1693-1703。KALININ, S.V., STRELCOV, E., BELIANINOV, A., SOMNATH, S., VASUDEVAN, R.K., LINGERFELT, E.J., ARCHIBALD, R.K., CHEN, C., PROKSCH, R. & LAANAIT, N.（2016）。扫描探针显微术中的大数据、深度数据与智能数据。ACS Publications。KALININ, S.V., ZIATDINOV, M., HINKLE, J., JESSE, S., GHOSH, A., KELLEY, K.P., LUPINI, A.R., SUMPTER, B.G. & VASUDEVAN, R.K.（2021b）。电子显微术与扫描探针显微术中的自动化和自主实验。《ACS Nano》15(8), 12604-12627。KELLEY, K.P., MOROZOVSKA, A.N., ELISEEV, E.A., SHARMA, V., YILMAZ, D.E., VAN DUIN, A.C., GANESH, P., BORISEVICH, A., JESSE, S. & MAKSYMOVYCH, P.（2022年）。氧空位

注入：增强铁电体机电响应的途径。《先进材料》34(2), 2106426. LEE, S., NEGISHI, M., URAKUBO, H., KASAI, H. & ISHII, S. (2020). Mu-net：用于双光子显微图像去噪与复原的多尺度 U-net。《神经网络》125, 92-103. MUSY, L., BULANADI, R., GAPONENKO, I. & PARUCH, P. (2021). Hystorian：用于扫描探针显微术及其他 n 维数据集的处理工具。《超显微学》228, 113345. NEWBURY, D.E. & RITCHIE, N.W. (2013). 通过扫描电子显微术-能量色散 X 射线光谱法（SEM-EDS）对微观结构进行元素映射：硅漂移探测器（SDD）带来的非凡进展。《分析原子光谱学杂志》28(7), 973-988. OPHUS, C. (2019). 四维扫描透射电子显微术（4D-STEM）：从扫描纳米衍射到叠层成像及更多进展。《显微术与微量分析》25(3), 563-582. OPHUS, C., ERCIUS, P., SARAHAN, M., CZARNIK, C. & CISTON, J. (2014). 在材料科学中记录和使用 4D-STEM 数据集。《显微术与微量分析》20(S3), 62-63. PENNYCOOK, S.J. & NELLIST, P.D. (2011). 扫描透射电子显微术：成像与分析。施普林格科学与商业媒体。 ROCCAPRIORE, K. M., ZIATDINOV, M., CHO, S.H., HACHTEL, J.A. & KALININ, S.V. (2021). 纳米粒子组装体中局域等离激元响应的可预测性。Small 17(21), 2100181. ROCKLIN, M. Dask：基于分块算法的并行计算与任务调度。载于第14届 Python 科学会议论文集，第136页。Citeseer。 SCHMIDT, J., MARQUES, M.R., BOTTI, S. & MARQUES, M.A. (2019). 机器学习在固态材料科学中的最新进展与应用。npj Comp. Mater. 5(1), 1-36. SOMNATH, S., COLLINS, L., MATHESON, M.A., SUKUMAR, S.R., KALININ, S.V. & JESSE, S. (2016). 通过完整的悬臂动态检测实现成像：扫描探针显微术中的通用动态模式成像与光谱学。《纳米技术》27(41), 414003. SOMNATH, S., SMITH, C.R., KALININ, S.V., CHI, M., BORISEVICH, A., CROSS, N., DUSCHER, G. & JESSE, S. (2018). 通过相似性搜索进行特征提取：在电子显微术和扫描探针显微术成像中的原子识别与去噪应用。Adv. Struct. Chem. Imag. 4(1), 1-10. SOMNATH, S., SMITH, C.R., LAANAIT, N., VASUDEVAN, R.K. & JESSE, S. (2019). USID 和 pycroscopy–用于存储和分析成像与光谱数据的开源框架。《显微术与微量分析》25(S2), 220-221. SPURGEON, S.R., OPHUS, C., JONES, L., PETFORD-LONG, A., KALININ, S.V., OLSZTA, M.J., DUNIN- BORKOWSKI, R.E., SALMON, N., HATTAR, K. & YANG, W.-C.D. (2021). 迈向数据驱动的新一代透射电子显微镜。《自然·材料》20(3), 274-279. STANSBERRY, D., SOMNATH, S., BREET, J., SHUTT, G. & SHANKAR, M. DataFed：通过联邦式数据管理实现可复现研究。载于2019年国际计算科学与计算智能会议（CSCI），第1312-1317页。IEEE。 VASUDEVAN, R.K., BELIANINOV, A., GIANFRANCESCO, A.G., BADDORF, A.P., TSELEV, A., KALININ, S.V. & JESSE, S. (2015a). 倒易空间中的大数据：用于确定周期性的滑动快速傅里叶变换。Appl. Phys. Lett. 106(9), 091601. VASUDEVAN, R.K., CAO, Y., LAANAIT, N., IEVLEV, A., LI, L., YANG, J.-C., CHU, Y.-H., CHEN, L.-Q., KALININ, S.V. & MAKSYMOVYCH, P. (2017). 铁电畴壁处电子电导的场增强。《自然通讯》8(1), 1-10. VASUDEVAN, R.K., TSELEV, A., GIANFRANCESCO, A.G., BADDORF, A.P. & KALININ, S.V. (2015b). 通过扫描隧道显微术研究锰酸盐表面的原子尺度电化学。Appl. Phys. Lett. 106(14), 143107.

VASUDEVAN, R.K., ZIATDINOV, M., JESSE, S. & KALININ, S.V. (2016). 基于实空间原子分辨数据的相与界面：基于物理的深度数据图像分析。《纳米快报》16(9), 5574-5581. ZHU, X., MANNAR, D., SRIVASTAVA, S.S., BEREZUK, A.M., DEMERS, J.-P., SAVILLE, J.W., LEOPOLD, K., LI, W., DIMITROV, D.S. & TUTTLE, K.S. (2021). 冷冻电子显微镜解析的N501Y SARS-CoV-2刺突蛋白结构：其与血管紧张素转换酶2及两种强效中和抗体形成的复合物。《公共科学图书馆·生物学》19(4), e3001237. ZIATDINOV, M., DYCK, O., MAKSOV, A., LI, X., SANG, X., XIAO, K., UNOCIC, R.R., VASUDEVAN, R., JESSE, S. & KALININ, S.V. (2017). 对原子分辨扫描透射电子显微图像进行深度学习：化学鉴定与局部转变跟踪。ACS Nano 11(12), 12742-12752. ZIATDINOV, M., LIU, Y., KELLEY, K., VASUDEVAN, R. & KALININ, S.V. (2022). 扫描探针显微术的贝叶斯主动学习：从高斯过程到假设学习。ACS Nano 16(9), 13492-13512.