



A Framework for the Optimal Selection of High-Throughput Data Collection Workflows by Autonomous Experimentation Systems

Rohan Casukhela¹ · Sriram Vijayan¹ · Joerg R. Jinschek^{1,2} · Stephen R. Niezgoda^{1,3}

Received: 13 June 2022 / Accepted: 5 October 2022 / Published online: 31 October 2022
© The Minerals, Metals & Materials Society 2022

Abstract

Autonomous experimentation systems have been used to greatly advance the Integrated Computational Materials Engineering paradigm. This paper outlines a framework that enables the design and selection of data collection workflows for autonomous experimentation systems. The framework first searches for data collection workflows that generate high-quality information and then selects the workflow that generates the highest-value information as per a user-defined objective. We employ this framework to select the optimal high-throughput workflow for the characterization of an additively manufactured Ti–6Al–4V sample using a deep-learning based image denoiser. The selected workflow reduced the collection time of backscattered electron scanning electron microscopy images by a factor of 5 times as compared to the case study's benchmark workflow, and by a factor of 85 times as compared to the workflow used in a previously published study.

Keywords ICME · Materials informatics · Autonomous experimentation systems · Decision science · Workflow design/engineering · High-throughput experimentation

Introduction

Autonomous experimentation (AE) (including autonomous simulation) is being explored as a strategy to accelerate materials design and reduce product development cycles [1–5]. Autonomous experimentation is defined by Stach et al. as “...an iterative research loop of planning, experiment, and analysis [that is] carried out autonomously.” [1]. Materials AE research is a rapidly advancing field. Powerful

applications of AE in materials science include implementing Bayesian optimization principles in AE systems to quickly optimize material properties of interest [6–11], as well as utilizing AE systems to perform high-throughput experimentation (HTE) for rapid materials discovery and optimization for polymers, metals, ceramics, and more [12–19].

We consider AE in a broader context and pose a futuristic scenario where scientific discovery proceeds from a human investigator giving a simple command to an autonomous system, such as identifying the likely root cause of failure for an example component. While this thought experiment borders on science fiction, it is instructive to consider the steps the autonomous system must complete to arrive at a final conclusion. For this autonomous exploration to be carried out, the system must:

1. Parse the verbal instructions into a quantifiable objective that meets the requirements of the user
2. Identify the necessary information required to achieve the objective
3. Design a workflow to collect relevant information. In the materials realm this might include, for example, a sequence of testing, characterization, and simulation steps

✉ Stephen R.
Niezgoda
niezgoda.6@osu.e
du Rohan Casukhela
rcasukhela@outlo
ok.com Sriram Vijayan
vijayan.13@osu.e
du Joerg R. Jinschek
jojin@dtu.dk

¹ Department of Materials Science and Engineering, The Ohio State University, 2041 College Rd N, Columbus, OH 43210, USA

² DTU Nanolab, Technical University of Denmark, Kgs. Lyngby, Denmark

³ Department of Mechanical and Aerospace Engineering, 201 W 19th Ave, Columbus, OH 43210, USA

4. Design a sequence of experiments using the workflow from Step 3 to optimally gain information about the system
5. Execute the experiments to collect data
6. Extract information from data and assess if the objective is met
7. Iterate on Steps 3–6 if the objective is not met

Steps 1 and 2 fall in the realm of knowledge discovery via natural language processing (NLP) [20–23]. Steps 4 through 7 fall in the realm of optimal experiment design [6–11]. Both tasks are active areas in research. Step 3 involves the design and engineering of workflows. The focus of this writing is on expanding the capabilities of modern AE systems to complete Step 3 independent of human guidance or intervention. Here, we define a workflow as: “*the set of procedures, methods, and models used to observe physical/virtual systems*”. The working assumption here is that a specific objective set in Step 1 (e.g., grain size measurement, tensile stress/strain curve....) is established and that potential experimental and/ or simulation-based tools identified in Step 2 (e.g., microscopy imaging, hardness testing devices, heat flow models, FEM simulations, etc.) to extract the required information are in place. Workflow design consists of determining how best to use these experimental and simulation tools to collect information relevant to the objective. From here, Steps 4 through 7 can proceed normally, returning to Step 3 when necessary.

During materials/process development cycles, Step 3 is challenging for AE systems as it requires human-like domain knowledge of materials systems, as well as engineering properties of interest. Once the type of information to be gathered is ascertained, the natural next question is “How do we actually collect that information?” Current AE efforts start with the adoption of a human-designed experimental workflow that remains static throughout the entire process [24–26]. While some of these workflow decisions may seem trivial to human experts, even the most basic experimental procedures are typically outlined by detailed standards and operating procedures. Often, investigators will simplify a complex procedure with the understanding that the quality of the results does not significantly change. These modifications to procedures are often made to maximize repeatability, reduce time or cost of data acquisition, or account for different sources of variability. For example, unless testing is being performed for certification and qualification, most tensile testing is not performed to ASTM E8 specifications, even though the output of these tests is largely acceptable and reliable; tensile testing of miniature specimens is one such instance [27]. Another example is materials characterization by scanning electron microscopy (SEM): a human scientist operating an SEM relies on their

prior experiences and intuition to determine the method of sample preparation, magnification, field of view, accelerating voltage, beam current, contrast/brightness, and so on to take an image with high-quality microstructural information. It is a non-trivial question for the AE system to assess when a given procedure is “good enough” for the scientific objectives given resource constraints, even when the data objectives and basic procedures have already been delineated in Step 2 above.

Current advances in materials AE involve *a priori* selecting one workflow to measure the system, despite the possible existence of other workflows that might yield higher-certainty, higher-accuracy information at a lower cost of acquisition. An analogue to this issue can be seen in autonomous vehicles (AVs). The degree to which human intervention is required to operate an AV is classified by the SAE levels of driving automation (LDA) [28]. The LDA describes 6 levels of autonomy for an AV, ranging from a Level 0 AV having no driving automation, to a Level 5 AV having full driving automation. AVs use sensor networks to make decisions without human intervention. Current AVs are around Level 2 or 3 automation and do not decide what sensors to use or ignore in a given scenario. Instead, the priority and properties of the signals are predefined through a series of algorithms by the engineer. Human scientists and engineers design both the sensor networks and the autonomous system, thus dictating how it gathers and uses information. This approach has three notable advantages:

1. The information stream from the workflow is controlled and predictable
2. The autonomous system can quickly iterate through the objective space without having to potentially change tooling or account for the difference in measurements as the workflow is changed
3. The measurement process is repeatable and thus allows autonomous systems to maximize data throughput

The notable disadvantages of using a static workflow are that:

1. Changes to the workflow or potential improvements cannot be quantified by the autonomous system
2. A human must define the best method for information collection rather than the autonomous system

Current approaches to the design of materials AE systems severely limit their potential application space to tasks that are strictly defined by human operators, which are repetitive in nature and limited in scope. Due to the growing activity in materials AE [1–11], they must quickly adapt to rapid advances made in experimental/simulation/data-processing technologies [20]. Hence, advancing the decision authority

of AE systems is crucial for their continued design and relevancy.

In order to enable AE systems to select high-value data collection workflows independent of human scientists and engineers, we pose a framework reminiscent of multi-objective optimization techniques to dynamically identify the highest-value workflow that generates structured materials information:

1. An objective is established by the user to guide workflow development
2. The procedures, methods, and models that will be considered in the workflow are listed by the user
3. A fast search over the space of possible user-defined workflows is conducted to quickly filter for high-quality workflows in the context of the objective
4. A fine search over high-quality workflows is conducted to select the optimal workflow

The concept of this framework will be described in detail, then illustrated in a case study. There, the impact of a deeplearning based denoising algorithm on a materials characterization workflow is examined. The framework was used to algorithmically select the optimal high-throughput workflow that collects backscattered electron scanning electron microscope (BSE-SEM) images on the material sample approximately 85 times faster compared to the previous study [29], and 5 times faster than the Ground-Truth workflow of the presented case study. Lastly, summary statistics for the information stream of the selected high-throughput workflow are provided.

Workflow Selection Framework

Motivation

All data collection efforts must begin with specifying an objective that needs to be met. A well-designed **Workflow** generates relevant **Information** that adds significant **Value** to the broader objective.

Workflow → **Information** → **Value** (1)

Extracted information is an objective-dependent summary of the raw data (number of pores within a sample, average number of cracks in a part, etc.), that describes the system under investigation. If the workflow generates high-value information, the use of black-box data processing/transformation methods (such as neural networks) is justified in any workflow that utilizes them. Thus, the subtleties of how information is extracted from data are ignored.

We use the direct product of the workflow, the extracted information, as a measure of the value of the workflow itself. The **Value** of information is proportional to the information's **Quality** and **Actionability**.

$$\text{Value} \propto \text{Quality} \cap \text{Actionability} \quad (2)$$

Actionability is a user-defined decision function that explains how useful information is in achieving a particular objective. High-actionability information is critical to making high-value decisions. For instance, the ground-truth defect density of a part to estimate the overall mechanical stability in a critical application is information of high-actionability. Sometimes, collecting highly-actionable information (such as the ground-truth) can be expensive or intractable. In these cases, the cost for collecting this information must be assumed to be infinite and other, lower-actionable types of information must be sought. For example, collecting 3-dimensional estimates of the volume fraction of a material sample can be very difficult, and so one may have to rely on estimates of area fraction obtained from 2-dimensional images instead. In contrast to high-actionability information, low-actionability information is less useful to making high-value decisions.

The **Quality** of information is proportional to its **Accuracy** with respect to a pre-determined ground truth and the number of unique data **Sources** from which it is harvested, while being inversely proportional to the **Cost** of acquisition.

$$\text{Quality} \propto \frac{\text{Accuracy} \cap \text{Sources}}{\text{Cost}} \quad (3)$$

In general, increasing the **Accuracy** and/or the **Sources** reduces the uncertainty about the system that is under investigation. Therefore, the two quantities are related to the amount of valuable information that the workflow generates. However, increased **Accuracy** and the **Sources** typically leads to increased **Cost** of acquisition. This is due to the extra time needed and the extra resources required to collect, structure and curate the data [30].

The approach for any effort should be to select the highest-value workflow from a set of high-quality data collection workflows. High-quality workflows generate information that strikes a balance between the information's accuracy, certainty, and cost. A high-quality workflow that generates high-actionability information is considered to be a highvalue workflow. As an example, high-throughput experimental workflow design aims to select a workflow that is considered high-quality and generates a given amount of information in the shortest amount of time at the lowest possible cost. In this study, we show that the setup and design of workflows can be addressed as a two-stage optimization problem.

Mathematical Definition of Framework

It is very challenging to find the highest-value experimental workflow in one step. Therefore, we conduct the search for the highest-value workflow in a two-stage approach: In *Stage I* we seek to filter for workflows that generate high-quality information. In *Stage II* we select the highest-value workflow from the set of workflows obtained in *Stage I* by maximizing the information's **Actionability** (see Fig 1).

Let x be the data obtained from the collection process, dependent on the data collection settings θ , A be the data processing sequence that is applied to data $x(\theta)$, with data processing parameters λ , the product $A_{\mathcal{X}}(\theta)$ be the extracted information from the workflow, and M be the design specification, or design parameter that we designate as ground-truth.

Stage I : $\theta^*, \lambda^* = \operatorname{argmin}_{\theta, \lambda} (C_1 ||A_{\mathcal{X}}(\theta) - M||$

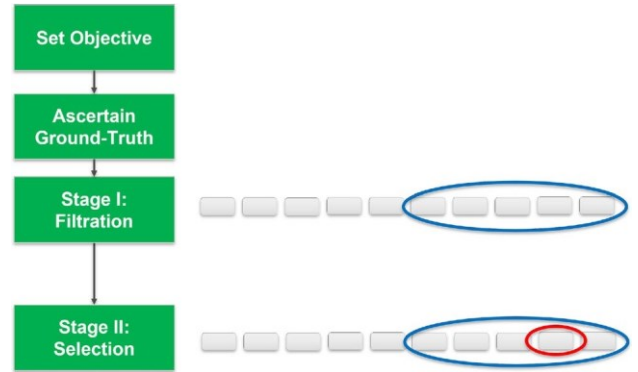
$$+ C_2 \mathbf{Cost}(x(\theta)) \quad (4)$$

$$+ C_3 \mathbf{Complexity}(A_{\mathcal{X}}(\theta)), \quad \{C_1, C_2, C_3 > 0\}$$

Stage II : $\operatorname{argmax}_{\theta, \lambda} (\mathbf{Actionability}(x(\theta), A_{\mathcal{X}}(\theta))) \quad (5)$

For *Stage II*, report the bias and change in standard deviation for the selected workflow's information stream when compared to Ground-Truth workflow.

- $A_{\mathcal{X}}(\theta)$ and M are compared in the *Stage I* objective function. This comparison is a measure of **Accuracy**. Bias is a measure of non-accuracy, and is used to characterize and compare workflows.
- **Cost** is a term that accounts for how expensive it is to collect the data x with an instance of data collection parameters θ . A cost value can be assigned by the user for each potential step in the workflow. The total **Cost** for a single workflow can be then determined by summing over each step's individual cost value.
- **Complexity** is a term that accounts for A 's complexity, computation time/resources, curation time (which can



increase due to the number of sources), and interpretability given an instance of data processing parameters λ . A simple example would be to use the term $p+q^2$, where p approximates the time/space complexity of algorithms used, and q corresponds to the number of sources of data.

- C_1, C_2 , and C_3 are *user-defined* weights that can be used to adjust the importance of each term in the objective function. Larger numbers imply more importance, and smaller numbers imply less importance.

In *Stage I*, we find a set of candidate experimental workflows with data collection settings θ^* and data processing parameters λ^* that minimize the objective function. This ensures that only workflows that generate high-quality information remain.

Stage II directly compares the candidate workflows with data collection settings θ^* , and data processing parameters λ^* , and finds the workflow among these that maximizes the user-defined **Actionability**. For high-throughput workflow design, **Actionability** can be defined as collecting the most amount of data in a given amount of time. We use this as part of the definition of **Actionability** for our *Stage II* criteria in our case study (see Sect. 3).

The *Stage I* objective function describes the quality of the information generated by the workflow: smaller scores of the function will correspond to a higher-quality workflow, while larger scores of the function will correspond to a lower-quality workflow. The goal is to minimize the objective function as much as possible. Regions of the same objective values, outline *iso-quality* regions. Iso-quality regions imply that both workflow settings produce information of the same quality.

Stage I's objective function is inherently “noisy,” meaning that repeated measurement of the experimental workflows will generate different *Stage I* objective values. This is the reason why *Stage II* is required: *Stage I* is a coarse search that informs the user of what regions to investigate further. *Stage II* is a fine search and guides the user to quan-

Fig. 1 Schematic of the framework showing the two-stage approach when searching for the highest-value workflow

processing parameters λ . For each data collection parameter θ , its collection cost must be defined and for each data processing parameter λ , its processing complexity must be defined. Second, experiments should be conducted to seek a set of well-established, well-understood “ground-truth” measurements that can be used to compare a workflow’s ability to produce accurate information. Third, the accuracy, cost, and analysis complexity term weights C_1, C_2, C_3 should be properly defined. Fourth, a *Stage I* search should be conducted. This can be done by collecting extracted information across all workflows being considered and using the extracted information to score each workflow. It is recommended that few measurements for each workflow are collected in this step, as the *Stage I* exploration space can be large. *Stage I* will filter the workflow space and produce a set of candidate workflows. Workflows that minimize the objective function or are close to the minimum should be included within the candidate set. Fifth, a *Stage II* data collection effort using the candidate workflows for finer comparison should be conducted. Because a *Stage II* search needs to be higher-resolution than *Stage I*, it will require more data per workflow than in the *Stage I* search. The *Stage II* search can be conducted via a formal statistical experimental design, such as one-way ANOVA. From here, the workflow that maximizes **Actionability** can then be selected. This fifth and final step will yield the highest-value data collection workflow.

Case Study: Designing a High-Throughput

tify the bias and variance of the different workflows. *Stage II* allows for the possibility that *Stage I*’s objective function values are random. In practice, *Stage II* takes candidate workflows with similar objective function values as input, compares the workflows, and selects the workflow maximizing **Actionability**. To give an example in the context of high-throughput workflow design, if a meaningful difference between the workflows is found, a decision can be made to select the workflow closest in mean to the GroundTruth workflow that also minimizes the acquisition time of information.

To use this framework, the user must first define the objective, the types of information to extract, all equipment/models/procedures to be potentially used for information extraction, and all data collection parameters θ and data

Workflow for Expediting Microstructural Characterization of AM Builds

We now present a case study in which we design a highthroughput workflow for the microstructural characterization of an additively manufactured (AM) sample using BSESEM images. The AM sample was fabricated in a previous study that examined and quantified the microstructural variation between builds fabricated using different beam scanning strategies [29].

Using SEM characterization is particularly interesting in the research of high-throughput workflows for AM applications because SEMs can be used to capture the subtle variation in AM builds and generate large amounts of high-value datasets. Thus, developing SEM-based high-throughput workflows for AM applications is a desirable goal.

An SEM image is acquired in a pixel-by-pixel approach, while the number of pixels has an upper limit determined by the SEM’s scan engine. The electron beam rasters in lines over an area of interest pixel-by-pixel with an operator-chosen dwell time at each of the steps. This dwell time has a huge impact on the signal-to-noise ratio (SNR) of the final image. An operator-chosen image magnification determines the number and size of the pixels in relation to the feature size in the sample (e.g., size of a grain or a defect) that is of interest. In principle, the image acquisition time is determined by the chosen number of pixels in an image multiplied by the chosen dwell time per pixel. Integration of multiple fast-scanned images (known as frame integration, or FI for short) can be used to boost the SNR of the image. The constraints are that the pixel size needs to be optimal in

Table 1 *Stage I* values for workflows with varying frame integration (FI) and dwell time (DT) settings

768 × 512 pixel resolution: dwell time (μs)											
	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.5	2	3
Frame integration											
1	4704	11,296	12,456	12,762	13,673	16,349	12,866	4649	1331	803	226
2	4700	9171	4866	3463	3204	2213	3818	357	636	174	42
4	2247	2984	834	2331	720	840	190	95	216	64	83
8	490	157	174	453	151	148	119	84	100	117	164
16	200	94	95	97	95	147	124	114	164	218	330
32	87	114	120	154	173	194	226	218	328	437	664
64	146	197	245	274	318	357	401	436	654	872	1313

Lower values indicate higher-quality workflows

The bolded values represent the workflows chosen for the *Stage II* selection process

relation to the size of a feature that needs to be resolved, and the dwell time needs to be sufficiently high to achieve a required SNR to recognize the feature in the subsequent data analysis process.

The trade-off between dwell time and FI vs. image quality is an example of a compromise one has to accept when attempting to utilize SEMs for high-throughput data collection (see Fig. 2): decreasing the cost of data acquisition (i.e., taking images faster with a shorter dwell time) decreases image quality and might therefore lead to greater deviations from a specified M (ground-truth).

Recent advances have shown that deep-learning algorithms can be used to boost the SNR of low-quality microscopy images and yield images with effectively higher SNR [31–37]. These methods have exhibited robust performance on images outside of their testing datasets when compared to more standard denoising techniques. Deep learning-based denoising algorithms provide a straightforward solution to the aforementioned problem of taking SEM images for high-throughput data collection purposes; one can take images with low SNR or low pixel resolution and still retain the quality of higher-SNR, higher-resolution images. The

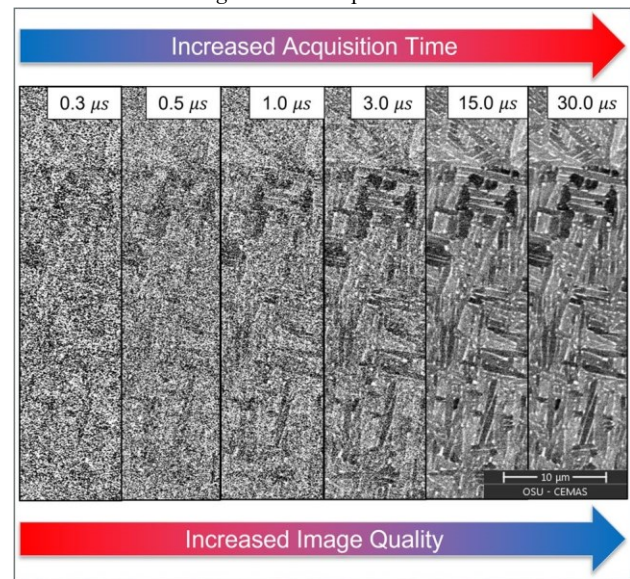


Fig. 2 Backscattered electron (BSE) scanning electron microscopy (SEM) images taken with increasing dwell time visualizing one of the many trade-offs between acquisition time vs. image quality question then becomes, “what is the lowest image quality we are willing to accept while increasing our throughput of data?”

The objective is to design and implement a high-throughput workflow utilizing an algorithm that denoises BSE-SEM images to systematically characterize the AM part examined in Shao et al. The extracted information for this case study will be the size of the Ti-6Al-4V (Ti64) α -lath (i.e., α -lath thickness).

Material Sample

Details about sample fabrication have previously been reported [29]. The center region of the center-top XY Ti64 sample fabricated using a linear scan (LS) strategy is used for this case study (see Fig. 3a in Shao et al).

Denoising Algorithm

The BSE denoiser model presented here is a convolutional neural network (CNN) with the U-Net architecture. Further model details can be found at the following references [31, 38].

Stage I Search

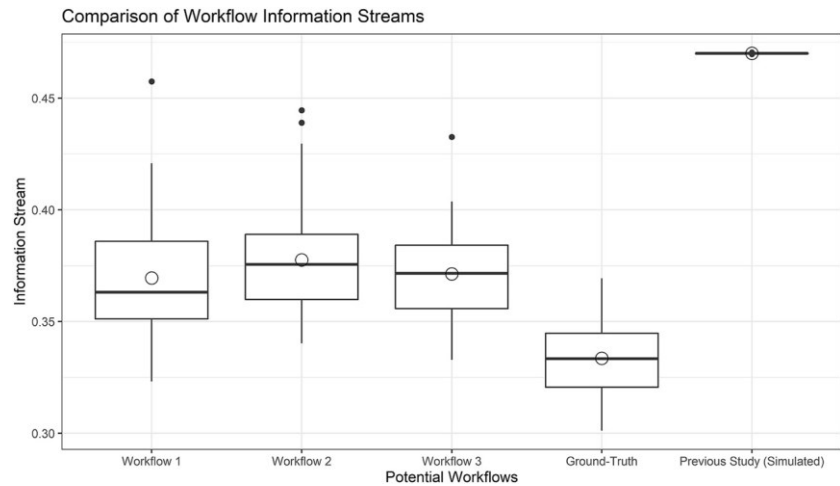
Here, C_1 is 500, the distance term between Ax and M is defined as the squared distance between the two terms, C_2 is 6.81, the cost term is defined by a combination of settings that relate to the acquisition time of a micrograph, and the model complexity term, C_3 , is set to 0. C_3 is set to 0 because all images for this case study have the same pixel resolution, and therefore the resources needed to process all images is the same. Therefore, the model complexity term is constant and does not affect the minimization of our *Stage I* objective function for this case study.

The optimization problem:

$$\begin{aligned} \text{Stage I : } \quad & \operatorname{argmin}_{\theta, \lambda} (500(Ax(\theta) - M)^2_{\theta, \lambda} \\ & + 6.81(\text{dweltime} \cdot \text{frameintegration})) \end{aligned} \quad (6)$$

Stage II : $\operatorname{argmin}_{\theta, \lambda} (\text{Actionability}(x(\theta), Ax(\theta)))$ (7)
was used for the study. For *Stage II*, we will report the bias and standard deviation change of the selected workflow's information stream when compared to the Ground-Truth and

Fig. 3 Boxplots of information streams for the candidate workflows as compared to the Ground-Truth workflow. An information stream for the Previous Study workflow was simulated using the statistics reported in [29]. The bolded lines represent the median α -lath thickness, while the hollow circles represent the average α -lath thickness



Previous Study workflows. Here, $x(\theta)$ is an image taken of the sample taken at image acquisition settings θ [29]. M is the agreed-upon estimate of what the α -lath thickness in the Ti64 sample is. The α -lath thickness was extracted from an

image at 30 μ s dwell time, 1 FI, and 768 $\times$ 512 pixel image resolution to act as the ground-truth setting M . A_λ is the imaging, denoising, and segmentation process being considered, and $Ax(\theta)$ is the α -lath thickness extracted from the image. We place the greatest emphasis on the comparison between $Ax(\theta)$ and M , as evidenced by the weight of 500 on the first term.

Given a fixed model complexity value, we define highlyactionable information as information that minimizes the cost to acquire data as well as the bias between collected data and agreed-upon metric M .

As given in Table 1, one image $x(\theta)$ for each workflow with settings (θ, λ) . The α -lath thicknesses $Ax(\theta)$ for each of the dwell times and FI were extracted from the images through an image processing sequence A_λ . Here, we fix λ by using the same image processing workflow for all images. The resulting scores were calculated using the objective function defined above.

Table 1 shows *Stage I* results of the workflow search. As the *Stage I* objective function is “noisy” (see Sect. 2.2), the workflow corresponding to the lowest objective function value in this figure should not immediately be taken to be the most optimal value. However, based on this *Stage I* search, it is obvious that some workflows are more preferable than others. One should conduct a *Stage II* search with workflows close to and equal to the lowest *Stage I* score to investigate which workflow is truly preferable. The number of potential workflows to check will vary based on the application.

Stage II Search

A total of 3 candidate experimental workflows were selected as per the *Stage I* results and compared against the

GroundTruth workflow. Using a fixed 768×512 pixel image resolution, these were:

- Workflow 1: 0.7 μ s dwell time, 16 times FI (4.5 s acquisition time)
- Workflow 2: 2 μ s dwell time, 4 times FI (3.2 s acquisition time)
- Workflow 3: 3 μ s dwell time, 2 times FI (2.4 s acquisition time)

Workflow 3 was selected as it had the lowest *Stage I* score in Table 1. Workflows 1 and 2 were selected because they had *Stage I* values close to Workflow 3's minimal *Stage I* value.

The Previous Study and Ground-Truth workflows are listed below:

- **Previous Study:** 30 μ s dwell time, 1 time FI at 3072×2048 pixel resolution (204.7 s acquisition time)
- **Ground-Truth:** 30 μ s dwell time, 1 time FI at 768×512 pixel resolution (11.9 s acquisition time)

A comparison of the above workflows' information streams can be seen in Fig. 3. 35 instances of information each were collected for Workflows 1–3 and the Ground-Truth workflow, and 10000 instances of simulated information were drawn to emulate the Previous Study workflow.

We note that Workflows 1–3 and the Previous Study workflow report a higher average α -lath thickness than the Ground-Truth workflow, and that Workflows 1–3 and the Ground-Truth workflow report a lower average α -lath thickness than what the Previous Study workflow reported for the same sample [29]. Additionally, Workflows 1–3 also report a higher α -lath thickness standard deviation than the GroundTruth and Previous Study workflows. Lastly, Workflows 1 and 2 exhibit some right-tailed skewness (as their mean is greater than their median), indicating that results achieved using these workflows would be influenced by the SNR lower limit for this system.

Based on our definition of **Actionability**, we sought to determine whether or not the information gathered using the three chosen candidate workflows were significantly different from one another. This would inform us if there is a meaningful difference in bias between Workflows 1–3 as compared to the Ground-Truth and Previous Study workflows. To investigate this, we conducted a repeated measures ANOVA (after checking all assumptions—no extreme outliers, normality in the response, and sphericity of variance) to examine differences in the workflows' information streams. Pairwise comparisons were conducted,

with the null hypothesis asserting that the mean α -lath thickness for the information streams of Workflows 1–3 are the same, and the alternative hypothesis asserting that at least one workflow's information stream had a mean α -lath thickness that was different compared to the others. The condition for rejection of the null hypothesis was derived using a family-wise significance level of 99% and a Bonferroni correction, which means that each of the 3 comparisons made would have to have a p value less than $(1 - 0.99)/3 = 0.003$ to be considered as statistically significant enough to reject the null hypothesis. None of the comparisons made yielded p values less than 0.003. Therefore, the null hypothesis was not rejected, and the information streams of the three workflows were considered to produce equivalent information. From this analysis, it was concluded that the workflows deliver similar results compared to each other (see Table 2), and Workflow 3 could be chosen as the most valuable high-throughput workflow,

as it had the lowest acquisition time of Workflows 1–3 (Workflow 3: average α -lath thickness = 0.37 μ m, standard deviation = 0.02 μ m (see Table 3 for comparison statistics of Workflow 3’s information stream)).

Discussion

Comparison of Workflows with Different Data Dimensionality

The previous study used a workflow that generated images with 3072 $\times$ 2048 pixel resolution [29]. The presented case study examined workflows that generated images with only 768 $\times$ 512 pixel resolution. The case study is an example where the **Actionability**, cost, and complexity associated with acquiring information equivalent to the previous study is very high. When this happens, a workflow with an information stream of lower **Actionability**, cost, and complexity must be used in order to proceed. As the *Stage I* objective function yields workflows that balance accuracy with cost and complexity, the difference in determined Ti64 α -lath thickness between Workflow 3, the Ground-Truth workflow, and the Previous Study workflow is accounted for. In this case study, we focus on calibrating our potential high-throughput workflows to the specified Ground-Truth workflow for this case study, and simply report the bias and variance of each workflow based on the reported GroundTruth workflow as well as the Previous Study workflow (see Table 3). As long as data collection proceeds with an understanding of the bias and variance between workflows, continuity and reproducibility between both studies is achieved.

Applying the Framework Beyond the Case Study

Understanding **Actionability** in the Context of Decision-Making

Actionability is how one encodes decision-making intuition into an AE system. The AE system makes decisions about workflows by maximizing the **Actionability** function. However, different objectives will yield different defi-

nitions of **Actionability**, and therefore, the result of a *Stage II* search will vary based on the user-defined objective. The framework also can be generalized to yield high-

Table 2 Results of the statistical comparisons conducted between Workflows 1–3

Comparison	<i>p</i> value
------------	----------------

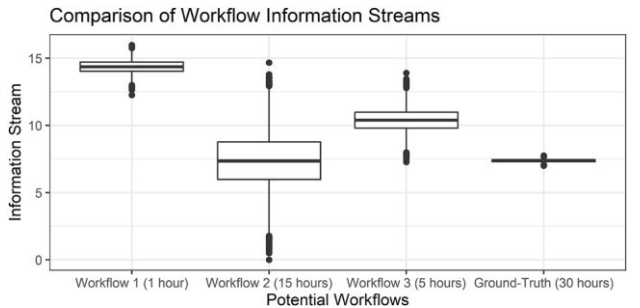
Table 3 information streams of Workflow 3 and the information streams of Wor workflow

Workflows being compared	Bias (μ m)	Standard deviation change (μ m)
Workflow 3—Ground-Truth	+ 0.03	+ 0.005
Workflow 3—Previous Study	– 0.10	+ 0.01

accuracy workflows and low-variance workflows due to the uncertainty quantification measures and accuracy estimates that the *Stage II* selection process yields. The schematic in Fig. 4 shows potential workflows that could be selected based on different objectives. Workflow 1’s variance and acquisition time are low, but the bias is large—this would have to be taken into consideration if Workflow 1 is selected for use. Workflow 2’s variance is high, but has no bias and has a lower acquisition time compared to the Ground-Truth workflow. This workflow might be considered if accurate results are required at a faster pace than the Ground-Truth workflow. Workflow 3 strikes a balance between the previous two workflows, with a bias, variance, and acquisition time between Workflow 1 and Workflow 2. Workflow 3 might be considered as a moderate choice between Workflows 1–3. Additionally, it may be decided that very precise measurements regardless of acquisition time are required and so the Ground-Truth workflow may also be selected.

Selection of the *Stage I* Objective Function Weights

The weights for the *Stage I* objective function are currently chosen by human beings, and not the automated system. This is intentional; changing the weights of the parameters of the *Stage I* objective function filters for workflows that prioritize different facets of informational quality. This allows the framework to yield a set of high-value workflows with



Workflow 1–Workflow 2	0.01
Workflow 1–Workflow 3	1.00
Workflow 2–Workflow 3	0.50

varying accuracy, cost, and complexity. Larger weights on a term will lead to more importance of that term in the overall *Stage I* objective function. For example, assigning larger weights to FI will skew the optimal results to conditions that incorporate lower FI. In practice, this choice of weights adds flexibility and allows one to prioritize conditions that best fit their requirements. As an example, investigations of critical components in aerospace applications demand a very accurate estimate of the ground-truth. These applications can still use the framework presented to select high-throughput workflows. In such a case, one may assign a relatively large weight to the bias term, while comparatively smaller terms are assigned to the cost and complexity terms. The *Stage I* objective function will naturally select workflows that collect high-accuracy information, while prioritizing the cost to acquire data and complexity to extract information from the data, less.

Exploring the *Stage I* Space

We opted for a grid-based search to highlight the use of a framework for a simple case study with a 2-dimensional *Stage I* search space. One of the limitations of the search method we used is that it does not generalize well to large parameter spaces. As an example, having 10 possible pieces of equipment each with n different possible settings will yield a *Stage I* search space of up to 10^n workflows; this space cannot be practically explored using an exhaustive grid-based search if n is even moderately large. We can make workflow selection more efficient in four ways. First, by being more selective with the candidate workflows in the final step of *Stage I*, one can choose workflows exactly at the minimum or workflows with scores having very small differences from the minimum, leading to a smaller set of candidate workflows to examine in *Stage II*. Second, using prior knowledge, the *Stage I* exploration space can be dramatically reduced to workflows that only make sense based on experience. Third, the weights of the *Stage I* objective function can be adjusted in such a way that it penalizes expensive workflows heavily and restricts the exploration space even further. Fourth, the *Stage I* search task can be completed using adaptive search methods such as active/ sequential learning, leading to vast improvements in *Stage I*'s execution speed [39–41].

Fig. 4 Examples of workflows that can each be considered optimal given different definitions of **Actionability**

Using the Framework to Select Workflows With Differing Structures

The presented case study demonstrates the use of the framework for a typical SEM characterization procedure, whose steps are well defined and understood. However, this framework can also be used to determine the optimal workflow for a scenario where numerous pieces of equipment, data processing pipelines, and types of models can be used in various combinations to arrive at the same kind of information, provided that the objective and ground-truth are still clearly defined by a human researcher. This framework is applicable for any objective where every possible sequence of steps in a workflow can be explicitly parametrized and defined.

To better illustrate this, we conceive of a scenario where an AE system is tasked to find the optimal workflow given a set of workflows with different structures and parametrizations. ARES is an AE system that synthesizes carbon nanotubes (CNTs) [24]. In this study, the authors chose (among other factors) the growth catalyst along with its film and support thickness, the dimensions for the silicon pillars that the catalyst pre-seeded, the number of silicon pillars per wafer for testing, and the method of determining the CNT growth rate. However, consider the possibility that some of the elements of the workflow can be changed, which would change the structure of the workflow used to measure the growth rate of CNTs. Each change that can be made to the experimental workflow requires the consideration of completely different sets of equipment and parameters. In general, this scenario can be posed as the following question: “if all assumptions of the framework were satisfied, and a given AE system could use all equipment/models/procedures in the lab, change all parameters accordingly, extract information from the potential workflows’ information streams, compare this information to the ground-truth, and evaluate workflows using a mathematically valid definition of **Actionability**, could the AE system find the optimal highthroughput workflow as defined by the *Stage I* and *Stage II* functions?” The answer is yes.

Using the Framework to Address Data Reproducibility Issues

Lastly, we envision that this framework will assist with the growing problem of data continuity and reproducibility within the scientific community. Historically, results from scientific studies could not be reproduced, either by the authors of the study or by other members in the community. This was partially due to the differences in the workflows that were employed: methods, capabilities, and tools are subject to change across research groups and across time. This complicates efforts to establish processing and property standards for materials systems, which is a significant impediment to expediting and optimizing materials development cycles. The presented framework creates an interface for different workflows to be compared and evaluated, which is a point of crucial importance as we enter a more mature era of the Integrated Computational Materials Engineering (ICME) initiative. Using this framework, different workflows can be compared on the *Stage I* objective space as long as they produce the same information for the same objective. Additionally, the framework's bias and uncertainty measures provide an easily interpretable set of metrics to judge the efficacy of workflows based on the information they produce.

Conclusion

It has been demonstrated in the literature that autonomous experimentation (AE) systems have incredible value to add to the community. For AE systems' potential to be fully realized, we recognize that their decision authority must be expanded in a practically implementable manner for experimenters. To achieve this, we designed a robust algorithmic framework for the selection of high-throughput workflows that can be completed by AE systems with minimal human intervention. We used the framework to select the optimal high-throughput workflow for material characterization on an AM Ti64 sample using a deep-learning based image denoiser in a case study. The collection time of BSE-SEM images was reduced by a factor of 5 and by a factor of 85 as compared to the Ground-Truth and Previous Study workflows, respectively. The bias and increase in standard deviation for the selected high-throughput workflow as compared to the Ground-Truth and Previous Study workflows were also reported.

Future work will involve utilizing the presented framework in further studies on metal AM components to develop databases that are critically important to understanding underlying process–structure–property relationships. Particularly, focus will be placed on

automating the image processing steps when performing materials characterization experiments to a greater extent.

Acknowledgements The authors would like to acknowledge support from ORNL's Manufacturing Demonstration Facility (MDF), the Australia–US Multidisciplinary University research Initiative (MURI) program under ONR award number N00014-18-1-2794 and project “Rationalization of Liquid/Solid and Solid/Solid Interphase Instabilities During Thermal-Mechanical Transients of Metal Additive Manufacturing”, and from the NSF Engineering Research Center for Hybrid Autonomous Manufacturing Moving from Evolution to Revolution (ERCHAMMER) under Grant GR128688 and Award AWD-113142.

Electron microscopy was performed at the Center for Electron Microscopy and Analysis (CEMAS) at The Ohio State University.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Stach E, DeCost B, Kusne AG, Hattrick-Simpers J, Brown KA, Reyes KG, Schrier J, Billinge S, Buonassisi T, Foster I et al (2021) Autonomous experimentation systems for materials development: a community perspective. *Matter* 4(9):2702–2726
2. Flores-Leonar MM, Mejía-Mendoza LM, Aguilar-Granda A, Sanchez-Lengeling B, Tribukait H, Amador-Bedolla C, AspuruGuzik A (2020) Materials acceleration platforms: on the way to autonomous experimentation. *Curr Opin Green Sustain Chem* 25:100370
3. Szymanski NJ, Zeng Y, Huo H, Bartel CJ, Kim H, Ceder G (2021) Toward autonomous design and synthesis of novel inorganic materials. *Mater Horiz* 8(8):2169–2198
4. Roch LM, Häse F, Kreisbeck C, Tamayo-Mendoza T, Yunker LPE, Hein JE, Aspuru-Guzik A (2018) Chemos: orchestrating autonomous experimentation. *Sci Robot* 3(19):eaat5559
5. Montoya JH, Aykol M, Anapolsky A, Gopal CB, Herring PK, Hummelshøj JS, Hung L, Kwon H-K, Schweigert D, Sun S et al (2022) Toward autonomous materials research: recent progress and future challenges. *Appl Phys Rev* 9(1):011405
6. Ling J, Hutchinson M, Antono E, Paradiso S, Meredig B (2017) High-dimensional materials and process optimization using datadriven experimental design with well-calibrated uncertainty estimates. *Integr Mater Manuf Innov* 6(3):207–217
7. Lookman T, Balachandran PV, Xue D, Hogden J, Theiler J (2017) Statistical inference and adaptive design for materials discovery. *Curr Opin Solid State Mater Sci* 21(3):121–128
8. Balachandran PV (2020) Adaptive machine learning for efficient materials design. *MRS Bull* 45(7):579–586
9. Xu H-Y, Shi T-N, Zhao H, Jin L-G, Wang F-C, Wang C-Y, Qi S-Y (2016) Heterogeneous Fenton-like discoloration of methyl orange using $\text{Fe}_3\text{O}_4/\text{MWCNTs}$ as catalyst: process optimization by response surface methodology. *Front Mater Sci* 10(1):45–55
10. Christensen M, Yunker LPE, Adedeji F, Häse F, Roch LM, Gensch T, dos Passos GG, Zepel T, Sigman MS, Aspuru-Guzik A et al (2021) Data-science driven autonomous process optimization. *Commun Chem* 4(1):1–12
11. Häse F, Matteo A, Hickman RJ, Roch LM, Christensen M, Liles E, Hein JE, Aspuru-Guzik A (2021) Olympus: a benchmarking

- framework for noisy optimization and experiment planning. *Mach Learn Sci Technol* 2(3):035021
12. Ong SP (2019) Accelerating materials science with high-throughput computations and machine learning. *Comput Mater Sci* 161:143–150
13. Calderon CE, Plata JJ, Toher C, Oses C, Levy O, Fornari M, Natan A, Mehl MJ, Hart G, Nardelli MB et al (2015) The AFLOW standard for high-throughput materials science calculations. *Comput Mater Sci* 108:233–238
14. Maier WF (2019) Early years of high-throughput experimentation and combinatorial approaches in catalysis and materials science. *ACS Comb Sci* 21(6):437–444
15. Vasudevan RK, Choudhary K, Mehta A, Smith R, Kusne G, Tavazza F, Vlcek L, Ziatdinov M, Kalinin SV, Hattrick-Simpers J (2019) Materials science in the artificial intelligence age: high-throughput library generation, machine learning, and a pathway from correlations to the underpinning physics. *MRS Commun* 9(3):821–838
16. Curtarolo S, Setyawan W, Wang S, Xue J, Yang K, Taylor RH, Nelson LJ, Hart GLW, Sanvito S, Buongiorno-Nardelli M et al (2012) AFLOWLIB.org: a distributed materials properties repository from high-throughput ab initio calculations. *Comput Mater Sci* 58:227–235
17. Hattrick-Simpers JR, Gregoire JM, Kusne AG (2016) Perspective: composition-structure-property mapping in high-throughput experiments: turning data into knowledge. *APL Mater* 4(5):053211
18. Liu YH, Hu ZH, Suo ZG, Hu LZ, Feng LY, Gong XQ, Liu Y, Zhang JC (2019) High-throughput experiments facilitate materials innovation: a review. *Sci China Technol Sci* 62(4):521–545
19. Pollice R, dos Passos Gomes G, Matteo A, Hickman RJ, Krenn M, Lavigne C, Lindner-D'Addario M, Nigam AK, Ser CT, Yao Z et al (2021) Data-driven strategies for accelerated materials design. *Acc Chem Res* 54(4):849–860
20. Choudhary K, DeCost B, Chen C, Jain A, Tavazza F, Cohn R, Park CW, Choudhary A, Agrawal A, Billinge SJL et al (2022) Recent advances and applications of deep learning methods in materials science. *npj Comput Mater* 8(1):1–26
21. Venugopal V, Sahoo S, Zaki M, Agarwal M, Gosvami NN, Krishnan NMA (2021) Looking through glass: knowledge discovery from materials science literature using natural language processing. *Patterns* 2(7):100290
22. Shetty P, Ramprasad R (2021) Automated knowledge extraction from polymer literature using natural language processing. *Iscience* 24(1):101922
23. Dieb S, Amano K, Tanabe K, Sato D, Ishii M, Tanifuji M (2021) Creating research topic map for NIMS SAMURAI database using natural language processing approach. *Sci Technol Adv Mater Methods* 1(1):2–11
24. Nikolaev P, Hooper D, Webber F, Rao R, Decker K, Krein M, Poleski J, Barto R, Maruyama B (2016) Autonomy in materials research: a case study in carbon nanotube growth. *npj Comput Mater* 2(1):1–6
25. Gongora AE, Xu B, Perry W, Okoye C, Riley P, Reyes KG, Morgan EF, Brown KA (2020) A Bayesian experimental autonomous researcher for mechanical design. *Sci Adv* 6(15):eaaz1708
26. Kusne AG, Yu H, Wu C, Zhang H, Hattrick-Simpers J, DeCost B, Sarker S, Oses C, Toher C, Curtarolo S, Davydov AV (2020) On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat Commun* 11(1):1–11
27. Kumar K, Pooleery A, Madhusoodanan K, Singh RN, Chakravartty JK, Dutta BK, Sinha RK (2014) Use of miniature tensile specimen for measurement of mechanical properties. *Procedia Eng* 86:899–909
28. SAE Taxonomy (2021) Definitions for terms related to driving automation systems for on-road motor vehicles. Publication J3016_202104. Society of Automotive Engineers
29. Shao M, Vijayan S, Nandwana P, Jinschek JR (2020) The effect of beam scan strategies on microstructural variations in Ti–6Al–4V fabricated by electron beam powder bed fusion. *Mater Des* 196:109165
30. Casukhela R, Vijayan S, Jacobsen M, Fourman J, Hepler K, Shao M, Gupta A, Jinschek J (2021) Towards the development of a multi-modal community-based am database. *Microsc Microanal* 27(S1):1090–1091
31. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 234–241
32. Mohan S, Manzorro R, Vincent JL, Tang B, Sheth DY, Simoncelli EP, Matteson DS, Crozier PA, Fernandez-Granda C (2020) Deep denoising for scientific discovery: a case study in electron microscopy. [arXiv: 2010. 12970](https://arxiv.org/abs/2010.12970)
33. de Haan K, Ballard ZS, Rivenson Y, Wu Y, Ozcan A (2019) Resolution enhancement in scanning electron microscopy using deep learning. *Sci Rep* 9(1):1–7
34. Roels J, Vermaillan F, Kremer A, Gonçalves A, Aelterman J, Luong HQ, Goossens B, Philips W, Lippens S, Saey Y (2020) An interactive ImageJ plugin for semi-automated image denoising in electron microscopy. *Nat Commun* 11(1):1–13
35. Zhang C, Han R, Zhang AR, Voyles PM (2020) Denoising atomic resolution 4d scanning transmission electron microscopy data with tensor singular value decomposition. *Ultramicroscopy* 219:113123
36. Zeltmann S, Minor A, Ophus C (2020) Denoising of sparse three- and four-dimensional hyperspectral electron microscopy data using a total variational method. *Microsc Microanal* 26(S2):1724–1726
37. Mohan S, Manzorro R, Vincent JL, Tang B, Sheth DY, Simoncelli E, Matteson DS, Crozier PA, Fernandez-Granda C (2022) Deep denoising for scientific discovery: a case study in electron microscopy. *IEEE Trans Comput Imaging* 8:585–597
38. Thermo Fisher Scientific (2019) Thermo fisher scientific xtra library for amira, avizo, and pergeos software. <https://xtras.amiravizo.com/xtras/bse-sem-denoiser>
39. Chaloner K, Verdinelli I (1995) Bayesian experimental design: a review. *Stat Sci* 10:273–304
40. Chernoff H (1959) Sequential design of experiments. *Ann Math Stat* 30(3):755–770
41. Cohn DA, Ghahramani Z, Jordan MI (1996) Active learning with statistical models. *J Artif Intell Res* 4:129–145