

Benchmarking autonomous scattering experiments illustrated on TAS

M. Teixeira Parente^{1,*}, A. Schneidewind¹, G. Brandl¹, C. Franz¹, M. Noack²,
M. Boehm³, and M. Ganeva¹

¹Jülich Centre for Neutron Science (JCNS) at Heinz Maier-Leibnitz Zentrum (MLZ), Forschungszentrum Jülich GmbH,
Garching, Germany

²Center for Advanced Mathematics for Energy Research Applications (CAMERA), Lawrence Berkeley National Laboratory,
Berkeley (CA), USA

³Institut Laue-Langevin, Grenoble, France

Abstract

With the advancement of artificial intelligence and machine learning methods, autonomous approaches are recognized to have great potential for performing more efficient scattering experiments. In our view, it is crucial for such approaches to provide thorough evidence about respective performance improvements in order to increase acceptance within a scientific community. Therefore, we propose a benchmarking procedure designed as a cost-benefit analysis that is applicable to any scattering method sequentially collecting data during an experiment. For a given approach, the performance assessment is based on how much benefit, given a certain cost budget, it is able to acquire in predefined test cases. Different approaches thus get a chance for comparison and can make their advantages explicit and visible. Key components of the procedure, i. e., cost measures, benefit measures, and test cases, are made precise for the setting of three-axes spectrometry (TAS) as an illustration. Finally, we discuss neglected aspects and possible extensions for the TAS setting and comment on the procedure’s applicability to other scattering methods. A Python implementation of the procedure to simplify its utilization by interested researchers from the field is also provided.

Keywords: autonomous experiment, data-driven analysis, cost-benefit analysis, performance, three-axes spectrometry, inelastic neutron scattering, materials analysis

1 Introduction

Scattering experiments have so far been carried out in a manual or semi-automated way, i. e., experimenters had to organize the measuring process and determine what and where to measure (next). With the rise of artificial intelligence and machine learning techniques, it is natural to ask whether there are autonomous approaches that allow performing experiments in a more efficient way. In other words, it is worthwhile to see if, for a fixed cost budget like experimental time available, autonomous approaches can perform “better” experiments. In the following, by “autonomous approach” and related phrases, we refer to a decision-making algorithm that is combined with an automated communication and analysis infrastructure to create a closed

*Corresponding author: Lichtenbergstraße 1, 85748 Garching, Germany, m.teixeira.parente@fz-juelich.de.

loop with an instrument control system and thus is, after initialization, able to conduct measurements without human intervention.

Indeed, there are already autonomous approaches that have recently been developed for scattering experiments [1, 2, 3, 4, 5, 6]. A way to compare and assess the performance of different approaches (manual as well as autonomous approaches) is, however, currently lacking. From our perspective, benchmarking and measuring performance is of the utmost importance at this stage for the establishment and progress of autonomous approaches in the field of materials analysis by scattering methods. As examples, similar efforts already developed a benchmark for oxygen evolution reaction catalyst discovery [7] or evaluated the performance of Bayesian optimization [8] across several materials science domains [9].

In this work, we propose a benchmarking procedure which is designed as a cost-benefit analysis and can be applied to any scattering method sequentially collecting data of a certain quantity of interest during an experiment. Key components of the procedure that mainly drive the performance assessment are cost measures, benefit measures, and test cases. Cost measures specify the type of cost that is to be minimized and benefit measures characterize how “success” is defined. Test cases describe particular scenarios depending on the scattering method and determine which aspects the approaches are tested on. We emphasize that benchmarking processes in the general field of machine learning are potentially fragile [10] and therefore need to be defined carefully. After the general formulation of the procedure, all of the mentioned components are made precise for the setting of three-axes spectrometry (TAS).

TAS is an established technique for materials analysis by inelastic neutron scattering [11]. For decades, three-axes spectrometers have measured the dynamic properties of solids, e. g., phonons and magnetic excitations, for a wide range in both energy transfer (E) and momentum (\mathbf{Q}) space and provide the opportunity to detect weak signals with high resolution. Compared to techniques like time-of-flight spectroscopy (TOF), profiting from large detector assemblies, the sequential, point-by-point, measurements in TAS experiments come with the cost of moving the three instrument axes—around a monochromator, a sample, and an analyzer—individually in real space, which is rather slow (seconds to minutes). Furthermore, instead of operating classically with a single detector, recent developments attempt to use multi-detectors for higher data acquisition speed at the expense of reduced measuring flexibility [12, 13, 14, 15]. However, as a first step, we concentrate on classical TAS experiments with a single detector here.

Semi-automated TAS experiments are often systematically organized on a grid in \mathbf{Q} - E space with predefined steps. This procedure is frequently inefficient as it collects a substantial amount of measurement points in the background, i. e., areas with erroneous signals most often coming from undesired scattering events inside the sample itself, the sample environment, or components of the instrument. If autonomous approaches, however, were able to collect information on intensities in signal regions and their shape faster, they could enable a more efficient (manual or autonomous) assessment of the investigated system’s behaviour.

The manuscript is divided into sections as follows. Sec. 2 gives fundamental definitions of the key components. The benchmarking procedure is specified in Sec. 3 and Sec. 4 contains an illustration of the key components in a TAS setting. Finally, we provide a discussion in Sec. 5 and a conclusion in Sec. 6.

2 Definitions

This section provides definitions for the central components of the benchmarking procedure (Sec. 3), i. e., for *test cases*, *experiments*, *cost measures*, and *benefit measures*. For TAS experiments, these notions are made precise in Sec. 4.

Definition (Test case). A *test case* t is a collection of details necessary to conduct a certain scattering experiment.

Here, a scattering experiment is viewed as a sequential collection of data points the meaning of which depends on the particular scattering method. Recall that an N -tuple, $N \in \mathbf{N}$, is a finite ordered list of N elements.

Definition (t -Experiment). Let t be a test case. A t -experiment $\mathcal{A} = \mathcal{A}_t$ is an N -tuple of data points

$$\mathcal{A} = (z_1, \dots, z_N), \quad (1)$$

where z_j denote data points and $|\mathcal{A}| := N \in \mathbf{N}$ is the number of data points.

In order to measure costs and benefits of experiments in the context of a certain test case, we need a formalization of cost and benefit measures.

Definition (Cost/Benefit measure). Let t be a test case. Both, a *cost measure* $c = c_t$ and a *benefit measure* $\mu = \mu_t$, are real-valued functions of t -experiments \mathcal{A} .

3 Benchmarking procedure

In this section, we use the notions from Sec. 2 to formulate the main outcome of this manuscript and suggest a step-by-step procedure for benchmarking scattering experiments. The result of a benchmark is a collection of sequences with benefit values (cf. Tab. 1) allowing to evaluate the performance of experiments in the context of predefined test cases.

Benchmarking procedure:

1. Specify $L \in \mathbf{N}$ test cases $t^{(\ell)}$, $\ell = 1, \dots, L$.
2. Specify a cost measure c and define $c^{(\ell)} := c_{t^{(\ell)}}$.
3. Specify a benefit measure μ and define $\mu^{(\ell)} := \mu_{t^{(\ell)}}$.

For each test case $t^{(\ell)}$, perform the following steps:

4. Specify $M^{(\ell)} \in \mathbf{N}$ ascending “milestone values”

$$0 < C_1^{(\ell)} < \dots < C_m^{(\ell)} < \dots < C_{M^{(\ell)}}^{(\ell)} < \infty. \quad (2)$$

5. Conduct a $t^{(\ell)}$ -experiment $\mathcal{A}^{(\ell)} := \mathcal{A}_{t^{(\ell)}} = (z_1^{(\ell)}, \dots, z_{|\mathcal{A}^{(\ell)}|}^{(\ell)})$ such that

$$c^{(\ell)}(\mathcal{A}^{(\ell)}) \geq C_{M^{(\ell)}}^{(\ell)}. \quad (3)$$

6. For $J \in \{1, \dots, |\mathcal{A}^{(\ell)}|\}$, let

$$\mathcal{A}_J^{(\ell)} := (z_1^{(\ell)}, \dots, z_J^{(\ell)}) \quad (4)$$

be the collection of the first J data points in $\mathcal{A}^{(\ell)}$. In particular, $|\mathcal{A}_J^{(\ell)}| = J$. Furthermore, let

$$J^{(\ell)}(C) := \max\{J \in \{1, \dots, |\mathcal{A}^{(\ell)}|\} \mid c^{(\ell)}(\mathcal{A}_J^{(\ell)}) \leq C\} \quad (5)$$

$t^{(1)}$	$\mu(\mathcal{A}_{J_1^{(1)}})$	\cdots	$\mu(\mathcal{A}_{J_{M^{(1)}}^{(1)}})$
\vdots	\vdots	\cdots	\cdots
$t^{(L)}$	$\mu(\mathcal{A}_{J_1^{(L)}})$	\cdots	$\mu(\mathcal{A}_{J_{M^{(L)}}^{(L)}})$

Table 1: Result of a benchmark. Each row, representing an experiment $\mathcal{A} = \mathcal{A}^{(\ell)}$ for a test case $t^{(\ell)}$, consists of benefit values (measured with $\mu = \mu^{(\ell)}$) that can be achieved using milestone values $C_m^{(\ell)}$ in column m as cost budgets.

be the maximum number of first data points in $\mathcal{A}^{(\ell)}$ possible with a cost budget $C \in \mathbf{R}$.

Then, for each milestone value $C_m^{(\ell)}$, compute benefit values using the first $J_m^{(\ell)} := J^{(\ell)}(C_m^{(\ell)})$ data points, i. e., compute

$$\mu^{(\ell)}(\mathcal{A}_{J_m^{(\ell)}}^{(\ell)}). \quad (6)$$

Summarizing, a benchmark requires the specification of test cases, a cost measure, a benefit measure, and a sequence of ascending milestone values for each test case. The result is a collection of sequences with benefit values (cf. Tab. 1) that can be achieved using the milestone values as cost budgets.

Using their results, different approaches can now be compared by, for example, plotting the function

$$m \mapsto \mu^{(\ell)}(\mathcal{A}_{J_m^{(\ell)}}^{(\ell)}) \quad (7)$$

for each test case $t^{(\ell)}$ and each approach. In Sec. 4.4, Fig. 1 provides a demonstration for a particular test case from the TAS context. Also, in the context of the specified cost and benefit measure, the results could allow to make strong claims of the following kind:

“Approach A lead to better experiments in $p\%$ of all test cases compared to approaches B, C , etc.”

4 Illustration on TAS

In this section, the general formulation of the benchmarking procedure and its components is made precise for TAS experiments.

4.1 TAS setting

In TAS experiments, the concept of a three-dimensional *reciprocal space* \mathbf{Q} is essential. For a given *sample* (or *material*), an element of \mathbf{Q} is denoted by $\mathbf{q} = (h, k, l)^\top$. An element of the *energy transfer space* E is denoted by ω . Furthermore, we call $(\mathbf{q}, \omega)^\top \in \mathbf{R}^r$, $r = 4$, the *\mathbf{Q} - E variables*. For details, we refer to [11].

Since TAS experiments are carried out only along $n \leq r$, $n \in \mathbf{N}$, predefined directions in \mathbf{Q} - E space, we introduce *experiment variables* $\mathbf{x} = (x_1, \dots, x_n)^\top \in \mathbf{R}^n$ and a corresponding affine transformation $T : \mathbf{R}^n \rightarrow \mathbf{R}^r$ to \mathbf{Q} - E variables, i. e.,

$$\begin{pmatrix} \mathbf{q} \\ \omega \end{pmatrix} = W\mathbf{x} + \mathbf{b} =: T(\mathbf{x}) \quad (8)$$

for a full rank matrix $W \in \mathbf{R}^{r \times n}$ and an *offset* $\mathbf{b} \in \mathbf{R}^r$. It follows that

$$\mathbf{x} = (W^\top W)^{-1} W^\top \left[\begin{pmatrix} \mathbf{q} \\ \omega \end{pmatrix} - \mathbf{b} \right] =: T^{-1}(\mathbf{q}, \omega) \quad (9)$$

for $(\mathbf{q}, \omega)^\top \in \mathbf{R}^r$. Note that $T^{-1}(T(\mathbf{x})) = \mathbf{x}$ for each $\mathbf{x} \in \mathbf{R}^n$, but $T(T^{-1}(\mathbf{q}, \omega)) = (\mathbf{q}, \omega)^\top$ only for $(\mathbf{q}, \omega)^\top \in T(\mathbf{R}^n)$.

Each experiment variable x_k only ranges between respective *limits of investigation* $x_k^\pm \in \mathbf{R}$, i. e., we have that

$$x_k^- \leq x_k \leq x_k^+. \quad (10)$$

The corresponding *domain of interest* $\mathcal{X} \subseteq \mathbf{R}^n$ is defined as

$$\mathcal{X} := [x_1^-, x_1^+] \times \cdots \times [x_n^-, x_n^+]. \quad (11)$$

The representation of a TAS instrument is divided into two structures, an *instrument configuration* and an *instrument*.

Definition (Instrument configuration). The tuple

$$\Pi = (\text{sm}, \text{sc}, \text{se}) \quad (12)$$

for a scan mode $\text{sm} \in \{\text{“constant } k_i\text{”}, \text{“constant } k_f\text{”}\}$, a scan constant (k_i or k_f) $\text{sc} > 0$, and a vector of scattering senses $\text{se} = (\text{se}_{\text{mono}}, \text{se}_{\text{sample}}, \text{se}_{\text{ana}})^\top \in \{-1, +1\}^3$ is called an *(TAS) instrument configuration*.

To measure intensities at a certain point in \mathbf{Q} - E space, a TAS instrument needs to steer its axes to six related angles. However, it is enough to regard only a subset of four angles due to dependencies.

Definition (Instrument). The tuple

$$\text{Ins} = (\mathbf{v}, d_{\text{mono}}, d_{\text{ana}}) \quad (13)$$

for a vector of angular velocities $\mathbf{v} = (v_1, \dots, v_4)^\top \in [0, \infty)^4$, and lattice parameters of the monochromator, $d_{\text{mono}} > 0$, and the analyzer, $d_{\text{ana}} > 0$, is called an *(TAS) instrument*.

The lattice parameters of the monochromator and the analyzer are necessary for a well-defined translation of points in \mathbf{Q} - E space to associated angles of instrument axes. The so-called *angle map*

$$\Psi : \text{dom}(\Psi) \rightarrow [0, \pi)^4, (\mathbf{q}, \omega) \mapsto (\Psi_1(\mathbf{q}, \omega), \dots, \Psi_4(\mathbf{q}, \omega))^\top \quad (14)$$

is induced by a sample, its *orientation*, and an instrument Ins and its configuration Π . Note that angular velocities v_k are related to angles $\Psi_k(\mathbf{q}, \omega)$. The domain $\text{dom}(\Psi) \subseteq \mathbf{Q} \times E$ denotes the set of points in \mathbf{Q} - E space for which Ψ is well-defined, i. e., points that are reachable by the instrument Ins . Furthermore, we define

$$\mathcal{X}^* := T^{-1}(\text{dom}(\Psi) \cap T(\mathcal{X})) \quad (15)$$

as the set of points in \mathcal{X} such that

$$\Psi(\mathbf{x}) := \Psi(T(\mathbf{x})) \quad (16)$$

is well-defined for $\mathbf{x} \in \mathcal{X}^*$. The limits of investigation for the domain of interest \mathcal{X} have to be set such that $\Psi : \mathcal{X}^* \rightarrow [0, \pi)^4$ becomes an *injective* function.

For a given configuration Π and instrument Ins , we can specify a *resolution function* [11]

$$\varphi_{\Pi, \text{Ins}} : T(\mathcal{X}^*) \rightarrow (T(\mathcal{X}^*) \rightarrow [0, \infty)), \quad (17)$$

i. e., each point in $T(\mathcal{X}^*) \subseteq \mathbf{R}^r$ gives an individual resolution function defined over $T(\mathcal{X}^*)$. Of course, the exact resolution function depends on additional parameters such as distances between instrument components or beam divergence allowed by collimators. However, these parameters are fixed during an experiment and hence can be omitted here. If Π and Ins are known from the context, we write $\varphi = \varphi_{\Pi, \text{Ins}}$. Also, for $\mathbf{x} \in \mathcal{X}^*$, we define

$$\varphi_{\Pi, \text{Ins}}(\mathbf{x}) := \varphi_{\Pi, \text{Ins}}(T(\mathbf{x})). \quad (18)$$

A sample and its orientation induce a so-called *scattering function* $s : T(\mathcal{X}) \rightarrow [0, \infty)$. Again, for $\mathbf{x} \in \mathcal{X}$, we define

$$s(\mathbf{x}) := s(T(\mathbf{x})). \quad (19)$$

Given a resolution function φ , we get an associated *intensity function* $i : \mathcal{X}^* \rightarrow [0, \infty)$ defined by a convolution of s with φ , i. e.,

$$i(\mathbf{x}) := (s * \varphi)(\mathbf{x}). \quad (20)$$

For benchmarking, we assume that i can be exactly evaluated. Of course, this is not possible in real experiments due to background and statistical noise.

For our suggestion of a benefit measure (see Sec. 4.3), we additionally need an *intensity threshold* $\tau > 0$.

The *experimental time* is defined as the sum of the cumulative counting time at measurement points in \mathbf{Q} - E space and the cumulative time for moving the instrument axes. For simplicity, we assume here that the *single counting time*, i. e., the counting time at a single measurement point, denoted by $T_{\text{count}} \geq 0$, is constant for each point.

Now, we can specify the notion of a *TAS test case* and a *TAS experiment*.

Definition (TAS test case). A tuple

$$t = (\text{sample}, \text{orientation}, W, \mathbf{b}, \mathcal{X}, \Pi, \text{Ins}, \tau, T_{\text{count}}) \quad (21)$$

for a sample and its orientation, an affine transformation induced by a full rank matrix $W \in \mathbf{R}^{r \times n}$ and an offset $\mathbf{b} \in \mathbf{R}^r$ (see Eq. (8)), a domain of interest $\mathcal{X} \subseteq \mathbf{R}^n$ (see Eq. (11)), an instrument configuration Π , an instrument Ins , an intensity threshold $\tau > 0$, and a single counting time $T_{\text{count}} \geq 0$ is called a *TAS test case*.

Note that a TAS test case t induces an intensity function $i : \mathcal{X}^* \rightarrow [0, \infty)$.

In the context of a certain test case, a TAS experiment is defined as a collection of intensities $i(\mathbf{x})$ at locations $\mathbf{x} \in \mathcal{X}^*$.

Definition (TAS t -Experiment). Let t be a TAS test case. A *TAS t -experiment* $\mathcal{A} = \mathcal{A}_t$ is an N -tuple of location-intensity pairs

$$\mathcal{A} = ((\mathbf{x}_1, i_1), \dots, (\mathbf{x}_N, i_N)), \quad (22)$$

where $\mathbf{x}_j \in \mathcal{X}^*$ denote measurement locations and $i_j = i(\mathbf{x}_j) \geq 0$ are corresponding values of the intensity function i induced by t .

4.2 Cost measure

We need to align our proposition of a cost measure with the limited experimental time, which is the critical quantity in a TAS experiment. Recall that the experimental time is defined as the sum of the cumulative counting time and the cumulative time for axes movement (see Sec. 4.1).

For the cumulative counting time, we define

$$c_{\text{count}}(\mathcal{A}) := \sum_{j=1}^{|\mathcal{A}|} T_{\text{count}} = |\mathcal{A}| \cdot T_{\text{count}}, \quad (23)$$

where $T_{\text{count}} \geq 0$ denotes the constant single counting time.

The cost measure representing the cumulative time for moving the instrument axes is defined as

$$c_{\text{axes}}(\mathcal{A}) := \sum_{j=1}^{|\mathcal{A}|-1} d(\mathbf{x}_j, \mathbf{x}_{j+1}) \quad (24)$$

for a metric $d : \mathcal{X}^* \times \mathcal{X}^* \rightarrow [0, \infty)$. The metric $d = d_t$ measures the maximum time of axes movement between corresponding angles, i. e.,

$$d(\mathbf{x}, \mathbf{x}') = \max_{1 \leq k \leq 4} \left| \frac{\Psi_k(\mathbf{x}) - \Psi_k(\mathbf{x}')}{v_k} \right|, \quad (25)$$

where $\Psi = (\Psi_1, \dots, \Psi_4)^\top$ denotes the angle map from Eq. (14) and $\mathbf{v} = (v_1, \dots, v_4)^\top$ is the vector of the instrument's angular velocities (see Eq. (13)). Note that the metric d is fully determined by the TAS test case t . Also, it is indeed a metric [16] since the angle map Ψ was chosen to be injective.

The measures c_{count} and c_{axes} can be used either individually or additively to form a cost measure representing the entire experimental time. For the latter, we finally define

$$c(\mathcal{A}) := c_{\text{count}}(\mathcal{A}) + c_{\text{axes}}(\mathcal{A}). \quad (26)$$

In our opinion, this cost measure is the most suitable to reflect time costs in a general TAS experiment.

4.3 Benefit measure

We propose a benefit measure that measures a type of weighted L^2 approximation error between a benchmark intensity function $i = i_t$ and an approximation $\hat{i} = \hat{i}(\mathcal{A})$ resulting from an experiment $\mathcal{A} = \mathcal{A}_t$. For example, \hat{i} can be constructed with location-intensity pairs from \mathcal{A} by linear interpolation or other approximation methods. To compare benefit values across experiments relating to different test cases, benefit measures should be “normalized”, i. e., we regard *relative* errors.

Let us define

$$\mu(\mathcal{A}) := \frac{\|i - \hat{i}(\mathcal{A})\|}{\|i\|}, \quad (27)$$

where $\|\cdot\|$ is a norm that enables to control the error measurement. Note that the notion of a “benefit” refers to an “error” in this case, i. e., *reducing* the benefit measure $\mu = \mu_t$ leads to an *increase* in benefit.

A suitable error norm $\|\cdot\|$ needs to reflect that a TAS experimenter is more interested in regions of signal than in the background. This suggests that we use i itself in a suitable definition. However, an important constraint is that signal regions with different intensities are weighted equally since they might be equally

interesting. For this, we use the intensity threshold $\tau > 0$ from the TAS test case t (see Eq. (21)) and define

$$i_\tau(\mathbf{x}) := \min\{i(\mathbf{x}), \tau\} \geq 0 \quad (28)$$

for $\mathbf{x} \in \mathcal{X}^*$, i. e., we cut i to a maximum intensity value of τ . As i_τ is a nonnegative function, its normalization

$$\rho_{i,\tau}(\mathbf{x}) := \frac{i_\tau(\mathbf{x})}{\int_{\mathcal{X}^*} i_\tau(\mathbf{x}') d\mathbf{x}'} \quad (29)$$

is a probability density function and can be used for weighting. Finally, we set

$$\|\cdot\| = \|\cdot\|_{i,\tau} := \|\cdot\|_{L^2(\mathcal{X}^*, \rho_{i,\tau})}, \quad (30)$$

where

$$\|h\|_{L^2(\mathcal{X}^*, \rho)} = \left(\int_{\mathcal{X}^*} h(\mathbf{x})^2 \rho(\mathbf{x}) d\mathbf{x} \right)^{1/2} \quad (31)$$

for a function h and a density function ρ . Note that $\|\cdot\|_{L^2(\mathcal{X}^*, \rho)}$ can be approximated by numerical quadrature rules or estimated by a Monte Carlo approach.

4.4 Test cases

A useful set of test cases represents the variety of different scenarios that can occur in a TAS experiment. It is the set of intensity functions that is particularly important here. Therefore, we suggest to create test cases such that the corresponding induced intensity functions are composed of one or more of the following structures known in the field:

- non-dispersive structures (e. g., crystal field excitations),
- dispersive structures (e. g., spin waves or acoustic phonons),
- (pseudo-)continua (e. g., spinons or excitations in frustrated magnets).

Particular intensity functions can be created by mathematical expressions, physical simulations, or experimental data.

As an example for an outcome of our benchmarking procedure, Fig. 1 displays a comparative benchmark result of two approaches for a test case including an intensity function that reflects a soft mode of the ferroelectric phase transition at ~ 60 K of a transverse optical phonon measured on SnTe [17]. This figure contains exemplifications of each benchmark component: experimental time (x -axis in Fig. 1c)) as cost measure (Sec. 4.2), approximation error (y -axis in Fig. 1c)) as benefit measure (Sec. 4.3), a phonon-type intensity function (Fig. 1a,b)) as part of the test case (Sec. 4.4), and milestone values (ticks on x -axis in Fig. 1c)).

5 Discussion

In this section, we discuss some open aspects of the TAS setting from Sec. 4 and the applicability of the benchmarking procedure from Sec. 3 to other scattering methods. Finally, we refer to a repository containing a software implementation.

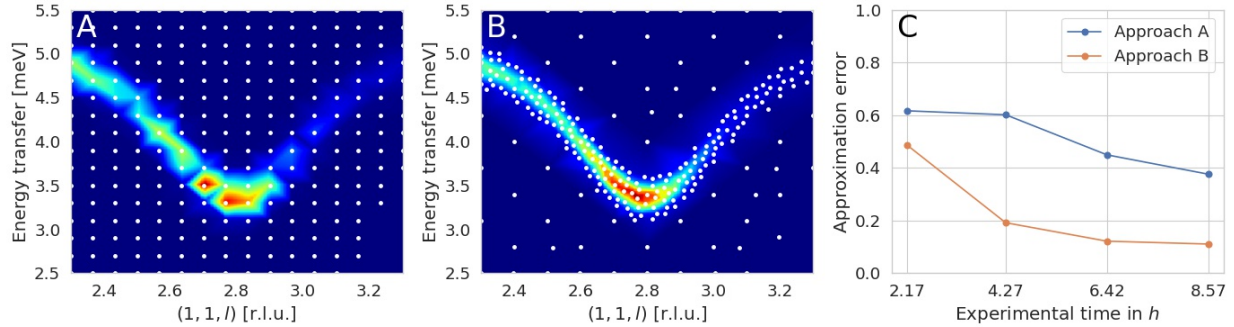


Figure 1: Example for an outcome of our benchmarking procedure. Two approaches (in a) and b)) perform an experiment in the context of a test case including an intensity function of a phonon defined on $\mathcal{X} = [2.3, 3.3] \times [2.5, 5.5]$. In c), the benchmark result for this test case shows that approach b) is able to reduce the approximation error from Sec. 4.3 quicker than approach a) since it puts the majority of measurement points in the region of signal.

5.1 Neglected aspects and possible extensions for TAS

The TAS setting does not comprise each detail occurring in a real experiment. Indeed, we neglected some aspects which we, however, see as acceptable deviations.

Certainly the most prominent neglected aspect is background and statistical noise. We find it difficult to include them in the benchmarking procedure due to their non-deterministic nature since a comparison of approaches needs to be done in a deterministic setting. The benefit measure from Sec. 4.3, for instance, makes use of a “true” benchmark intensity function which would not be available in the presence of background or statistical noise.

We assumed that the single counting time is constant for each measurement point. However, since single counting times are determined by the physics of the sample and the requested statistics in real experiments, they might vary at different measurement locations. In our opinion, this variation is negligible in a first step, but can be taken into account in more complex benchmarking setups if desired.

Next, the cost measures from Sec. 4.2 are rather general and can be applied to any TAS instrument. They can, however, be arbitrarily extended by more details if benchmarking is to be done in the context of a specific instrument. For example, real TAS instruments may differ in more (PANDA [18]) or less (ThALES [19]) time-consuming procedures for moving k_i values.

Furthermore, the intensity functions induced by the test cases described in Sec. 4.4 are derived in the context of *fixed* environmental parameters of the sample (such as temperature, external magnetic or electric field, etc.). Future work might extend the four-dimensional \mathbf{Q} - E space with these parameters, i.e., $r > 4$, to allow for more complex intensity functions that would, however, require adjusted cost measures.

Finally, to compute all benefit values for Tab. 1, the benchmarking procedure requires an approach to perform an experiment that is large enough (see Eq. (3)). Hence, an autonomous stopping criterion is not tested although we consider it a crucial part of a fully autonomous approach.

5.2 Applicability to other scattering methods

The benchmarking procedure is formulated in a modular way and depends only on abstract components like cost measures, benefit measures, and test cases. For TAS experiments, we specified these components in

Sec. 4, but feel that the overall procedure is also applicable for scattering methods other than TAS such as diffraction, reflectivity, SAS/GISAS, or TOF. Indeed, experimenters for each of these methods measure costs and benefits in their own way and investigate different kinds of intensity functions.

Diffraction experiments [20], for example, may have a setting similar to TAS experiments as the experimenter is interested in intensities defined over \mathbf{Q} variables and many diffractometers need to move their components (sample, detector). Therefore, the cost and benefit measures presented above might also be useful in this context. The set of test cases, however, would need to be composed differently since regions of signal are mainly separated and small in shape. Also, additional aspects (such as shadowing, overlapping reflections, missing knowledge of symmetry, etc.) are to be taken into account.

5.3 Data repository

Since the benchmarking procedure has algorithmic structure, we decided to provide an implementation in the form of Python code that computes sequences of benefit values for given cost and benefit measures (cf. Tab. 1). Also, benchmark components for the TAS setting are already implemented. The repository along with instructions on how to run the code is publicly available [21]. It also contains descriptions of test cases that can be complemented in the future.

6 Conclusion

In this manuscript, we have developed a benchmarking procedure for scattering experiments which is designed as a cost-benefit analysis and based on key components like cost measures, benefit measures, and test cases.

Although we have provided first suggestions for all these components in a TAS setting, the process of finding a suitable benchmark setting for the scattering community in general as well as the TAS community in particular is certainly not finished.

As an outlook for the TAS community, a useful next step could be the inclusion of non-constant single counting times since it has the potential of further savings of experimental time that we do not account for in the current setting. Also, extending the \mathbf{Q} - E variables with environmental parameters that were assumed to be fixed would lead to a more comprehensive setting. Finally, we see the contribution to the dynamical set of test cases as another future task for the community.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

MG and MT contributed to the conception and design of the study. All authors took over the implementation of the study. MT wrote the first draft of the manuscript. CF, MG, MN, AS, and MT wrote paragraphs of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version. GB and MT wrote source code for implementing the benchmarking procedure.

Funding

MG and MT received support through the project *Artificial Intelligence for Neutron and X-Ray Scattering* (AINX) funded by the Helmholtz AI unit of the German Helmholtz Association. MN is funded through the Center for Advanced Mathematics for Energy Research Applications (CAMERA), which is jointly funded by the Advanced Scientific Computing Research (ASCR) and Basic Energy Sciences (BES) within the Department of Energy’s Office of Science, under Contract No. DE-AC02-05CH11231.

Acknowledgments

We thank Frank Weber and Rolf Heid (both Institute for Quantum Materials and Technologies, Karlsruhe Institute of Technology) for providing scattering data for Fig. 1. Also, we acknowledge discussions with Thomas Kluge (Helmholtz-Zentrum Dresden-Rossendorf).

Supplemental Data

None.

Data Availability Statement

Code for the implementation of the benchmarking procedure and files for TAS test cases are open source and publicly available at <https://jugit.fz-juelich.de/ainx/base>.

References

- [1] J. H. Durant, L. Wilkins, K. Butler, and J. F. K. Cooper, “Determining the maximum information gain and optimizing experimental design in neutron reflectometry using the Fisher information,” *Journal of Applied Crystallography*, vol. 54, no. 4, pp. 1100–1110, 2021.
- [2] J. H. Durant, L. Wilkins, and J. F. Cooper, “Optimising experimental design in neutron reflectometry,” *arXiv preprint arXiv:2108.05605*, 2021.
- [3] P. M. Maffettone, L. Banko, P. Cui, Y. Lysogorskiy, M. A. Little, D. Olds, A. Ludwig, and A. I. Cooper, “Crystallography companion agent for high-throughput materials discovery,” *Nature Computational Science*, vol. 1, no. 4, pp. 290–297, 2021.
- [4] M. M. Noack, G. S. Doerk, R. Li, J. K. Streit, R. A. Vaia, K. G. Yager, and M. Fukuto, “Autonomous materials discovery driven by Gaussian process regression with inhomogeneous measurement noise and anisotropic kernels,” *Scientific Reports*, vol. 10, no. 1, p. 17663, 2020.
- [5] M. M. Noack, P. H. Zwart, D. M. Ushizima, M. Fukuto, K. G. Yager, K. C. Elbert, C. B. Murray, A. Stein, G. S. Doerk, E. H. Tsai, *et al.*, “Gaussian processes for autonomous data acquisition at large-scale synchrotron and neutron facilities,” *Nature Reviews Physics*, vol. 3, no. 10, pp. 685–697, 2021.

- [6] M. Teixeira Parente, G. Brandl, C. Franz, A. Schneidewind, and M. Ganeva, “Autonomous experiments for neutron three-axis spectrometers (TAS) with log-Gaussian processes,” *arXiv preprint arXiv:2105.07716*, 2021.
- [7] B. Rohr, H. S. Stein, D. Guevarra, Y. Wang, J. A. Haber, M. Aykol, S. K. Suram, and J. M. Gregoire, “Benchmarking the acceleration of materials discovery by sequential learning,” *Chemical Science*, vol. 11, no. 10, pp. 2696–2706, 2020.
- [8] P. I. Frazier and J. Wang, “Bayesian optimization for materials design,” in *Information science for materials discovery and design*, pp. 45–75, Springer, 2016.
- [9] Q. Liang, A. E. Gongora, Z. Ren, A. Tiihonen, Z. Liu, S. Sun, J. R. Deneault, D. Bash, F. Mekki-Berrada, S. A. Khan, *et al.*, “Benchmarking the performance of Bayesian optimization across multiple experimental materials science domains,” *npj Computational Materials*, vol. 7, no. 1, p. 188, 2021.
- [10] M. Dehghani, Y. Tay, A. A. Gritsenko, Z. Zhao, N. Houlsby, F. Diaz, D. Metzler, and O. Vinyals, “The benchmark lottery,” *arXiv preprint arXiv:2107.07002*, 2021.
- [11] G. Shirane, S. M. Shapiro, and J. M. Tranquada, *Neutron scattering with a triple-axis spectrometer: basic techniques*. Cambridge University Press, 2002.
- [12] F. Groitl, D. Graf, J. O. Birk, M. Markó, M. Bartkowiak, U. Filges, C. Niedermayer, C. Rüegg, and H. M. Rønnow, “CAMEA—A novel multiplexing analyzer for neutron spectroscopy,” *Review of Scientific Instruments*, vol. 87, no. 3, p. 035109, 2016.
- [13] F. Groitl, R. Toft-Petersen, D. L. Quintero-Castro, S. Meng, Z. Lu, Z. Huesges, M. D. Le, S. Alimov, T. Wilpert, K. Kiefer, *et al.*, “MultiFLEXX - The new multi-analyzer at the cold triple-axis spectrometer FLEXX,” *Scientific Reports*, vol. 7, no. 1, pp. 1–12, 2017.
- [14] M. Kempa, B. Janousova, J. Saroun, P. Flores, M. Boehm, F. Demmel, and J. Kulda, “The FlatCone multianalyzer setup for ILL’s three-axis spectrometers,” *Physica B: Condensed Matter*, vol. 385, pp. 1080–1082, 2006.
- [15] J. Lim, K. Siemensmeyer, P. Čermák, B. Lake, A. Schneidewind, and D. Inosov, “BAMBUS: a new inelastic multiplexed neutron spectrometer for PANDA,” *Journal of Physics: Conference Series*, vol. 592, no. 1, p. 012145, 2015.
- [16] “Metric,” *Encyclopedia of Mathematics*, 1999. <https://encyclopediaofmath.org/wiki/Metric> (Accessed: 2021-06-30).
- [17] F. Weber and R. Heid, “Private communication,” 2021.
- [18] A. Schneidewind and P. Čermák, “PANDA: Cold three axes spectrometer,” *Journal of large-scale research facilities JLSRF*, vol. 1, p. A12, 2015.
- [19] M. Boehm, P. Steffens, J. Kulda, M. Klicpera, S. Roux, P. Courtois, P. Svoboda, J. Saroun, and V. Sechovsky, “ThALES – Three Axis Low Energy Spectroscopy for highly correlated electron systems,” *Neutron News*, vol. 26, no. 3, pp. 18–21, 2015.
- [20] D. S. Sivia, *Elementary scattering theory: for X-ray and neutron users*. Oxford University Press, 2011.

- [21] M. Teixeira Parente and G. Brandl, “BASE: Benchmarking Autonomous Scattering Experiments.” <https://jugit.fz-juelich.de/ainx/base>, 2021.