



Cite this: DOI: 10.1039/d5dd00518c

# Cost-aware Bayesian optimization of real-world nanoindentation workflows for accelerated mechanical characterization

Vivek Chawla,<sup>a</sup> Stephen Pupilampu,<sup>a</sup> Haochen Zhu,<sup>b</sup> Philip D. Rack,<sup>b</sup> Dayakar Penumadu<sup>a</sup> and Sergei Kalinin<sup>b</sup>

Accelerating the discovery of mechanical properties in combinatorial materials requires autonomous experimentation that accounts for both instrument behavior and experimental cost. Here, an automated nanoindentation (AE-NI) framework is developed and validated for adaptive mechanical mapping of combinatorial thin-film libraries. The method integrates heteroskedastic Gaussian-process modeling with cost-aware Bayesian optimization to dynamically select indentation locations and hold times, minimizing total testing time while preserving measurement accuracy. A detailed emulator and cost model capture the intrinsic penalties associated with lateral motion, drift stabilization, and reconfiguration-factors often neglected in conventional active-learning approaches. To prevent kernel-length-scale collapse caused by disparate time scales, a hierarchical meta-testing workflow combining local grid and global exploration is introduced. Implementation of the workflow is shown on an experimental Ta–Ti–Hf–Zr thin-film library. The proposed framework achieves nearly a thirty-fold improvement in property-mapping efficiency relative to grid-based indentation, demonstrating that incorporating cost and drift models into probabilistic planning substantially improves performance. This study establishes a generalizable strategy for optimizing experimental workflows in autonomous materials characterization and can be extended to other high-precision, drift-limited instruments.

Received 21st November 2025  
Accepted 28th May 2026

DOI: 10.1039/d5dd00518c

rsc.li/digitaldiscovery

## 1 Introduction

Materials discovery is a cornerstone of technological advancement. As the demand for performance, functionality, and sustainability grows, so too does the need for new processing techniques and compositions that satisfy the growing needs. Over the last twenty years, considerable effort has been directed towards theory-based materials prediction,<sup>1,2</sup> with companies like Google,<sup>3</sup> Meta,<sup>4</sup> and Microsoft<sup>5</sup> joining the effort recently. However, the advancements in theory belied the need for experimental realization of theoretical predictions and associated need to accelerate experimental discovery cycles. This led to a wave of interest in the automated synthesis of functional materials *via* microfluidics<sup>6</sup> and pipetting robotics,<sup>7–10</sup> and rapid adaptation of additive manufacturing to the combinatorial discovery.

It is important to note that this wave of interest in high throughput synthesis is but the most recent one. A notable earlier wave emerged in the late 1990s and early 2000s, when advances in thin-film deposition methods, particularly pulsed

laser deposition (PLD),<sup>11,12</sup> enabled the creation of large-area composition spreads. The primary lesson of these combinatorial winters was that one of the primary challenges associated with material discovery is not only the generation of candidate materials but also their effective characterization. To keep up with the pace of synthesis, characterization methods must also be altered for speed and precision.

In the realm of mechanical characterization, nanoindentation has emerged as a particularly powerful tool for evaluating combinatorial samples. Originally used for hardness testing and reduced modulus<sup>13–15</sup> evaluation, nanoindentation has evolved into a comprehensive method capable of probing elastic–plastic behavior,<sup>16–19</sup> creep,<sup>20–24</sup> fracture toughness,<sup>25–29</sup> strain rate effects,<sup>30–32</sup> and viscoelastic behavior.<sup>33,34</sup> More recent developments have enabled the extraction of full plasticity parameters, including yield strength and ultimate tensile strength.<sup>35</sup>

Here, we develop a cost aware framework for active learning in nanoindentation applied to rapid mechanical property discovery in combinatorial libraries using cost-aware Bayesian optimization (BO) with the (structured) Gaussian process surrogate models. Here, the ML-enabled nanoindentation or automated nanoindentation (AE NI) system, dynamically plans the experiment, deciding the location of subsequent indentation based on a combination of mean predictions, uncertainty

<sup>a</sup>Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996, USA. E-mail: vchawla@vols.utk.edu

<sup>b</sup>Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN 37996, USA



and time required for these measurements. A unique aspect of nanoindentation is that it presents intricate measurement costs associated with the manufacturer-specified engineering controls in the system. Here, we calibrate these measurement costs and derive the universal cost function structure for BO. By coupling the Gaussian Process (GP) model with different noise priors (classical and heteroskedastic) with a custom acquisition function that accounts for mean, uncertainty, and the cost, a cost aware decision-making algorithm is defined. We report the model simulated workflows and experimental implementation on the model KLA nanoindenter systems, overall enabling guidelines for discovery.

This study advances the concept of automated, adaptive mechanical characterization by operationalizing a cost-aware Bayesian optimization framework for nanoindentation. The proposed workflow unifies probabilistic modeling, emulator-based cost calibration, and experiment planning within a single decision-making loop that accounts for both measurement uncertainty and instrument-imposed overheads such as reconfiguration, positioning, and drift stabilization. To address the kernel-length-scale collapse arising from measurements across different temporal regimes, a hierarchical meta-testing strategy is introduced that combines local grid exploration with global optimization. The resulting AE NI system dynamically balances exploration and efficiency, reducing total experimental time by more than an order of magnitude compared to conventional grid mapping, while maintaining accuracy in reconstructed property fields. Beyond a proof of concept, the framework generalizes to diverse material systems and provides a practical foundation for intelligent automation in high-throughput mechanical characterization.

## 2 Emulator for thin films

Fig. 1 illustrates the envisioned operation of the automated experimental nanoindentation (AE-NI) framework. In this approach, a surrogate regressor continuously learns from previously measured indentation data to predict the evolving property landscape and its associated uncertainty. Based on these predictions, the controller identifies the next most informative measurement point and commands the nanoindenter to execute it, forming a closed loop between computation and the experiment. This adaptive cycle enables efficient exploration of thin-film libraries by prioritizing measurements that maximize information gain while minimizing time and repositioning costs.

To illustrate the workflows for the automated structural property discovery *via* ML-enabled nanoindentation and given that Bayesian optimization to our knowledge has not been used for automated nanoindentation before, we first illustrate the emulator for workflow. In addition to illustration purposes, such emulators can significantly facilitate the initial hyperparameter tuning for the ML algorithms, required given the high cost of experimental measurements compared to the theoretical workflows.

Here, the simulated combinatorial library, representative of a ternary alloy system, is created by diagonally varying the

composition of elements A, B and C (SI Fig. S1) onto a two-dimensional spatial domain thereby emulating a realistic deposition pattern often seen in combinatorial synthesis. The ground-truth response was generated using a synthetic model that incorporates multiple compositional effects, including baseline mixture behavior, pairwise interactions, and higher-order concentration-dependent terms. The resulting hardness was the sum of the three mechanisms as shown in eqn (1). More details on the choice of ground truth can be found in the SI. The resulting hardness variation can be found in the SI (Fig. S1).

$$H = \sum_{i=A}^C c_i H_i + \sum_{i,j=A}^C \sigma_{ij} c_i c_j + k \times c_C^2 \times (1 - c_C) \quad (1)$$

where  $H$  is the overall hardness.  $H_i$  are baseline elemental contributions,  $c_i$  is the concentration of element  $i$ ,  $\sigma_{ij}$  are the interaction coefficients, and  $k$  modulates the nonlinear concentration dependent term.

To enable automated experiments, we use Bayesian Optimization (BO) based on a surrogate Gaussian process based model. BO is a probabilistic framework designed to efficiently optimize or explore functions that are expensive to evaluate, often referred to as black box functions. In the context of automated experiments, BO has emerged as a powerful strategy to reduce the overall cost of discovery or optimization of the black box function. In the current emulator, the black-box function corresponds to the unknown spatial variation of hardness across the  $X$ - $Y$  space. BO helps navigate the trade-off between exploration (sampling the regions of high uncertainties) and exploitation (sampling regions that based on current analysis are most likely to yield the optimal values) to achieve the experimental objective. It is important to note that a key aspect of BO compared to other decision-making frameworks (reinforcement learning, dynamic programming, *etc.*)<sup>36,37</sup> is that the decisions are myopic, ideally matching the exploration of combinatorial libraries where the results of measurements in a certain location do not depend on previous measurements elsewhere.

To achieve this objective, Bayesian optimization requires a surrogate model that relates the input feature space ( $X$ ,  $Y$ ) to the target property (hardness). In this study, a Gaussian Process (GP) as a surrogate model is used. GPs are probabilistic models that assume the value at any given location is statistically correlated with values at nearby locations. Mathematically, a Gaussian Process (GP) is defined in eqn (2). Specifically, nearby points have stronger influence, while the influence of distant points diminishes with increasing distance. The spatial correlation is captured through a covariance matrix, whose entries are determined by a kernel function. For example, the Radial Basis Function (RBF) kernel (eqn (3)), scales this correlation based on a kernel length ( $l$ ). A smaller kernel length implies rapid function variation over short distance or noise effects. Similarly, a large length scale indicates slow variation over large distances.

$$f(x) \sim \text{GP}(\mu(x), k(x, x')) \quad (2)$$



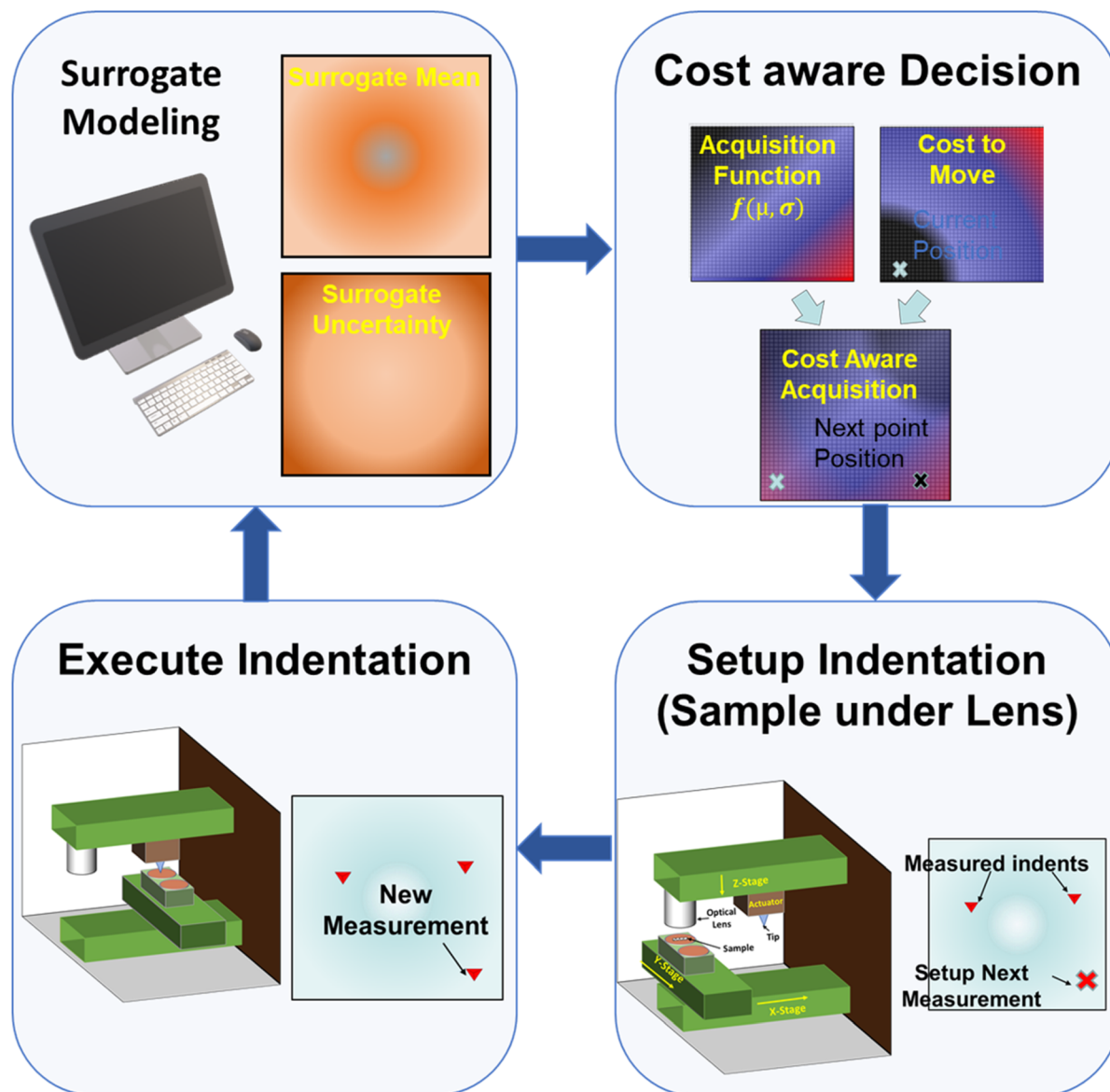


Fig. 1 Schematic of the automated experimental nanoindentation (AE-NI) framework. A surrogate model continuously updates its mean and uncertainty from measured indents, guiding the selection of the next measurement based on uncertainty and cost. The controller converts model decisions into stage movements, closing the loop between data acquisition, model update, and physical indentation for adaptive mechanical mapping.

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2l^2}\right) \quad (3)$$

During GP training, both the mean and kernel parameters are optimized to best represent the underlying black-box function. The key advantage of GPs is that they provide not only a prediction at each point but also an estimate of the uncertainty associated with that prediction.

This enables strategic decision-making in Bayesian optimization. Depending upon the objective, whether to explore uncertain regions or exploit the regions with high predicted values, the next sampling point can be chosen. This BO

decision-making process is governed by the acquisition function. The acquisition function balances the exploration and exploitation and guides the selection of the next query point. Some of the most commonly used acquisition functions include Upper Confidence Bound (UCB), Expected Improvement (EI), and Uncertainty Estimation (UE). UCB favors points where either the predicted mean is high or the uncertainty is large. EI quantifies the expected amount by which the new point will improve over the current best which inherently balances the mean and variance by giving more value to points that are likely to improve performance. The UE acquisition strategy simply chooses the points of highest uncertainty. Mathematically the UCB, EI and UE are defined in eqn (4)–(6) respectively. In the present study, the objective is explicitly defined as learning the



underlying function. Accordingly, unless otherwise specified, uncertainty exploration (UE) is used as the acquisition function for the baseline implementation.

$$\text{UCB}(x) = \mu(x) + \beta^{1/2} \cdot \sigma(x), \beta \text{ controls the trade off} \quad (4)$$

$$\text{EI}(x) = E[\max(0, f(x) - f^+)], f^+ \text{ is best observed so far} \quad (5)$$

$$\text{UE}(x) = \sigma(x) \quad (6)$$

From the perspective of this study, there are two possible choices for the feature space used in the surrogate model: one can use either the spatial coordinates ( $X, Y$ ) or the compositional variables (fractions of A, B and C) as input features. Although the composition spans three components, it is constrained by  $\chi_A + \chi_B + \chi_C = 1$ , reducing the degrees of freedom to two. Therefore, the compositional space can be fully parameterized by the two spatial coordinates ( $X, Y$ ). This remains valid even for higher-dimensional systems ( $n \geq 4$ ), as the library is physically realized on a two-dimensional substrate and composition can vary only along two spatial degrees of freedom (hence the full phase diagram cannot be explored in 2D thin-film libraries for  $n \geq 4$ ). Therefore, at any given time, the compositional space is sampled along two spatial directions on the substrate. Hence, the  $X$ - $Y$  coordinates are used as input features.

Note that BO can be based on more complex surrogate models, as long as they allow interpolation over combinatorial space and either allow for prediction and uncertainty, or allow for multiple probabilistic function prediction (Thompson sampling). Therefore, the workflow proposed here can be elementarily updated to include such alternative models.

Another important aspect of BO is that typically in theoretical implementations the costs of measurements everywhere in the parameter space are assumed to be constant. Practically, however, the cost structure (*e.g.*, time of measurement, computational cost, *etc.*) can be readily incorporated into the acquisition function. In nanoindentation measurements, the measurement cost can be very complex and is determined by the engineering controls established by the manufacturers, and is discussed below.

### 3 Experimental implementation

Implementation of ML-driven active learning in combinatorial nanoindentation requires careful consideration of actual experimental workflows. In this work, the analysis is specific to a KLA nanoindenter. For different instruments from the same manufacturer, the procedure can be reproduced through calibration, while for instruments from other manufacturers the overall workflow of automated CSM indentation remains the same but individual details may differ. For instance, grid positioning and setup can vary between systems, as can testing details such as the method of surface detection and the timing and duration of drift evaluation. Nevertheless, the methodology is analogous. All results here are presented for the continuous stiffness measurement (CSM) technique, which is the standard approach for measuring hardness and modulus,<sup>14</sup> although the same ideas

can be extended to other testing modes such as fracture toughness or displacement-controlled experiments. Inevitably, the individual details of implementation will change. Key considerations are the cost of stage movements, drift effects, and manufacturer-imposed constraints, which together form the basis for benchmarking and uncertainty quantification.

#### 3.1 Automated nanoindentation workflow

A KLA iMicro Nanoindenter system is used in this current study. In our previous work,<sup>38</sup> an automated nanoindentation framework was developed that enables the direct control of the indenter through a Python environment. The automated system allows users to specify indentation locations based on  $X$ - $Y$  coordinates, the indenter optical image, or an external optical image. For this work, the focus is exclusively on  $X$ - $Y$  coordinate-based indentation, as the emulator exhibits continuous variation of composition across the physical  $X$ - $Y$  space with no distinct microstructural features.

The nanoindentation workflow follows a standardized sequence. First, samples are mounted on a sample holder and loaded within the nanoindenter. The holder then moves from the 'safe-position' to a location beneath the optical lens *via*  $X$ - $Y$  movement, while the  $Z$ -stage is subsequently adjusted to bring the sample into focus. This step is called the initialization step and it remains the same across both manual and automated nanoindentation workflows. Following the initialization, indentation grids (series of indents that follow a consistent spacing in  $x$  and  $y$ ) are defined. The definition of grids consists of moving to the position of the grid underneath the optical lens, setting up the grid (setup step) and repeating the procedure until all the grids are defined. Subsequently, the indentation procedure is carried out. The indentation procedure follows a structured sequence: the  $Z$ -stage rises (frame with the indenter and optical lens moving away from the sample) and the  $X$ - $Y$  stage shifts to the location of the first indent in the first grid under the indenter (lift step), and the  $Z$ -stage then lowers to detect the surface contact (engage step). Subsequently, the system waits for stabilization of the drift (drift settlement step) before proceeding to test each indent within the grid (measurement step). Henceforth, the combined sequence of  $Z$ -stage upward movement,  $X$ - $Y$  stage movement, surface engagement, and drift stabilization is referred to as reconfiguration. During automated grid indentation, lateral movements between sites may encounter local surface height variations that can pose a risk of tip-sample contact. To mitigate this, the indenter tip is retracted by a fixed clearance (20  $\mu\text{m}$ ) prior to any lateral movement, ensuring sufficient separation between the tip and the sample surface. After finishing a grid, the reconfiguration happens as the sample moves between consecutive grids. Once the final grid is completed, the  $Z$ -stage rises and the sample moves from underneath the indenter back to the optical lens.

#### 3.2 Modeling experimental cost

To implement active learning on a nanoindenter, it is first essential to quantify the cost associated with each action the nanoindenter performs. Here, we build a detailed cost model



that captures the real-world time penalties tied to each step in the indentation workflow.

The time associated with the lift step ( $t_{\text{step}}$ ), engage step ( $t_{\text{engage}}$ ), and comeback step ( $t_{\text{comeback}}$ ) was recorded to be approximately 30 s, 65 s, and 35 s, respectively. In addition to the above, there are costs that are variable during indentation. These costs include drift stabilization time ( $t_{\text{drift}}$ ), measurement time ( $t_{\text{measurement}}$ ) and movement time ( $t_{\text{move}}$ ). The average drift is set to 300 s ( $t_{\text{drift}} = 300$  s), although in practice it can vary and can be bounded. The  $t_{\text{measurement}}$  typically depends upon the type of test as well as the target depth.  $t_{\text{measurement}} = 120$  s is assumed based on the target indentation depth of 200 nm and continuous stiffness measurement (CSM) based indentation with a strain rate of 0.2. Lastly, the  $t_{\text{move}}$  is a function of Euclidian distance between indentation locations. Therefore,  $t_{\text{move}}$  is modeled by recording the time required to perform stage movements ranging between 2 and 50 000  $\mu\text{m}$ .

The movement time is then fit using a piecewise function consisting of an exponential model for short range movements and a linear model for long range movements. This functional choice is based on the behavior of stepper motor-controlled

stages, which often use PID control systems that behave differently depending on movement scale. Fig. 2a shows the cost of movement from one location to another. Fig. 2b shows the function during indentation. As can be seen there is a jump of  $\sim 400$  seconds in time required to move from one location to another if the distance between the two locations is greater than 1000  $\mu\text{m}$ . This is an important manufacturer-imposed safety constraint to prevent the indenter tip from accidentally striking the surface during extended movements. Therefore, to move such large distances, a reconfiguration penalty is applied *i.e.*, the system first raises the Z-stage ( $t_{\text{rise}}$ ), and then the sample is moved laterally to the new location ( $t_{\text{move}}$ ) followed by lowering the Z-stage ( $t_{\text{engage}}$ ) and drift stabilization ( $t_{\text{drift}}$ ). To demonstrate the impact of this constraint, two indentation grids are modeled (Fig. 2c). In the first grid, three indents are placed with the first two spaced by 100  $\mu\text{m}$ , and the third indent is located 990  $\mu\text{m}$  away from the second indent. In the second grid, the same positions are used except the third indent is placed 1000  $\mu\text{m}$  from the second indent—just crossing the reconfiguration threshold. Fig. 2d shows the resulting total time per configuration showcasing the  $\sim 400$  s discrepancy in the second case.

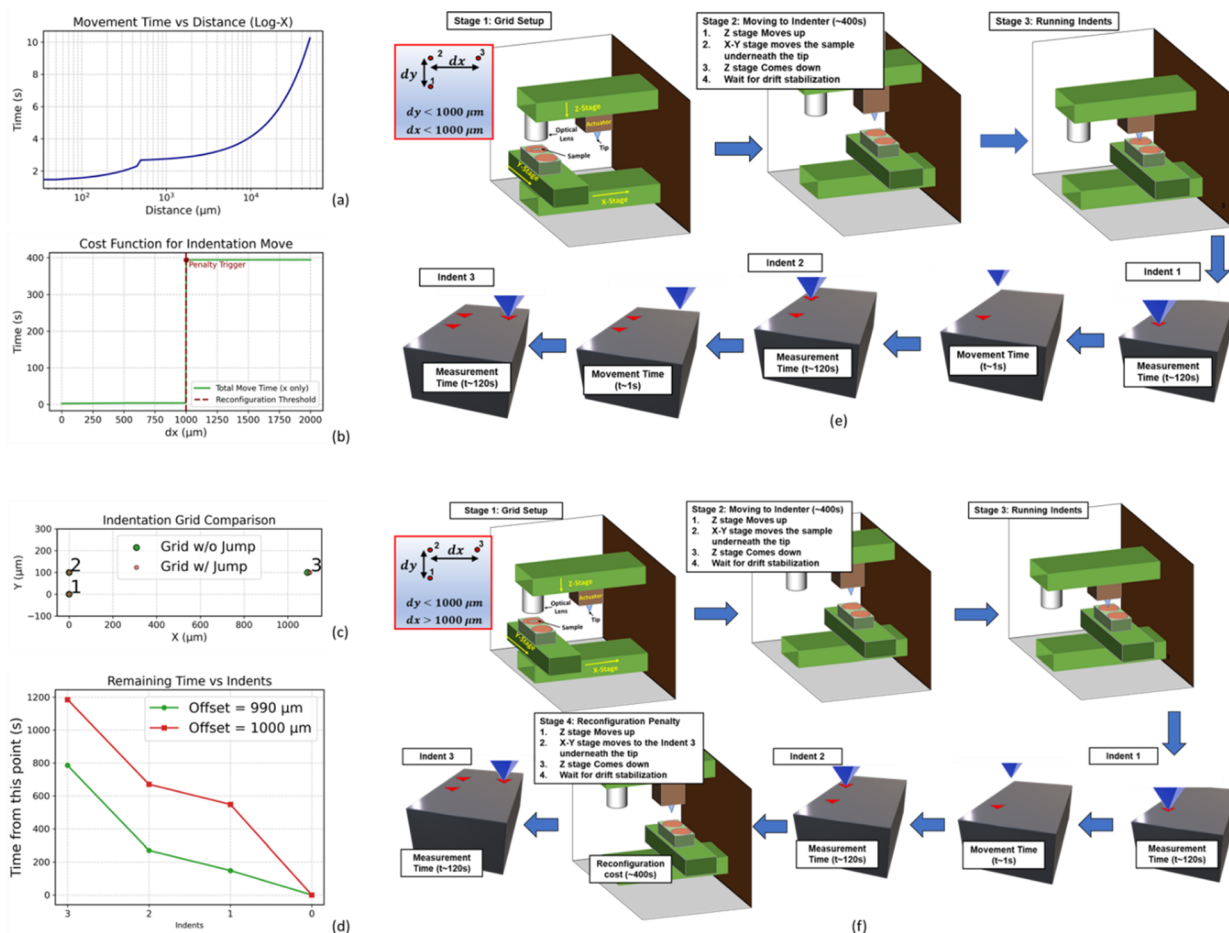


Fig. 2 Movement-based cost model for adaptive indentation. (a and b) Movement time and corresponding cost function showing a 400 s reconfiguration penalty for travel distances exceeding 1000  $\mu\text{m}$  during indentation. (c–f) Example grids and execution sequences demonstrating how the penalty alters total measurement time: when displacement exceeds the threshold, full Z-stage repositioning and drift stabilization are triggered before resuming indentation in (f) (travel distance > 1000  $\mu\text{m}$ ) while only the indenter moves in (e) (travel distance < 1000  $\mu\text{m}$ ).



Fig. 2e and f illustrate the full sequence of steps during indentation for the two cases. In the no-reconfiguration case (Fig. 2e), once the grid setup is completed, the z-stage rises and the sample moves underneath the indenter, and waits for the initial drift stabilization to be complete. All subsequent indents are then executed sequentially within the same region. As previously mentioned, the indenter tip retracts by 20  $\mu\text{m}$  and the x-y stage moves between points, preventing tip-sample collision, leading to minimal overhead and rapid measurement cycles. In contrast, when the displacement exceeds the 1000  $\mu\text{m}$  threshold (Fig. 2f), the system performs a complete reconfiguration sequence before the third indentation to account for potential height variations over larger travel distances. This additional sequence introduces a fixed time penalty ( $\sim 400$  s) per reconfiguration event, significantly increasing the total experiment duration. Together, these steps emphasize that the cost function captures not only lateral motion time but also the real operational overhead imposed by the instrument's safety and drift-control protocols.

### 3.3 Building baselines

**3.3.1 Grid search.** With the cost model in place, the next step is to establish a baseline experimental cost to allow for benchmarking of BO based active learning *vs.* a classical approach. In conventional nanoindentation workflows, the domain is often explored using a grid search approach. To mimic this common practice, a  $9 \times 9$  grid search over the emulator domain is performed, where each grid contains  $5 \times 5$

indentations, resulting in a total of 2025 indents. Using the previously defined cost model, this exhaustive search corresponds to a total testing time of approximately 2,863,900 s ( $\sim 80$  hours). Fig. 3a shows the grid search pattern on the true variation of hardness, while Fig. 3b and c display the ground truth hardness map and the reconstructed estimate, respectively. Note, the acquisition function in this case becomes relatively uniform at later stages of sampling, reflecting convergence of the GP under noisy observations. As a result, the remaining variations in uncertainty are small and can be influenced by boundary effects inherent to kernel-based models. Consequently, regions near previously sampled locations still exhibit comparable or slightly elevated uncertainty values, not due to lack of information, but due to the combined effects of noise and edge-related penalty (eqn S1 in the SI) Although the variation in mean absolute percentage error (MAPE) converges only marginally faster for the regression-based adaptive approach compared to the grid search (Fig. 3d), prior studies have demonstrated that active learning methods can improve convergence rates by a factor of two to three under more complex conditions. The relatively modest gain observed here likely stems from the simplicity of the emulator's ground truth surface. Nevertheless, despite the extensive number of indents and significant time investment, the grid search fails to identify the true maximum hardness located near the bottom-left region of the domain (Fig. 3e). While targeted edge indentations could, in principle, reveal this feature, such decisions require prior knowledge of the ground truth. Consequently, any sharp

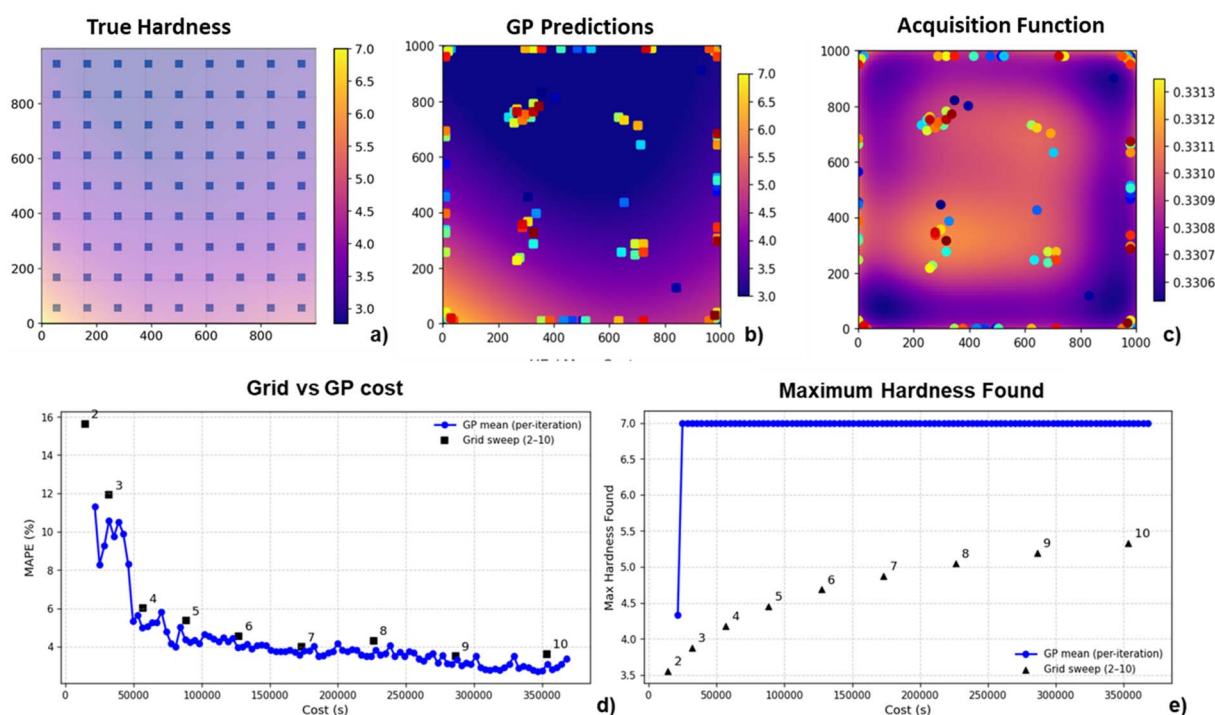


Fig. 3 Comparison of grid search and GP exploration. (a) True hardness map with a  $9 \times 9$  grid sampling pattern. (b and c) Adaptive regressor predictions showing measured points (recent in red; early in blue) and the corresponding uncertainty. (d and e) Mean absolute percentage error and maximum hardness found *versus* total cost, illustrating faster discovery and convergence of the adaptive method compared to grid-based search.



gradient or localized phase transition would remain undetected within a uniform grid scheme. Adaptive exploration strategies, by contrast, incorporate feedback from measured data, enabling dynamic refinement of sampling density in regions of high variability. This capability transforms data collection from a passive mapping exercise into an information-driven process, thereby reducing experimental cost while improving the likelihood of capturing critical microstructural or compositional transitions.

**3.3.2 Measurement cost models.** With the cost model established, we analyze the measurement bottlenecks, *i.e.*, the stages where most of the experimental time is spent during testing. For a given grid, total time ( $t_{\text{total}} = t_{\text{move}} + t_{\text{engage}} + t_{\text{drift}} + t_{\text{measurement}} + t_{\text{lift}}$ ). An obvious next step is to evaluate the percentage of time spent on actual indentation measurements. This percentage would depend upon testing parameters as well as grid size once the assumption is made on  $t_{\text{drift}}$  and  $t_{\text{measurement}}$ . The distribution of this cost is shown in Fig. 4a and b (for cases of grid spacing being below and above 1000  $\mu\text{m}$ ). For the single indent grids,  $\sim 60\%$  of the time is spent in drift calibration, making the test highly inefficient. In contrast, for larger grids, the majority of the time is spent on measurements. Of course, the total time per grid is also higher for larger grids. If the grid spacing is above 1000  $\mu\text{m}$  (Fig. 4b), most of the time is spent on reconfiguration cost. As the most time is spent on measurements for higher grid size, it is useful to break down the measurement duration itself. Following the Oliver-Pharr method and default Continuous Stiffness Measurement (CSM) settings,  $\sim 30$  s is spent on detecting the

surface,  $\sim 40$  s for measurement with a  $\sim 2$  s hold (crucial for elastic unloading for viscoelastic materials), and  $\sim 80$  s is allocated to post-hold drift monitoring, which is then used to correct the displacement data. Thus, nearly 50% of the measurement time is used just to evaluate drift. This is notable because a significant time is spent allowing drift to settle before testing begins ( $\sim 300$  s). Given this, there is a strong case for adaptive drift monitoring strategies, especially in exploratory workflows where the goal is to understand property variations. Since drift is often caused by external thermal or mechanical fluctuations, adapting its measurement dynamically can potentially reduce time without significantly affecting data quality as drift is often caused by an external event.

**3.3.3 Lateral drift.** An important factor in any experimental measurement is drift, *i.e.*, deviation between the intended measurement location and the actual one. In ML-enabled experimentation, the drift role is twofold. The negative aspect is the loss of positional precision, affecting the surrogate models that are traditionally much more sensitive to the uncertainties in parameter space than in the measured function. The positive is that accelerated discovery compared to the grid-based operations reduces the deleterious effects of drift. We note that in principle drift can be accounted for post-acquisition (simply by mapping the locations of indentation sites) or dynamically *via* the use of optical fiducials, but defer these studies to the future. Here, we explore the effects of drift on model performance.

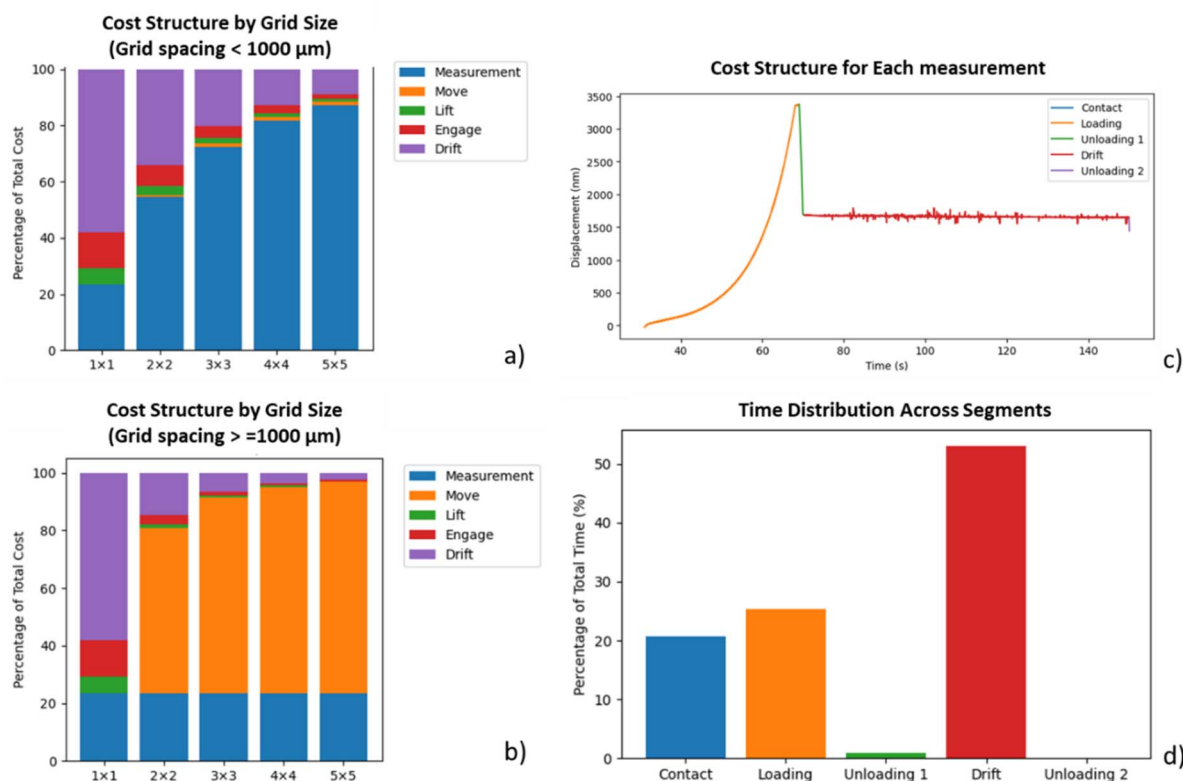


Fig. 4 Cost structure and measurement characteristics. (a) Cost distribution for grid spacing  $<1000$   $\mu\text{m}$ . (b) Cost distribution for grid spacing  $>1000$   $\mu\text{m}$ . (c) Measurement time distribution for fused silica indents with maximum target load. (d) Percentage time spent during measurement.



**3.3.4 Adaptive drift sampling.** Vertical drift in nano-indentation experiments is often caused by thermal relaxation, instrument instability and sample movement. Drift does not result in noisy experiments but causes a bias in the measured depth as given in eqn (7) (where DR is the measured drift rate) which in turn results in the bias in the area function and subsequently, the hardness (eqn (8)–(10)). Assuming the realistic values of a DR of  $0.1 \text{ nm s}^{-1}$  and loading segment of 30 s for an indent depth of 3000 nm, results in an error of less than 0.1 percent in hardness. Therefore, a long drift hold ( $\sim 80 \text{ s}$ ) typically used in traditional workflows is not necessary. While effective in eliminating bias, this hold is time consuming. This approach assumes drift is always significant, which may not hold in practice, especially in short or localized experiments.

$$h_{\text{measured}} = h_{\text{indent}} + t_{\text{measurement}} \times \text{DR} \quad (7)$$

Assuming,

$$h_{\text{measured}} = h_{\text{contact}}, \text{ i.e., } \varepsilon = 0 \quad (8)$$

$$H = \frac{P}{A_c} \quad (9)$$

where  $h_{\text{measured}}$ ,  $h_{\text{indent}}$ , and  $h_{\text{contact}}$  are the measured depth, the total indentation depth, and the contact depth.  $\varepsilon$  is the Oliver-Pharr<sup>14</sup> epsilon that accounts for pile-up and sink in effects.  $H$  is the hardness,  $P$  is the load, and  $A_c$  is the contact area.

Therefore, a more adaptive strategy is adopted that balances drift correction with experimental efficiency. We use an

adaptive holding scheme with two hold durations: 5 seconds and 80 seconds. The short 5 s hold is used by default, while the 80 s hold is activated when the estimated drift exceeds a defined threshold and remains until the drift stabilizes. Two drift scenarios are analyzed. In the first, the drift-induced error is small (0.3) and quickly recovers. In the second, the drift error is large ( $\sim 1$ ) and decays slowly, as shown in Fig. 5a. Fig. 5c–e show the GP predictions, uncertainty, and deviation from ground truth for the bad drift case. In Fig. 5c, similar to Fig. 3c, the acquisition function eventually saturates with very little spatial variation due to measurement noise. As a result, the remaining local variations are primarily influenced by noise, edge penalty and drift-induced effects rather than true lack of sampling. In the favorable drift case, the adaptive hold strategy preserves model accuracy while reducing total measurement time by up to a factor of about 2 compared to a fixed 80 s hold. In the high-drift case, the model automatically increases the hold duration, yielding smaller time savings but maintaining reliable convergence. The exact improvement factor depends on drift magnitude and parameters, but overall, the results demonstrate that adaptive holds can yield significant time gains while retaining predictive robustness (Fig. 5b). This also provides a practical criterion to define the drift cutoff based on the acceptable measurement error rather than an arbitrary threshold. Since drift-induced bias scales with  $\text{DR} \cdot t$  (eqn (7)), the resulting relative error in hardness can be estimated as  $\sim \left(\text{DR} \cdot \frac{t}{h}\right)^2$ , allowing the cutoff to be selected directly from

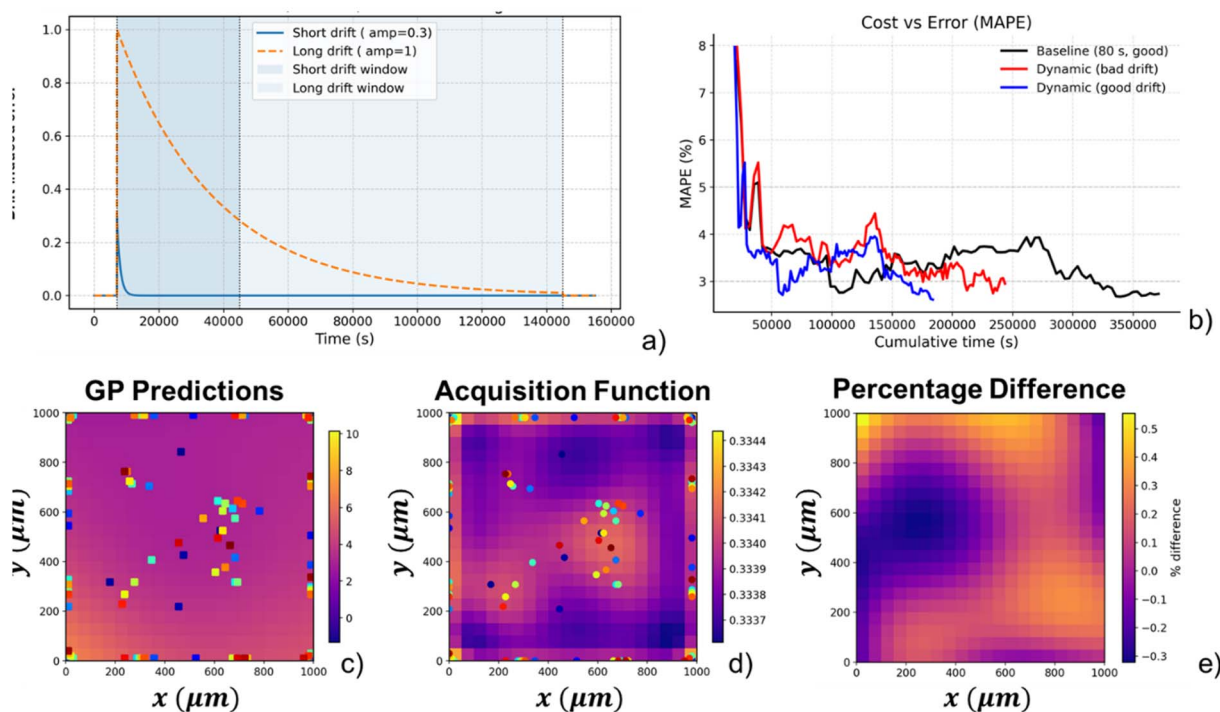


Fig. 5 Effect of drift-aware adaptive hold selection on mapping accuracy. (a) Short- and long-drift models showing exponential decay of drift rate over time. (b) Mean absolute percentage error (MAPE) versus cumulative time for baseline (fixed 80 s hold) and dynamic hold strategies under good and poor drift conditions. (c–e) Corresponding regressor predictions, uncertainty, and percentage difference maps, demonstrating that drift-adaptive scheduling maintains accuracy while reducing total experiment time.



a target accuracy. If the drift exceeds this threshold, the adaptive scheme automatically transitions to longer hold durations, ensuring that the accumulated bias remains within acceptable limits while maintaining experimental efficiency.

## 4 Building search strategies

In principle, the analysis of the cost function enables construction of a purely Bayesian optimization (BO) workflow, where the Gaussian Process (GP) surrogate suggests the next measurement location (based on the policy) under cost constraints. However, in practice such an approach often demonstrates instabilities, *e.g.*, the collapse of kernel length scales. This happens due to the noise in the surrogate function coupled with large time differences between repeated local measurements *vs.* measurements that require large stage movements. In principle, this can be corrected by the non-Gaussian (*e.g.*, student *t*-distribution) based noise models, but that can lead to introduction of a lot of additional hyperparameters.

While these are guaranteed to self-correct over long measurement times, practical considerations of AE NI can benefit from more robust schemes. One such approach is a human in the loop;<sup>39,40</sup> and another is introduction of more robust policies. Here, we do it by analyzing grid measurement *v.* local measurement grid.

### 4.1 Adaptive grid sizing

Often in indentation testing, an  $n_x \times n_y$  grid is used to get measurement at a ‘single’ location. These indents act as multiple tests at a single location to understand the noise associated with the measurement. Therefore, the numbers of indents are often decided by a probabilistic model, like the student-*t* distribution, such that the confidence in the measurement is high. Often,  $n_x = n_y = 5$  are set, resulting in a cost of  $\sim 3500$  s per grid. However, the grid size is pre-determined and does not adapt to the underlying variation within the sample.

In nanoindentation of combinatorial samples, spatially varying noise is not uncommon. By definition, combinatorial libraries exhibit compositional gradients, and each composition can lead not only to different target properties (*e.g.*, hardness) but also to different noise levels (arising from processing steps, layer adhesion, and sample curvature). In such cases, a heteroskedastic Gaussian Process (GP) is more appropriate as it not only adapts to the underlying trend but also learns a spatially varying noise function, treating noise as a variable rather than a fixed quantity. In heteroskedastic GP, two GPs are trained simultaneously: one modeling the mean of the data and the other models the input-dependent noise. When the noise is truly Gaussian and randomly distributed, the heteroskedastic GP naturally reduces to the behavior of a standard (homoscedastic) GP, effectively recovering the noiseless case.

To realize this, first noise is added to the nanoindentation measurements. Fig. 6a–c shows the ground truth of the emulator, the noise added to the measurements, and the

measurements after the added noise. For the purpose of this work, the maximum noise was kept low compared to the maximum measured hardness ( $\sim 1.4$  percent). Given the measurements, we actively train the heteroskedastic GP and compute the expected noise at any candidate point. Given a target standard error of the mean (target SEM), denoted by  $\epsilon$  and given by the user, the required number of indents can be determined at that location using eqn (10). To keep the grid size practical and avoid extremes, clipping and rounding is applied to the resulting grid size using eqn (11). This means the system will adaptively choose between  $2 \times 2$  to  $5 \times 5$  grid size, depending upon local noise levels.

$$n_{\text{req}}(x, y) = \left( \frac{\sigma_{\text{ext}}(x, y)}{\epsilon} \right)^2 \quad (10)$$

$$n_{\text{grid}} = \min(\max(\sqrt{n_{\text{req}}}, 2), 5) \quad (11)$$

where  $n_{\text{req}}$  is the minimum number of grid required,  $\sigma_{\text{ext}}(x, y)$  is the estimated noise at location  $(x, y)$ ,  $\epsilon$  is the target SEM, and  $n_{\text{grid}}$  is the chosen grid size. Fig. 6e and f shows the predicted mean, and the expected noise. In this case, even though the model was not exactly able to recreate the noise model (perhaps because the noise itself was kept low), the target SEM criterion was sufficient to guide intelligent grid resizing. The evolution of grid size with iteration is shown in Fig. 6d. Furthermore, the adaptive approach achieves approximately a fourfold reduction in total time compared to the fixed-grid case, while reaching a comparable error level (see the blue and black curves in Fig. 6j).

### 4.2 Meta grid

As further optimization of the automated indentation workflow is carried out, there is a clear bottleneck: reconfiguration cost. As shown in Fig. 4a, particularly for smaller grids (which become dominant in low noise regions), a significant portion of the total time is spent transitioning between grids rather than performing indentations. To address this bottleneck, a strategy is introduced that groups multiple local grids, thereby forming a meta grid, within a single global ‘safe zone’. For instance, considering a case where, based on the noise estimate from heteroskedastic GP, a  $2 \times 2$  grid is required at a global location of  $(g_{x1}, g_{y1})$ , a set of neighboring  $2 \times 2$  grids centered at  $(g_{xi}, g_{yi})$ , where  $i = 1:N$ , are predefined. Collectively, these  $2 \times 2 \times (N + 1)$  indents form what is subsequently referred to as the meta grid, with each  $2 \times 2$  set defined as the sub-grid. The next step is to determine the locations  $(g_{xi}, g_{yi})$  for the sub-grids. These can be chosen using various strategies, for example, selecting locations of highest uncertainty, combining uncertainty with distance-based weighting or using uniformly spaced points. Regardless of the selection strategy, the manufacturer introduced constraint of no consecutive indentation being more than 1000  $\mu\text{m}$  apart must be followed to avoid triggering the reconfiguration penalty. Another essential part of the strategy must include defining a safe zone in the *Z*-direction. In our previous work,<sup>38</sup> it has been demonstrated that the topography variation of the sample can be approximated by a linear smooth surface fitted using *X*, *Y*, *Z* coordinates under a microscope (or the



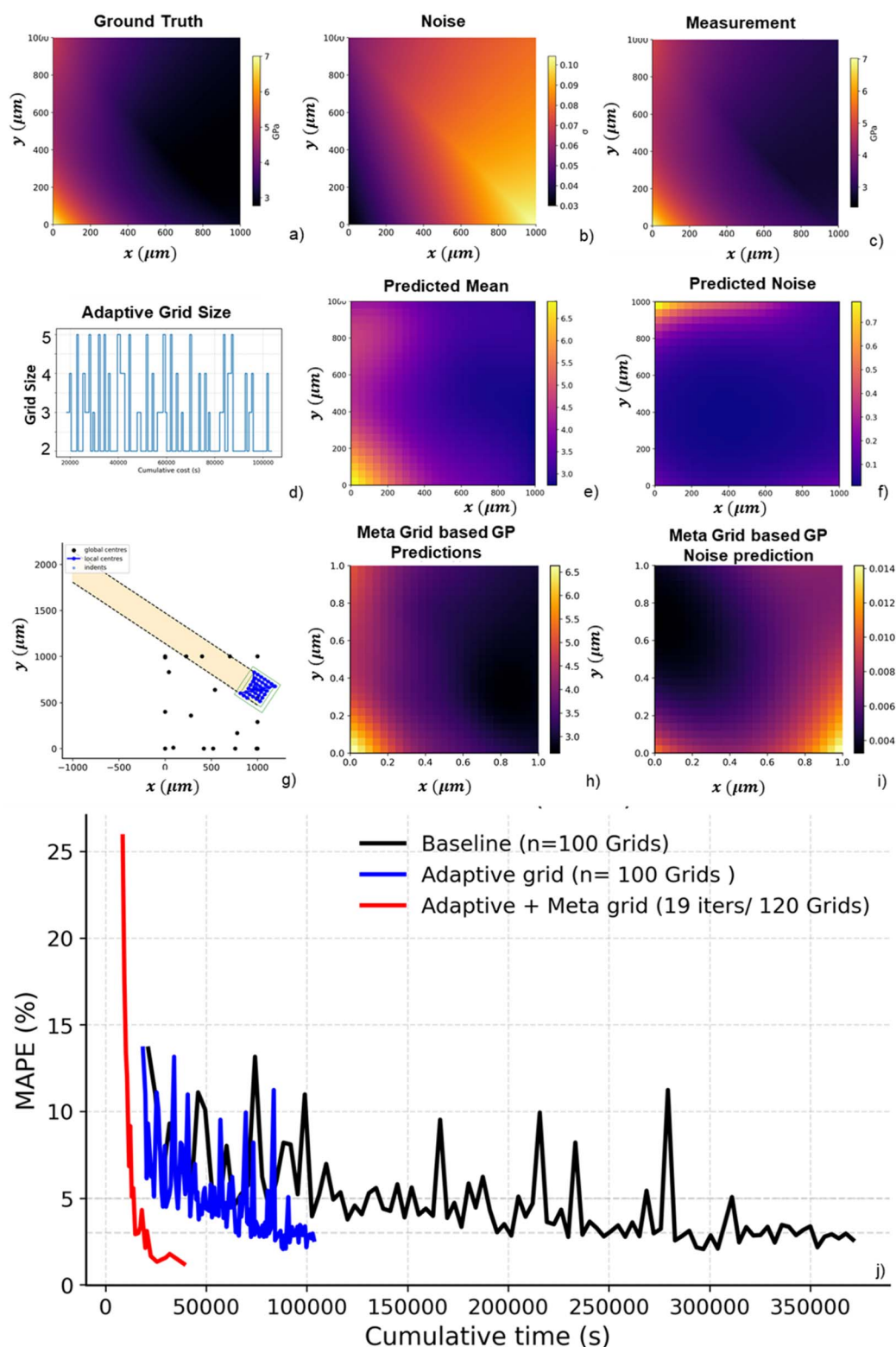


Fig. 6 Adaptive grid sizing and the meta-grid strategy under heteroscedastic noise. (a–c) Ground-truth hardness, spatially varying noise field, and corresponding noisy measurements. (d) Evolution of grid size with cumulative cost showing dynamic resizing between  $2 \times 2$  and  $5 \times 5$  based on local noise. (e and f) Predicted mean and noise fields from the trained heteroscedastic regressor. (g and h) Meta-grid configuration showing the transition from local to global exploration. (i and j) Comparison of mean absolute percentage error versus cumulative time for baseline, adaptive, and adaptive + meta-grid cases, demonstrating a fourfold reduction in total experiment time while maintaining accuracy.



indenter). The details are part of the broader automation framework and can be found elsewhere.<sup>38</sup> Once a local Z-gradient is known, a constraint that any movement from the global grid center ( $g_{x1}, g_{y1}$ ) stays within a maximum allowable vertical deviation, *i.e.*,  $\Delta Z < \Delta Z_{\text{safe}}$  can be enforced. This defines a safe operating region within which sub-grids can be explored.

In the current study, the above is shown using a predefined layout in which sub-grids are spaced 500  $\mu\text{m}$  apart, forming a rectangular pattern within the safe zone. The safe zone is approximated based on an assumed linear Z-surface in the emulator. After each sub-grid is completed, the algorithm evaluates whether to proceed to a new global location, which comes with a reconfiguration cost, or to continue within the current safe zone. This decision is made using  $\frac{\sigma}{\text{cost}}$ , where  $\sigma$  represents model uncertainty, and cost represents the time penalty of moving. Initially, the algorithm favors global moves, but as uncertainty reduces, it shifts to exploring local sub-grids. Fig. 6g illustrates the above algorithm, where black dots denote globally visited points and blue dots represent locally visited points during the current iteration. The center of each local grid represents the current global center, and the surrounding blue points form the sub-grid selected by the algorithm for investigation before moving to the next global center. The diagonal region enclosed by the dashed line represents the safe zone defined for the current global center. In this iteration, 8 sub-grids were explored, all without triggering reconfiguration penalties. Fig. 6h and i show the predicted mean hardness, and the predicted noise variation of the emulator, respectively. The model not only captured the spatial noise variation qualitatively but also achieved nearly a tenfold reduction in total time for a comparable error level (see the red and black curves in Fig. 6j). The inset of Fig. 6j illustrates the number of local grids executed within the meta-grid framework. Initially, only a single grid was explored; however, as global uncertainty decreased, an increasing number of local grids were adaptively investigated. As noted earlier, caution is warranted since excessive local exploration can lead to kernel length collapse, where the model begins interpreting noise as meaningful data. Therefore, incorporating a human-in-the-loop component may be essential to maintain model stability in those cases. The computational cost of the adaptive workflow is dominated by acquisition-related operations rather than GP training. Across all cases, acquisition accounts for approximately 65–75% of the total runtime; a detailed analysis is provided in the SI (Fig. S3).

The performance of the above framework is inherently tied to the validity of the Gaussian Process (GP) assumptions. When the underlying response is sufficiently smooth or can be guided by appropriate priors, the GP provides an effective surrogate for adaptive decision-making. However, in cases involving highly complex, discontinuous, or noise-dominated behavior, the GP may begin to interpret local noise as a signal, leading to excessive local refinement and reduced efficiency. In such scenarios, strategies that emphasize broader, global exploration in the early stages, followed by localized refinement, can improve robustness. Incorporating physics-informed priors or constraints can further enhance model fidelity. These

considerations are not specific to the present application but reflect general limitations of GP-based surrogate modeling.

## 5 Materials characterization via adaptive exploration

To experimentally demonstrate the proposed cost-aware adaptive indentation framework, the workflow was implemented on a co-sputtered Ta–Ti–Hf–Zr thin-film combinatorial library (Materials and Methods) using an automated iMicro Nano-indenter. This section translates the simulation-based active learning and cost-modeling concepts into a real experimental setting. The film serves as a model quaternary system with continuous composition gradients, allowing direct testing of the framework's ability to adaptively plan indentations under realistic instrument constraints. The thin-film indentation workflow was experimentally automated by combining stage-space parametrization, heteroskedastic Gaussian-process modeling (both mean and noise are fitted with a GP with constant mean prior), and cost-aware planning under instrument and drift constraints as described in previous simulations. Four calibrated corners-top-left (TL), top-right (TR), bottom-right (BR), and bottom-left (BL)-define an affine map from stage coordinates ( $x, y$ ) to unit-square coordinates ( $u, v$ ) using BL as the origin. The film is discretized into 2500 candidate centers ( $50 \times 50$ ) in stage space (Fig. 7a), while modeling and acquisition operate in ( $u, v$ ). Two initial  $5 \times 5$  local indentation grids are conducted at random centers with an 80 s hold. Each indentation grid is collapsed to one robust hardness measurement for a tested center given by the median of the measurements.

Prior to scheduling any local indentation grids, the “safe band” is computed to avoid topographic violation. Subsequently, the heteroskedastic GP models the hardness median in normalized space ( $u, v$ ) and then reports the mean predictions and uncertainty in physical  $x$ - $y$  space. Consequently, uncertainty is used to identify the next global or primary center and grid size, respectively. Candidate centers are ranked by uncertainty divided by time cost. The local cost scales with current grid size ( $g$ )  $\times$  (move + hold) while the global cost includes an additional reconfiguration of 520 s. Hold-time selection is drift-aware *i.e.*, using the last measured drift rate (DR) and a nominal maximum depth of 100 nm (target indentation depth), the percent area error is  $\left| \frac{h_{\text{meas}}^2 - h_{\text{true}}^2}{h_{\text{true}}^2} \right| \times 100$  with  $h_{\text{true}} = h_{\text{meas}} - \text{DR} \times t_{\text{measurement}}$  determined, and subsequently if the error is below 2% the method “hold = 5 s” is chosen; otherwise “hold = 80 s” is chosen.

After the two initial randomly chosen  $5 \times 5$  indentation grids, the active-learning loop begins. The early seed points exhibited high measurement noise (higher standard deviation within a grid). In such cases, the Gaussian Process (GP) model tends to remain highly uncertain, causing the exploration to become locally confined and potentially leading to kernel instability (“kernel crash”). To mitigate this, the first four iterations (the exploratory phase of active learning) were restricted



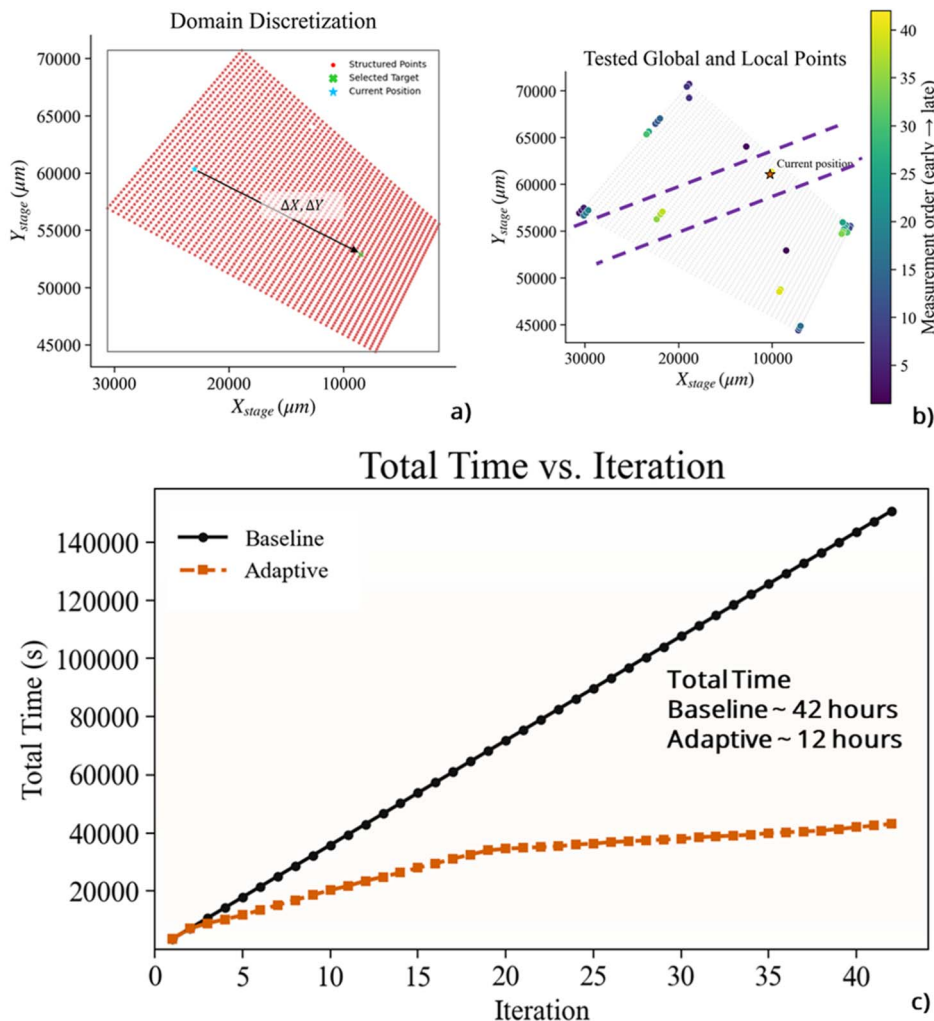


Fig. 7 Illustration of global-local exploration during adaptive indentation. (a) A  $50 \times 50$  sampling grid, shown with the x-axis reversed to match the indenter stage motion. (b) Example of executed centers (colored by order), the next global candidate (star), and planned local centers. (c) Total characterization time as a function of iteration for conventional grid-based characterization and adaptive GP-based characterization.

to a minimal number of local centers to promote stable kernel formation. Local centers were identified by estimating a planar surface model,  $z(x, y) = ax + by + c$ , fitted to three measured corner elevations at the top-left (TL), top-right (TR), and bottom-right (BR) points, providing an estimate of the local surface slope and offset. This model was used to define a safe band around each global center, within which indentation locations were permitted. During each global iteration, local centers were queued sequentially before testing. After the first four exploratory iterations, up to five local centers were allowed, subject to a maximum center-to-center spacing of  $\leq 500 \mu\text{m}$ . Candidate locations violating either the tested-point exclusion zone or the safe-band constraint were automatically skipped. In addition, an edge penalty was introduced to counteract the tendency of the GP to favor boundary regions toward the later stages of exploration. First, the domain  $(x, y)$  was normalized to  $(u, v) \in [0, 1]^2$ . A penalty based on the minimum distance to the nearest boundary,  $d_{\text{edge}}$ , was incorporated into the acquisition function via

$$P_{\text{edge}} = 1 + w_{\text{edge}}(1 - d_{\text{norm}}) \quad (12)$$

where  $d_{\text{norm}} = d_{\text{edge}}/0.5$  and the weight  $w_{\text{edge}} = 0.5$ . The final score was computed as  $\text{score} = \frac{\sigma_{\text{epistemic}}}{\text{cost} \times P_{\text{edge}}}$ , such that  $P_{\text{edge}}$  increases near the boundary, thereby down weighting edge-adjacent locations.

Each accepted local grid center is added as a  $g \times g$  grid at  $5 \mu\text{m}$  pitch as tentative points for indentation. After each the GP is refit and the stay-local *versus* jump-global decision is recomputed after every completed block. Upon completing each block of  $g^2$  indents the system exports data, refreshes the training table, refits the GP, and compares staying on the next local center *versus* jumping to the best global candidate by evaluating uncertainty per time cost, including the 520 s reconfiguration penalty for the global move. If the local score is larger, indentation continues locally; otherwise, tests are stopped and the plan is recomputed around the new global primary. An additional cost check under slow hold evaluates whether switching



to the fast hold (based on newly measured drift) would save at least the reconfiguration penalty ( $g^2(80-5) \geq$  reconfiguration cost); if so, tests are stopped and the new global center with fast hold is selected. Fig. 7a shows a  $50 \times 50$  (2500-point) grid with the  $x$ -axis reversed: in this setup, increasing stage  $x$  moves the sample to the right while the tip is fixed, so indentation occurs to the left. Hence  $x$  decreases toward the right. Fig. 7b shows an example of all visited centers (colored by recency).

Fig. 7c compares the total accumulated time *versus* iteration for the dynamic strategy. The black curve represents a baseline GP policy that ignores reconfiguration cost and thus repeatedly jumps to the most uncertain point after each block. The orange curve shows the proposed cost-aware policy, which weighs uncertainty against execution cost, remains within local neighborhoods when efficient, and pays the 520 s reconfiguration penalty only when globally justified. Over the full run, the baseline required 150 780 s ( $\approx 42$  h) to reach the same endpoint,

whereas the cost-aware policy was completed in 43 176 s ( $\approx 12$  h). While a plain GP policy already outperforms naïve grid sampling, the cost-aware approach reduces total experiment time by roughly  $3.5\times$  through efficient local exploration and adaptive hold-time adjustment.

Fig. 8 summarizes the Gaussian Process (GP) prediction, noise, and uncertainty fields. Fig. 8a shows the GP mean, with overlaid points marking the measurement centers. Several tight clusters correspond to locally executed  $g \times g$  grids; these clusters appear when the local “stay” score (posterior uncertainty divided by local cost) exceeds the best global “jump” score (uncertainty divided by global cost plus the 520 s reconfiguration penalty). The mean surface varies smoothly across the domain, exhibiting gentle gradients from Ta/Ti-rich to Hf/Zr-rich regions—consistent with compositional variation in the quaternary system. In the absence of any change in the physical mechanism, such smooth monotonic behavior is expected.

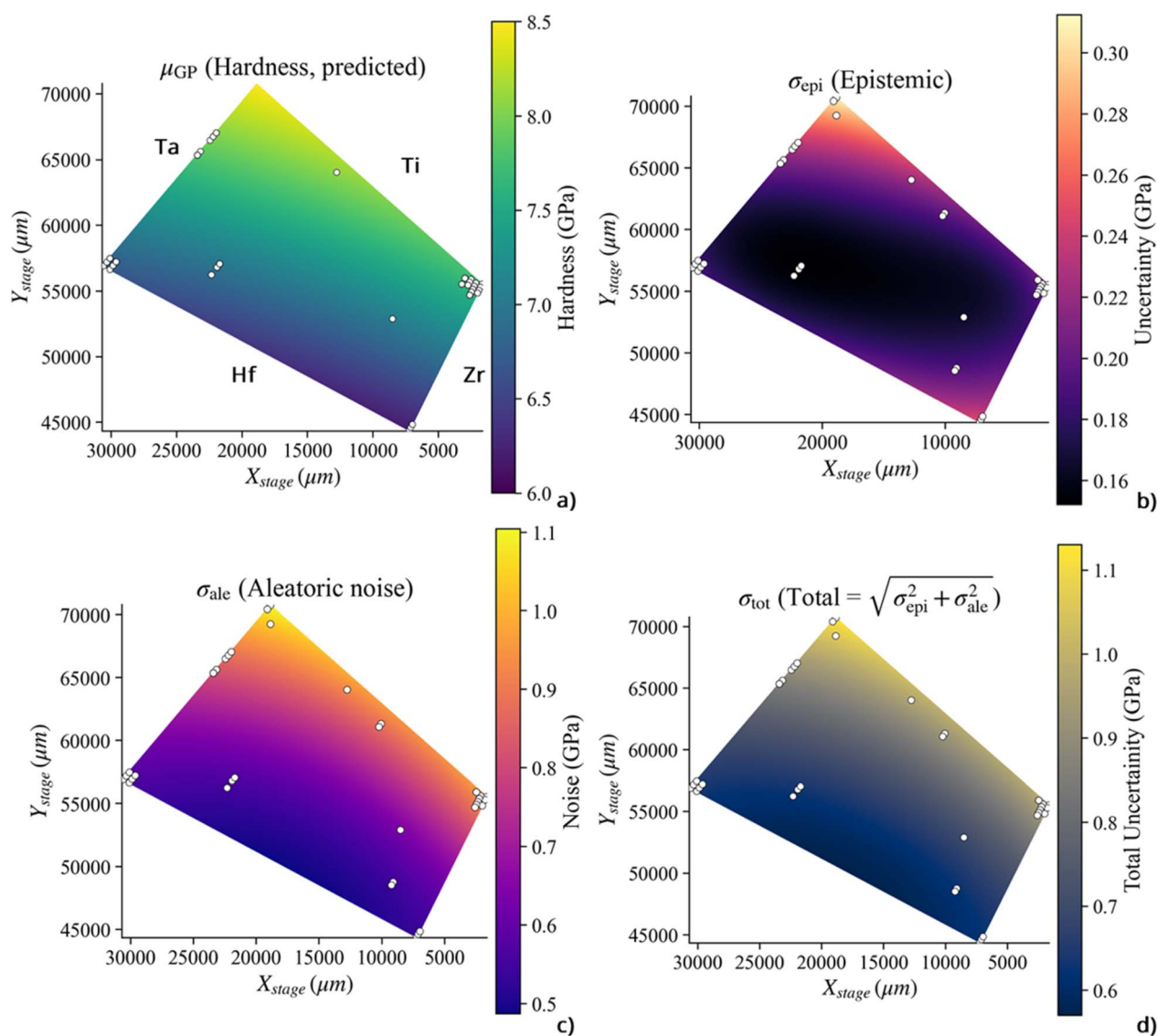


Fig. 8 (a–d) Gaussian process outputs: mean prediction, estimated noise, epistemic uncertainty, and total uncertainty. Local clusters appear where the local “stay” score exceeds the global “jump” score.



The predicted hardness decreases gradually from the upper-left (Ta–Ti-rich) to the lower-right (Hf–Zr-rich) region, aligning with theoretical expectations based on previous studies.<sup>41</sup>

Fig. 8b shows the epistemic uncertainty, derived as the standard deviation of the latent function variance returned by the heteroskedastic GP. This metric governs the selection of the next global centers. During the initial iterations, exploration naturally focuses on the corners (BL, TR, BR, and TL) despite the corner penalty, as global uncertainty dominates and no region yet warrants local refinement. As observations accumulate, exploration shifts toward the interior, followed by renewed corner exploration once the algorithm finds local regions worth revisiting. The uncertainty term alone determines where to measure next, while the noise term is used to size the local grid according to the target SEM criterion.

Fig. 8c shows the estimated noise field of the heteroskedastic GP. The model predicts noise exceeding 10% of the maximum hardness in some regions, varying similarly to the hardness field itself. Unlike epistemic uncertainty, the noise reflects irreducible variability arising from instrument drift, surface roughness, or film heterogeneity at the indentation scale. The noise map remains comparatively smooth and slowly varying across the domain, with slightly elevated values near the edges. Fig. 8d shows the combined uncertainty field across the entire domain.

## 6 Materials and methods

As a model system, a  $\text{Ta}_w\text{Ti}_x\text{Hf}_y\text{Zr}_z$  thin film was deposited by co-sputtering from four pure metal targets (Ta, Ti, Hf, and Zr) to fabricate a combinatorial composition library on a 100 mm-diameter fused silica substrate with a thickness of 500  $\mu\text{m}$ . The deposition chamber was evacuated to a base pressure of approximately  $3 \times 10^{-7}$  Torr prior to sputtering. During deposition, the sputtering powers applied to the Ta, Ti, Hf, and Zr targets were 30 W (DC), 190 W (DC), 165 W (RF), and 145 W (RF), respectively. The sputtering process was conducted for 70 minutes under a working argon pressure of 5 mTorr. After sputtering, the thin film was annealed at 1000 °C for 1 hour under vacuum. The details of the composition variation and the location of the extracted subsection within the wafer are provided in the SI.

The nanoindentation experiments were performed using an automated iMicro nanoindenter mounted on a vibration isolation table to minimize environmental disturbances during measurement. The instrument was equipped with a Berkovich diamond tip for all indentation tests. Automated control of the indenter, including stage movement and test execution, was achieved through a Python-based interface developed in our previous work.<sup>38</sup>

## 7 Conclusion

This study establishes a cost-aware, adaptive framework for automated nanoindentation and workflow optimization that substantially accelerates materials discovery. Conventional nanoindentation workflows—whether based on dense grid

searches or simple Gaussian-process exploration—are limited by the high cost of reconfiguration, drift stabilization, and redundant measurements. By explicitly modeling these costs and embedding them within a Bayesian-optimization framework, the work operationalizes a universally applicable automation strategy that accounts for instrument precision, positional repeatability, and drift behavior. The strategy combines heteroskedastic Gaussian processes, adaptive hold times, adaptive grid sizing, and meta-grid formation to simultaneously capture instrument constraints, local noise, and global uncertainty. The result is an intelligent system that balances exploration with practical execution costs, thereby improving both efficiency and data fidelity.

Through emulator studies and experimental validation on co-sputtered thin films, the proposed method demonstrates an up to thirty-fold improvement in effective throughput compared to traditional grid-based approaches. Even relative to standard Gaussian-process acquisition strategies, the incorporation of cost awareness approximately halves the total experimental time while preserving accuracy in the reconstructed hardness maps. The hierarchical meta-testing workflow, combining local-grid and global-exploration steps, prevents kernel-length collapse and ensures stable convergence across measurements with differing time scales. These gains are achieved not by compromising data quality but by strategically reducing redundant operations, adapting measurement protocols to local conditions, and minimizing high-overhead reconfigurations. The results highlight the potential of integrating decision theory with laboratory automation to overcome long-standing characterization bottlenecks in high-throughput materials research.

It is shown that the method can reduce the time taken to discover the variation of material properties within the film by a factor of 10 in comparison to the GP with a purely uncertainty-based acquisition function which itself has been shown to work 3–5 times faster in comparison to the standard grid search-based approaches. The methodology developed in this work is initially developed based on an emulator of a thin film. The results are experimentally confirmed using a co sputtered thin film. Therefore, the results presented are more than a proof of concept. They demonstrate that intelligent automation, when combined with practical cost considerations, can drastically accelerate materials discovery. Instead of relying on brute force mapping of the composition or process space, we leverage a principal framework that adapts to the landscape of property space and makes decisions that are both data efficient and cost effective.

The broader implication of this framework is the shift from passive, fixed protocols toward active and context-aware experimentation. By treating cost not as an afterthought but as a central component of experimental design, the workflow transforms nanoindentation into a scalable tool for combinatorial libraries, thin films, and complex microstructures. Because the emulator and cost models are architecture-agnostic, the same principles can be generalized to other autonomous or semi-autonomous instruments where positioning accuracy, stabilization time, or environmental drift



constrain throughput. Beyond indentation, the general principles outlined here are applicable to a wide class of automated characterization platforms where stage movement, stabilization, or calibration times dominate the measurement budget. As synthesis methods continue to generate increasingly complex libraries, adaptive cost-aware exploration will be essential for maintaining parity between fabrication and characterization. Cost-aware Bayesian optimization coupled with automated nanoindentation provides a pathway to drastically shorten the time required to extract reliable mechanical property maps. This approach enables data-efficient exploration of large parameter spaces and offers a generalizable blueprint for accelerating materials discovery.

## Author contributions

V. C. and S. V. K. conceived the overall idea and research direction. V. C. performed the initial simulations, carried out the experiments, and drafted the manuscript. S. V. K. provided core conceptual guidance, contributed to the development and refinement of the framework, and completed the first round of manuscript revision. S. P. contributed to experimental validation and subsequent manuscript revisions. H. Z. and P. D. R. provided the materials used in the experimental investigation and assisted with manuscript revisions. D. P. supported the overall direction of the project and provided guidance during manuscript preparation and revision.

## Conflicts of interest

All authors declare no financial or non-financial competing interests.

## Data availability

All codes, datasets, and notebooks supporting this study are openly available through the GitHub repository and archived on Zenodo. The latest archived version of the repository is available at <https://doi.org/10.5281/zenodo.20421606>, while the original archived release is available at <https://doi.org/10.5281/zenodo.20336715>. The GitHub repository is available at <https://github.com/VivekAChawla/AE-NI-Gradient-Library-Workflow>.

Supplementary information (SI) is available. See DOI: <https://doi.org/10.1039/d5dd00518c>.

## Acknowledgements

This research was primarily supported by the National Science Foundation Materials Research Science and Engineering Center (MRSEC) program through the UT Knoxville Center for Advanced Materials and Manufacturing (DMR-2309083).

## References

- 1 A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, Scaling deep learning for materials discovery, *Nature*, 2023, **624**(7990), 80–85.
- 2 J. Hafner, C. Wolverton and G. Ceder, Toward Computational Materials Design: The Impact of Density Functional Theory on Materials Research, *MRS Bull.*, 2006, **31**(9), 659–668.
- 3 A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, Scaling Deep Learning for Materials Discovery, *Nature*, 2023, **624**(7990), 80–85.
- 4 Meta AI., *Accelerating Materials Science with Large-Scale Machine Learning Models*, 2024.
- 5 Microsoft Research, *MatterGen: A New Paradigm of Materials Design with Generative AI*, 2025.
- 6 R. W. Epps, K. C. Felton, C. W. Coley and M. Abolhasani, Automated microfluidic platform for systematic studies of colloidal perovskite nanocrystals: Towards continuous nano-manufacturing, *Lab Chip*, 2017, **17**(23), 4040–4047.
- 7 A. Halder, M. B. Alghalayini, S. Cheng, N. Thalanki, T. M. Nguyen, A. R. Hering, D. K. Lee, S. Arnold, M. S. Leite, E. Barnard, A. Razumtcev, M. Wall, A. Gashi, Y. R. Liu, M. M. Noack, S. Sun and C. M. Sutter-Fella, AI-Driven Robot Enables Synthesis-Property Relation Prediction for Metal Halide Perovskites in Humid Atmosphere, *Adv. Energy Mater.*, 2025, **15**(34), 2502294.
- 8 K. Higgins, M. Ziatdinov, S. V. Kalinin and M. Ahmadi, High-Throughput Study of Antisolvents on the Stability of Multicomponent Metal Halide Perovskites through Robotics-Based Synthesis and Machine Learning Approaches, *J. Am. Chem. Soc.*, 2021, **143**(47), 19945–19955.
- 9 K. Higgins, S. Mani Valleti, M. Ziatdinov, S. V. Kalinin and M. Ahmadi, Chemical Robotics Enabled Exploration of Stability and Photoluminescent Behavior in Multicomponent Hybrid Perovskites *via* Machine Learning, *ACS Energy Lett.*, 2020, **5**(10), 3427–3436.
- 10 J. Zhang, J. A. Hauch and C. J. Brabec, Toward Self-Driven Autonomous Material and Device Acceleration Platforms (AMADAP) for Emerging Photovoltaics Technologies, *Acc. Chem. Res.*, 2024, **57**(9), 1434–1445.
- 11 X.-D. Xiang, X. Sun, G. Briceño, Y. Lou, K.-A. Wang, H. Chang, W. G. Wallace-Freedman, S.-W. Chen and P. G. Schultz, A Combinatorial Approach to Materials Discovery, *Science*, 1995, **268**(5218), 1738–1740.
- 12 I. Takeuchi, R. van Dover, H. Koinuma, I. Takeuchi, R. Dover, B. van and H. Koinuma, Combinatorial Synthesis and Evaluation of Functional Inorganic Materials Using Thin-Film Techniques, *MRS Bull.*, 2002, **27**(4), 301–308.
- 13 W. C. Oliver and G. M. Pharr, Measurement of Hardness and Elastic Modulus by Instrumented Indentation: Advances in Understanding and Refinements to Methodology, *J. Mater. Res.*, 2004, **19**(1), 3–20.
- 14 W. C. Oliver and G. M. Pharr, An Improved Technique for Determining Hardness and Elastic Modulus Using Load



- and Displacement Sensing Indentation Experiments, *J. Mater. Res.*, 1992, **7**(6), 1564–1583.
- 15 G. M. Pharr, W. C. Oliver and F. R. Brotzen, On the Generality of the Relationship among Contact Stiffness, Contact Area, and Elastic Modulus during Indentation, *J. Mater. Res.*, 1992, **7**(3), 613–617.
- 16 J. E. Campbell, R. P. Thompson, J. Dean and T. W. Clyne, Experimental and computational issues for automated extraction of plasticity parameters from spherical indentation, *Mech. Mater.*, 2018, **124**, 118–131.
- 17 S. Kuttler, H. Walter, A. Grams, M. Schneider-Ramelow, and S. Huber, Determination of Stress-Strain Properties Combining Small-Depth Nanoindentation and Numerical Simulation, *7th Electronic System-Integration Technology Conference (ESTC)*, DOI: [10.1109/ESTC.2018.8546464](https://doi.org/10.1109/ESTC.2018.8546464).
- 18 J. Weng, R. Lindvall, K. Zhuang, J. E. Ståhl, H. Ding and J. Zhou, A machine learning based approach for determining the stress-strain relation of grey cast iron from nanoindentation, *Mech. Mater.*, 2020, **148**, 103522.
- 19 K. D. Bouzakis, N. Michailidis, S. Hadjiyiannis, G. Skordaris and G. Erkens, The effect of specimen roughness and indenter tip geometry on the determination accuracy of thin hard coatings stress-strain laws by nanoindentation, *Mater. Charact.*, 2002, **49**(2), 149–156.
- 20 E. G. Herbert, K. E. Johanns, R. S. Singleton and G. M. Pharr, On the measurement of energy dissipation using nanoindentation and the continuous stiffness measurement technique, *J. Mater. Res.*, 2013, **28**(21), 3029–3042.
- 21 C. Su, E. G. Herbert, S. Sohn, J. A. LaManna, W. C. Oliver and G. M. Pharr, Measurement of power-law creep parameters by instrumented indentation methods, *J. Mech. Phys. Solids*, 2013, **61**(2), 517–536.
- 22 P. S. Phani, W. C. Oliver and G. M. Pharr, On the effective load during nanoindentation creep testing with continuous stiffness measurement (CSM), *J. Mater. Res.*, 2021, **36**(8), 1740–1750.
- 23 P. S. Phani, W. C. Oliver and G. M. Pharr, Understanding and modeling plasticity error during nanoindentation with continuous stiffness measurement, *Mater. Des.*, 2020, **194**, 108923.
- 24 P. S. Phani, W. C. Oliver and G. M. Pharr, Influences of elasticity on the measurement of power law creep parameters by nanoindentation, *J. Mech. Phys. Solids*, 2021, **154**, 104527.
- 25 M. Ghidelli, M. Sebastiani, K. E. Johanns and G. M. Pharr, Effects of indenter angle on micro-scale fracture toughness measurement by pillar splitting, *J. Am. Ceram. Soc.*, 2017, **100**(12), 5731–5738.
- 26 T. Zhang, Y. Feng, R. Yang and P. Jiang, A method to determine fracture toughness using cube-corner indentation, *Scr. Mater.*, 2010, **62**(4), 199–201.
- 27 V. Chawla, S. B. Piplampu, W. Hopper and D. Penumadu, Splitting Mode-I fracture toughness of carbon fibers using nanoindentation, *Carbon N. Y.*, 2024, **219**, 118777.
- 28 M. Sebastiani, K. E. Johanns, E. G. Herbert and G. M. Pharr, Measurement of fracture toughness by nanoindentation methods: Recent advances and future challenges, *Curr. Opin. Solid State Mater. Sci.*, 2015, **19**(6), 324–333.
- 29 M. Sebastiani, K. E. Johanns, E. G. Herbert, F. Carassiti and G. M. Pharr, A novel pillar indentation splitting test for measuring fracture toughness of thin ceramic coatings, *Philos. Mag.*, 2015, **95**(16–18), 1928–1944.
- 30 B. L. Hackett, P. Sudharshan Phani, C. C. Walker, W. C. Oliver and G. M. Pharr, On the Strain Rate Sensitivity Measured by Nanoindentation at High Strain Rates, *JOM*, 2024, **76**(6), 2936–2945.
- 31 P. Sudharshan Phani, B. L. Hackett, C. C. Walker, W. C. Oliver and G. M. Pharr, High strain rate nanoindentation testing: Recent advancements, challenges and opportunities, *Curr. Opin. Solid State Mater. Sci.*, 2023, **27**(1), 101054.
- 32 P. S. Phani, B. L. Hackett, C. C. Walker, W. C. Oliver and G. M. Pharr, On the measurement of hardness at high strain rates by nanoindentation impact testing, *J. Mech. Phys. Solids*, 2023, **170**, 105105.
- 33 E. G. Herbert, W. C. Oliver and G. M. Pharr, Nanoindentation and the dynamic characterization of viscoelastic solids, *J. Phys. D Appl. Phys.*, 2008, **41**(7), 074021.
- 34 E. G. Herbert, W. C. Oliver, A. Lumsdaine and G. M. Pharr, Measuring the constitutive behavior of viscoelastic solids in the time and frequency domain using flat punch nanoindentation, *J. Mater. Res.*, 2009, **24**(3), 626–637.
- 35 J. L. Reed, J. Dean, G. Aldrich-Smith and T. W. Clyne, A methodology for obtaining plasticity characteristics of metallic coatings via instrumented indentation, *Int. J. Solids Struct.*, 2016, **80**, 128–136.
- 36 W. B. Powell. *Sequential Decision Analytics and Modeling: Modeling with Python, Part I*, Wiley, Hoboken, NJ, 2022.
- 37 W. B. Powell. *Sequential Decision Analytics and Modeling: Modeling with Python, Part II*, Wiley, Hoboken, NJ, 2022.
- 38 V. Chawla, D. Penumadu and S. Kalinin, Automating Nanoindentation: Optimizing Workflows for Precision and Accuracy, *arXiv*, 2025, preprint, arXiv: 2504, p. 18525, DOI: [10.48550/arXiv.2504.18525](https://doi.org/10.48550/arXiv.2504.18525).
- 39 Y. Liu, M. Ziatdinov, R. Vasudevan and S. V. Kalinin, *Post-Experiment Forensics and Human-in-the-Loop Interventions in Explainable Autonomous Scanning Probe Microscopy*, GitHub Repository, 2023, <https://github.com/yongtaoliu/Forensics-DKL-BEPS>.
- 40 U. Pratiush, K. M. Roccapriore, Y. Liu, G. Duscher, M. Ziatdinov and S. V. Kalinin, *Building Workflows for Interactive Human in the Loop Automated Experiment (HAE) in STEM-EELS*, GitHub Repository, 2025, <https://github.com/utkarshp1161/hAE/tree/main/hAE>.
- 41 R. Emery, S. B. Piplampu, A. Wood, D. Penumadu, E. A. Lass, J. Lasseter, K. A. Unocic, S. Chen, Y. Zhao, S. J. Zinkle, T. Liang, H. Xu, D. A. Gilbert, C. C. Buchanan, P. K. Liaw and P. D. Rack, Thin Film Combinatorial Sputtering of TaTiHfZr Refractory Compositionally Complex Alloys for Rapid Materials Discovery, *Mater. Des.*, 2025, **250**, 113643.

