Construction of the Synthetic Hardness Emulator

A synthetic composition–property landscape was created to evaluate the automated nanoindentation workflow without relying on any specific material class. The purpose of this emulator is to generate a smooth but nontrivial property field that contains gradients, curvature, and local interactions so that the adaptive measurement algorithm can be tested in a controlled setting. The three composition fields $C_A(x,y)$, $C_B(x,y)$, and $C_c(x,y)$ were assigned spatial distributions that vary diagonally across the domain, which resembles the type of gradients produced in combinatorial sputtering. Figure S1 shows the resulting spatial maps for the three components along with the corresponding synthetic hardness field used as ground truth.

The hardness model itself is not taken from any specific alloy system. Instead, it is a phenomenological construction designed purely for generating a property landscape with realistic complexity. The first term in the model is a simple rule-of-mixtures expression that provides a baseline contribution proportional to the local compositions. The second term consists of pairwise products C_iC_j scaled by interaction coefficients and is included as an artificial representation of solid-solution type interactions to introduce controlled nonlinear coupling between components. This form is not a standard description of alloy hardening but was introduced here to ensure that the hardness varies not only with individual components but also with how components coexist locally. These pairwise terms create smooth nonlinear coupling between the composition fields, which helps produce a richer landscape for testing the adaptive workflow. The final term, applied to component C, penalizes very high amounts of this component and produces a maximum at intermediate concentration. This ensures that the overall hardness field contains regions of both monotonic and non-monotonic behavior.

Figure S1 shows the three composition maps generated for the emulator and the resulting hardness surface produced by the combined effect of the rule-of-mixtures, interaction, and penalty contributions. Component A increases toward the upper left, component B toward the lower right, and component C forms a diagonal gradient across the domain. The hardness field reflects all three patterns and shows the expected smooth curvature and directional variation. This synthetic landscape served as the ground truth against which the performance of the Bayesian-optimized automated nanoindentation workflow was evaluated.

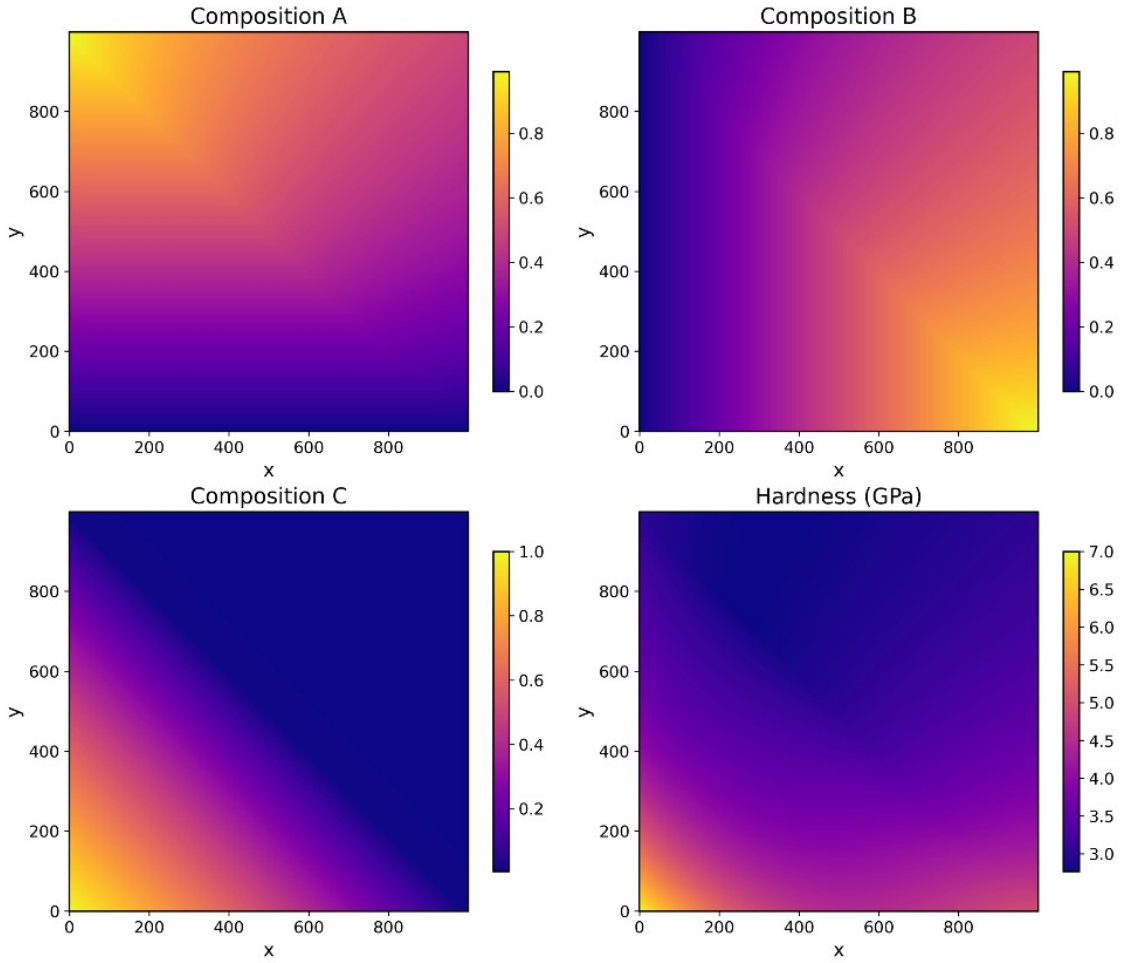

Figure S1. Thin film emulator for a ternary system. (a, c) show the spatial variation of elements A, B and C respectively, while d) shows the corresponding hardness distribution in (x, y).

Experimental thin film composition variation

Figure S2 shows the spatial distribution of Ti, Ta, Hf, and Zr across the entire wafer and within the selected subsection used in this study. The subsection corresponds to the region bounded by $0 \leq X \leq 1.25$ and $2.34 \leq Y \leq 5$, as indicated in red. Among the four elements Ta shows the strongest variation, whereas Zr varies the least within this area.

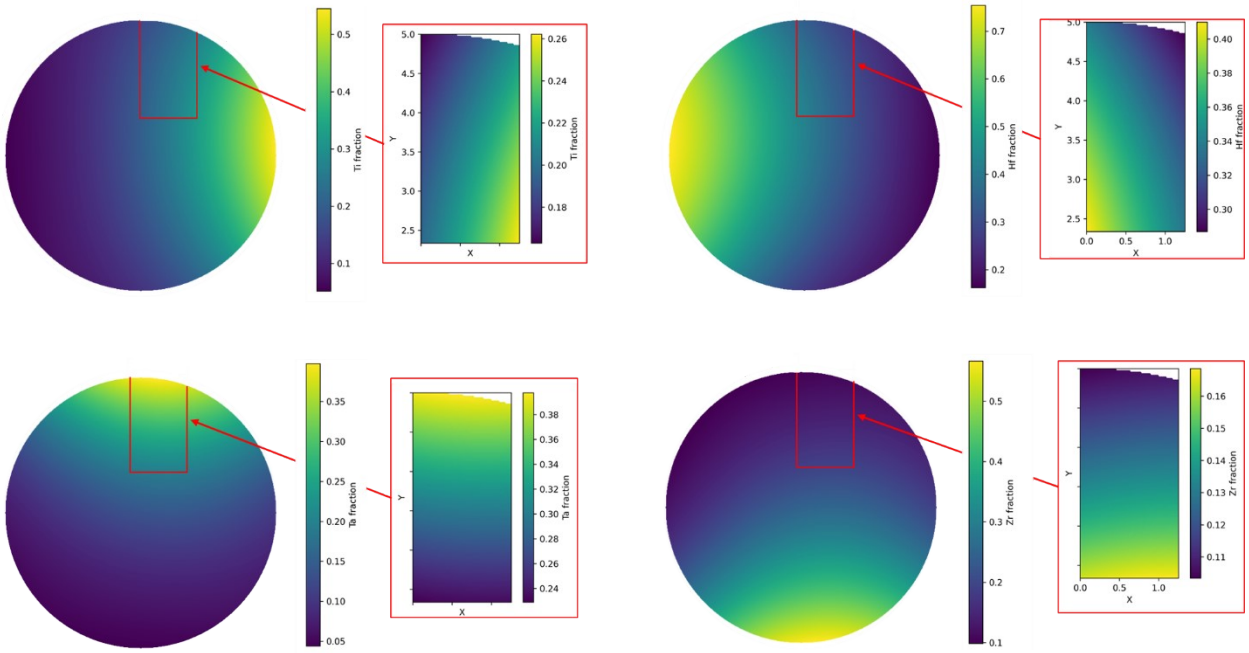

Figure S2: Spatial distribution of Ti, Ta, Hf, and Zr within the selected subsection of the composition wafer. Each panel shows the measured elemental fraction for all points inside the defined rectangular region

Computational cost analysis

To quantify the computational overhead of the adaptive workflow, we measured the wall-clock time associated with Gaussian Process (GP) training and acquisition at each iteration. The acquisition score is defined by equation S1. The corresponding acquisition time includes both the global evaluation of candidate points and the local acquisition decision steps during refinement. The acquisition time is defined using equation S2.

$score = \frac{\sigma_{epistemic}}{cost.\, P_{edge}}$	(S1)
$t_{acq} = t_{global} + t_{local\ decision}$	(S2)
$H_{baseline}(A,B,C) = 3A + 5B + 7C - 4AB - 3BC - 2CA - 6C2(1 - C)$	(S3)
$H_{smooth}(A,B,C) = 3A + 5B + 7C - 2AB - 1.5BC - AC + 0.25sin(2\pi A)cos(2\pi B)$	(S4)
$H_{sharp}(A,B,C) = 3A + 5B + 7C - 4AB - 3BC - 2CA + H_{jump}(A,B) + H_{band}(A)$	(S5)
$H_{jump}(A,B) = 1.21_{A > 0.55} - 1.01_{B > 0.60} + 0.81_{A + B < 0.45}$	(S6)
$H_{band}(A) = 0.8exp(-\frac{2(A - 0.65)^2}{2(0.01)^2})$	(S7)

To assess the influence of landscape complexity, three synthetic hardness fields were considered (Fig. S3 a-c): a smooth field with low spatial variation, a baseline field with moderate nonlinear interactions, and a sharp-gradient field containing discontinuities and localized features. The equations for each case are given by equations S3-S7. All cases were evaluated under identical conditions, including domain size, noise model, candidate pool, and number of adaptive iterations, allowing for a consistent comparison of computational scaling.

The reconstruction accuracy for the three synthetic cases is shown in the percentage error maps in Fig. S3(d–f). The smooth case exhibits the lowest error, with a maximum percentage error of approximately 0.87%, indicating effective modeling of continuous variations. The baseline case shows moderate error levels, reaching up to 1.75%, reflecting increased nonlinearity in the response surface. In contrast, the sharp-gradient case exhibits higher localized errors, with maximum values of approximately 6%, primarily concentrated near discontinuities and narrow high-gradient regions.

Across all cases, the total computational cost per iteration is dominated by acquisition-related operations rather than GP training. For the baseline case, the mean GP training time is approximately 23.9 s per iteration, compared to 27.6 s for global acquisition, 166.2 s for local GP updates, and 314.0 s for local acquisition decision steps, yielding a total of 531.7 s per iteration. For the smooth case, these values increase to 55.0 s (GP training), 80.8 s (global acquisition), 254.4 s (local GP update), and 624.7 s (local decision), with a total iteration time of 1014.9 s. The sharp-gradient case exhibits intermediate behavior, with average times of 40.3 s (GP training), 60.6 s (global acquisition), 199.8 s (local GP update), and 524.6 s (local decision), corresponding to a total of 825.3 s per iteration. Figure S3e–f summarizes these results. Figure S3e shows the time associated with different components as a function of iteration for the three cases, while Fig. S3f shows the mean time for each component.

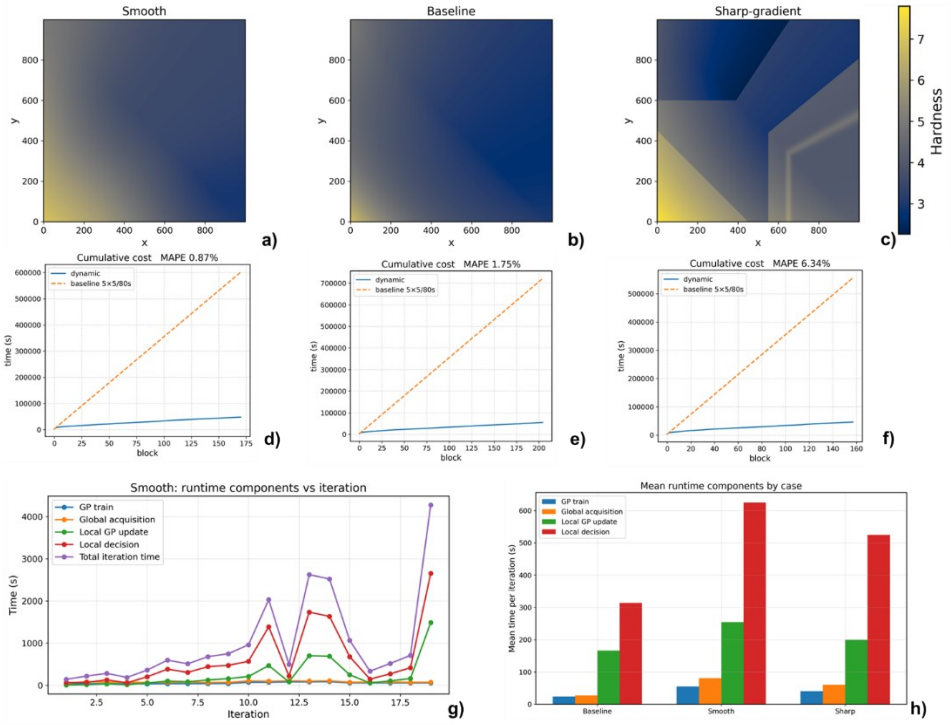

Figure S3: Synthetic hardness fields (a–c), corresponding percentage error maps (d–f), and computational cost analysis. (e) Runtime components as a function of iteration and (f) mean time for each component, showing that acquisition-related operations dominate the computational cost.