A Appendix

A.1 Active Learning Acquisition Functions

Traditional BO acquisition functions, such as EI and LCB, use the computed means and variances from a surrogate model to compute an acquisition value; maximizing this acquisition value guides sampling of the manifold [7, 32, 12]. However, these traditional acquisition functions are static, such that they do not actively use any information about the performance of previously sampled experiments to guide sampling. Hence, we implement an active learning approach into the acquisition functions to develop two novel functions, EI Abrupt and LCB Adaptive, that dynamically adapt their sampling based on the quantity and quality of previously sampled experiments. In contrast to a static acquisition function, these dynamic acquisition functions are initialized with an initial set of hyperparameter values to guide their search but then tune these values as sampling progresses. The developed EI Abrupt and LCB Adaptive functions are used within the `ZoMBI` framework to further accelerate optimization and avoid pigeonholing, see line 9 of Algorithm 1.

EI Abrupt uses actively learned information about the quality of previously measured experiment target values, y, to change sampling policies. For example, if the value of y plateaus for three or more experiments in a row, EI Abrupt will abruptly switch from a greedy sampling policy to a more explorative sampling policy. Specifically, this information feedback received by the function determines if the current round of sampling should exploit the surrogate mean values, $\mu(X)$, or explore the surrogate variances, $\sigma(X)$. EI Abrupt computes an acquisition value, $a \in [0, 1]$, for a given X, wherein the X with the highest a is selected by the acquisition function as the next suggested experiment to measure. EI Abrupt is implement for a minimization problems as:

$$
\begin{aligned}
a_{\text{EI Abrupt}}(X, y; \beta, \xi, \eta) &= \begin{cases} \left(\mu(X) - y^* - \xi\right)\Phi(Z) + \sigma(X)\psi(Z), & \text{if } |\nabla\{y_{n-3\ldots n}\}| \leq \eta \\ \mu(X) - \beta\sigma(X), & \text{otherwise} \end{cases} \\
Z &= \frac{\mu(X) - y^* - \xi}{\sigma(X)},
\end{aligned}
\tag{3}
$$

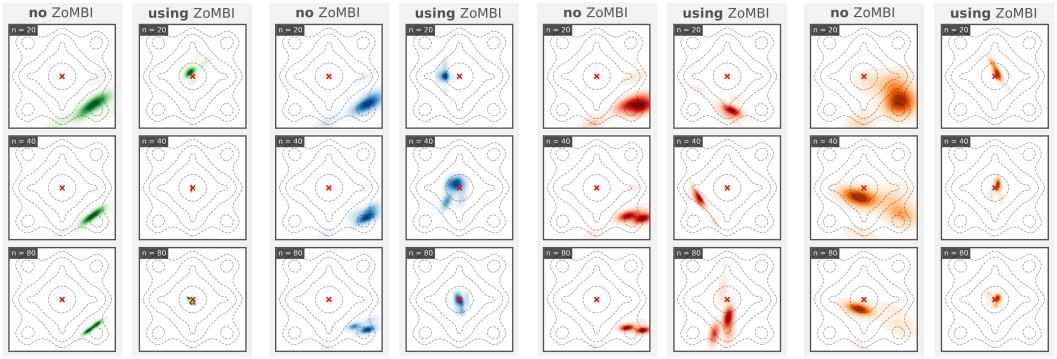
where y^* is the lowest measured target value thus far (*i.e.*, the running minimum), $\Phi(\cdot)$ is the cumulative density function of the normal distribution, $\psi(\cdot)$ is the probability density function of the normal distribution, and $|\nabla\{y_{n-3\ldots n}\}|$ is the absolute value of the gradient of the set of target values of the last three sampled experiments; a gradient of 0 indicates a plateau. Moreover, $\beta = 0.1$, $\xi = 0.1$, and $\eta = 0$ are hand-tuned initialization hyperparameters used for the rest of the paper for EI Abrupt. These hyperparameters were selected based on *a priori* domain knowledge of EI Abrupt performance on a variety of different problems. The most important hyperparameter for efficient sampling is β, whose ideal value is non-obvious, but it is found that $\beta = 0.1$ allows EI Abrupt to switch into an explorative sampling policy while still having a strong weight on the surrogate means, implying that exploration does not veer far.

LCB Adaptive uses actively learned information about the quantity of previously sampled experiments, n, to tune its hyperparameter. For example, as the n increases, LCB Adaptive decays its β hyperparameter value to become less explorative and more exploitative. Specifically, this information feedback received by the function determines the contribution of both $\mu(X)$ and $\sigma(X)$ to the acquisition value, a. Similar to EI Abrupt, LCB Adaptive computes an acquisition value, $a \in [0, 1]$, for a given X, wherein the X with the highest a is selected by the acquisition function as the next suggested experiment to measure. LCB Adaptive is implemented for a minimization problem as:

$$
a_{\text{LCB Adaptive}}(X, n; \beta, \epsilon) = \mu(X) - \epsilon^n \beta\sigma(X), \tag{4}
$$

where n is the number of experiments sampled, and $\beta = 3$ and $\epsilon = 0.9$ are hand-tuned initialization hyperparameters selected based on *a priori* domain knowledge of the function's performance on a variety of different problems. Having a large β and an ϵ close to 1 supports a gradual decay from very explorative to very exploitative, rather than a rapid decay. In the following section (Section A.2), the dynamic EI Abrupt and LCB Adaptive are shown to both discover optima in fewer experiments and avoid pigeonholing into local minima better than their static counterparts by actively balancing the ratio of exploitation to exploration using learned information about the quality and quantity of previously sampled experiments.

A.2 Acquisition Function Performance



(a) LCB (b) LCB Adaptive (c) EI (d) EI Abrupt

Figure A.1: **Acquisition Function Sampling Density.** The colored heatmaps indicate the regions of a 2D slice from a 5D Ackley function where sampling density is high for each respective acquisition function: (a) LCB, (b) LCB Adaptive, (c) EI, and (d) EI Abrupt. The contour lines indicate the manifold topology with local minima as the circular and pointed regions of the contours. The red "x" indicates the global minimum. For each acquisition function, the left panel shows the sampling density after $n = \{20, 40, 80\}$ evaluated experiments without the use of `ZoMBI` while the right panel shows the sampling density after $n = \{20, 40, 80\}$ evaluated experiments with the use of `ZoMBI`.

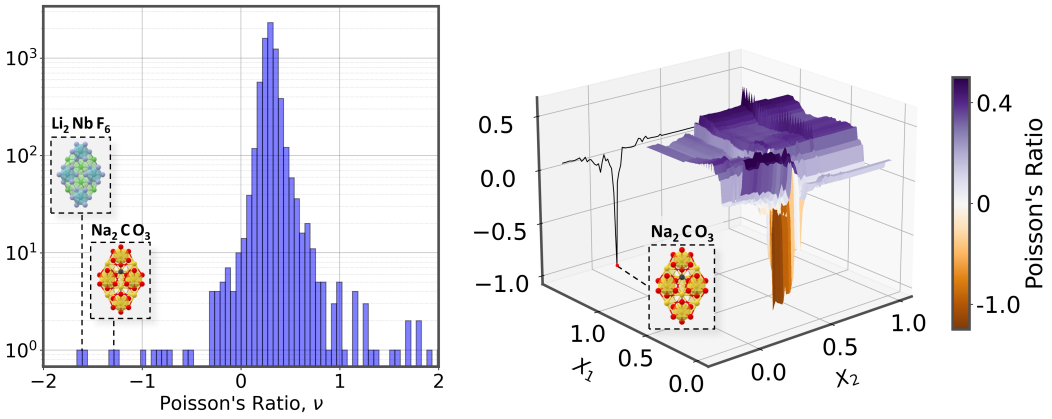
Pigeonholing into the local minima of a function occurs when an optimization algorithm has insufficient learned knowledge of the manifold topology to continue exploring potentially profitable regions or when the algorithm's hyperparameters are improperly tuned, leading to overly exploitative tendencies [1, 9]. The `ZoMBI` algorithm's anti-pigeonholing capabilities are two-fold: (1) the zooming search bounds help the acquisition function to quickly stop sampling local minima once a better performing data point is found and (2) actively learned acquisition function hyperparameters use knowledge about the domain to help exit a local minimum. Figure A.1 demonstrates the anti-pigeonholing capabilities of `ZoMBI` on optimizing a 5D Ackly function with both static and dynamic acquisition functions, compared to that of traditional BO. The needle-like global minimum is indicated by the red "x" and the local minima are indicated by the circular and pointed regions of the contour lines. The sampling density of each acquisition function is illustrated by the heatmap, where the darker colors indicate higher sampling density regions. The goal is to get high sampling density near the red "x". It is shown that without `ZoMBI` being activated, the LCB, LCB Adaptive, and EI acquisition functions all end up pigeonholing into local minima. However, EI Abrupt initially pigeonholes into a local minima but then switches from an exploitative to an explorative mode to jump out of the local minimum and converge closer to the global. Conversely, when running the optimization procedure with `ZoMBI` active, all of the acquisition functions except the most exploitative, EI, converge onto the global minimum fast. LCB Adaptive and EI are shown to initially start sampling towards a local minima, but as `ZoMBI` is iteratively activated, the search bounds zoom in closer to the global minimum. Thus, with the combination of active learning dynamic acquisition functions and zooming search bounds, pigeonholing into sub-optimal local minima can be more readily avoided while optimizing NiaH problems, although avoidance is not guaranteed, as shown by the sampling density of EI.

A.3 Rare Material Search Space Topology

Table A.1: Description of variables from the two real-world Needle-in-a-Haystack materials science datasets [20].

Training Variable	**Units**	**Description**
Density	g/cm^3	Density of the entire molecule.
Formation Energy	eV/atom	Normalized change of energy to form target phase.
Energy Above Hull	eV/atom	Normalized energy to decompose into stable phase.
Fermi Energy	eV	Highest energy level at absolute zero.
Band Gap	eV	Valence to conduction band electron excitation energy

Target Variable	**Units**	**Description**
Poisson's Ratio, ν	Unitless	Mechanical deformation perpendicular to the loading direction.
Thermoelectric Merit, ZT	Unitless	Electrical and thermal potential of a material to produce current.

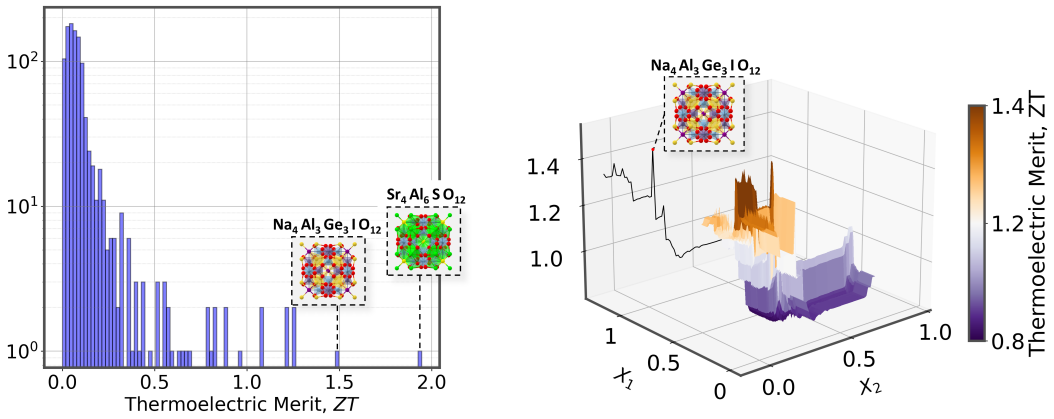


(a) Raw Dataset Histogram

(b) Computed Random Forest Manifold

Figure A.2: **Histogram and Random Forest of Poisson's Ratio Dataset.** (a) The Poisson's ratio histogram of all 146k materials in the Materials Project Dataset [20, 21] with y-axis in log-scale. The two needles are called out, indicating the locations of minimum Poisson's ratio values $\nu_{\text{min}} = \{-1.2, -1.7\}$. (b) The noisy, non-convex manifold topology generated by an RF regression of 500 trees on the Poisson's ratio dataset. A projected 2D slice of the 5D space is illustrate with z-axis and colorbar indicating the Poisson's ratio. The slice of space shown indicates the narrow basin of attraction region containing the $\nu_{\text{min}} = -1.2$ needle.

A.4 ZoMBI Algorithm



(a) Raw Dataset Histogram

(b) Computed Random Forest Manifold

Figure A.3: **Histogram and Random Forest of Thermoelectric Merit Dataset.** (a) The Thermoelectric Merit, ZT, histogram of all 1k materials in the dataset computed by BoltzTraP [55, 20] with y-axis in log-scale. The two needles are called out, indicating the locations of maximum ZT values $ZT_{\text{max}} = \{1.4, 1.9\}$. (b) The noisy, non-convex manifold topology generated by an RF regression of 500 trees on the ZT dataset. A projected 2D slice of the 5D space is illustrate with z-axis and colorbar indicating the ZT value. The slice of space shown indicates the narrow basin of attraction region containing the $ZT_{\text{max}} = 1.4$ needle.

Algorithm 1: Zooming Memory-Based Initialization (`ZoMBI`)

Input : $\mathbf{X}$: Set of data points $\{X_1, X_2, \ldots, X_n\}$, where $X_j \in \mathbb{R}^d$,
$\mathbf{y}$: Set of target values $\{y_1, y_2, \ldots, y_n\}$, where $y_j \in \mathbb{R}$,
α: Number of `ZoMBI` activations,
ϕ: Number of forward experiments per activation,
$\boldsymbol{\gamma}$: Set of acquisition function hyperparameters $\{\beta, \xi, \epsilon, \eta\}$,
AF: An acquisition function selected by the user

Output : The next experimental condition $X_{n+1} \in \mathbb{R}^d$ and measured target value $y_{n+1} \in \mathbb{R}$

for α activations **do**
  Compute bounds $\{\mathcal{B}_d^l, \mathcal{B}_d^u\} \leftarrow \{\min, \max\}_{X \in \mathbf{X}^{(m)}}\{[X]_d\}$
  Initialize with i LHS data points $\{X_i\} := \{X_1, X_2, \ldots, X_i\}$, where $X_j \in \mathbb{R}^d, [\mathcal{B}_d^l, \mathcal{B}_d^u]$ and target values $\{y_i\} := \{y_1, y_2, \ldots, y_i\}$, where $y_j \in \mathbb{R}$
  Overwrite memory $\mathbf{X} \leftarrow \{X_i\}$ and $\mathbf{y} \leftarrow \{y_i\}$
  for ϕ forward experiments **do**
    Let $n = i + \phi$
    Retrain surrogate model $f(\mathbf{X})$ using target values $\mathbf{y}$
    Extract set of surrogate means $\{\mu\}$ and variances $\{\sigma\}$
    Compute set of acquisition values $\{a\} \leftarrow AF(\{\mu\}, \{\sigma\}; \boldsymbol{\gamma})$
    Find the best new experimental condition $X_{n+1} \leftarrow \arg\max\big[\{a\}\big]$
    Measure target value of new experimental condition $y_{n+1} \leftarrow f(X_{n+1})$
    Append outputs to sets $\mathbf{X}$`.append`(X_{n+1}) and $\mathbf{y}$`.append`(y_{n+1})
  end
end