WHEN FEEDBACK IS NOT REWARD: ACTIVE SEARCH UNDER POOL-DEPENDENT UTILITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Active search typically assumes that querying an item reveals the label used by the search objective. In many scientific settings, however, a query returns a local measurement, while whether the queried item counts as a discovery depends on a shared latent pool. We formalize this setting as active search with pool-dependent utility. Under additive terminal utility and equal query costs, the one-step Bayes action ranks candidates by posterior terminal-success probability. With two queries remaining, an exact decomposition shows that lookahead improves the first action only when observation-induced decision value offsets the posterior-utility gap from leaving a higher-ranked candidate. Repeated greedy search can be arbitrarily suboptimal in the worst case, while a weak-posterior-movement condition bounds its loss. On 230 MatPES PBE-to-r2SCAN systems, using complete-pool rather than history-visible membership to define terminal success improves first-query utility by 0.657 confirmations and budget-curve area by 6.617. Once the terminal event is fixed, posterior integration provides little additional utility over a deterministic complete-pool score. An anchored two-step rollout selects a different action in 91.7% of systems at $B = 6$, but improves terminal utility by only 0.039 confirmations per system at $46.9\times$ the measured wall-clock cost. These results separate the effect of specifying the terminal event from the value of deeper planning.

1 INTRODUCTION

Active search typically assumes that a query reveals the label counted by the search objective. High-throughput materials screening exposes a different feedback interface: an expensive calculation returns a structure's energy, while its eventual stability is relational, determined by the competing phases used to construct the convex hull (Jain et al., 2013; Bartel et al., 2020; Merchant et al., 2023; Riebesell et al., 2025; Novick et al., 2024). This dependence is central to large-scale stability screening, where a locally promising structure can be displaced after another candidate is evaluated.

For an item x, we distinguish the query observation O_x from the terminal utility label Y_x. The observation arrives immediately; the label remains latent because it is a function of the shared pool state. An unqueried candidate can therefore change the reward assigned to an item that has already been queried. We call this decision structure *active search with pool-dependent utility*; Figure 1 shows the resulting decision structure on a real O–Te state. The distinction is decision-theoretic rather than a consequence of correlation alone: modern Bayesian experimental design derives acquisition decisions from an explicit downstream objective, of which information gain is only one choice (Rainforth et al., 2024; Huang et al., 2026).

Pool-dependent utility separates three choices that are usually tied together in active search: what a query reveals, what event determines terminal success, and how far ahead the policy plans. In convex-hull screening, a query reveals a target-fidelity energy, whereas reward is assigned only after the queried structure is compared with the completed candidate pool. This objective differs from reconstructing the hull itself (Neiswanger et al., 2021; Cheng et al., 2024; Novick et al., 2024): the two tasks may use the same posterior over energies but can induce different acquisition decisions.

We make three contributions. First, we formulate pool-dependent active search and show that its one-step Bayes action ranks candidates by posterior terminal-success probability. Second, we characterize the value of planning. An exact two-step identity separates observation-induced decision value from the immediate utility cost of leaving a higher-ranked candidate; repeated greedy search can be

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

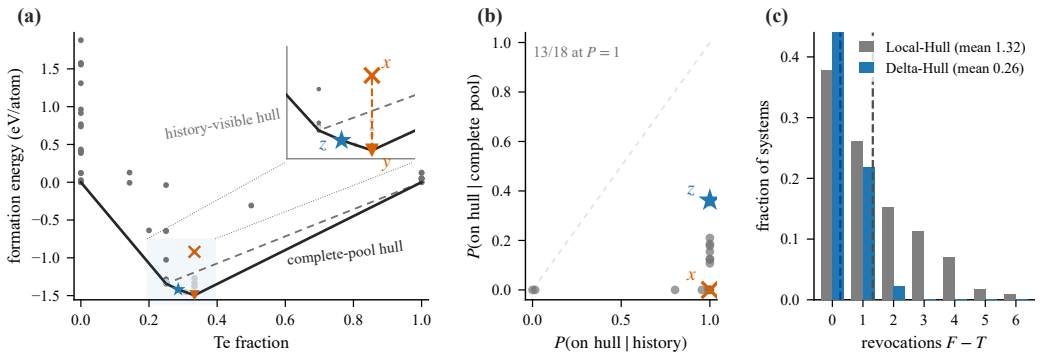


Figure 1: **Pool-dependent utility changes what a query certifies.** (a) Real O–Te state: all real candidates (light points), the history-visible hull (dashed), and the complete-pool hull (solid). The Local-Hull selection x ("mp-1077823") is revoked from the complete-pool hull by the unqueried same-composition competitor y ("mp-2125"), whereas the terminal-utility selection z ("mp-12177") remains a complete-pool hull vertex; the inset enlarges the local geometry. The observation $O_x = E_x$ is revealed at query time, but the terminal label Y_x cannot be determined from E_x alone. (b) Per-candidate posterior scores at the same state: history-visible membership probability versus terminal complete-pool membership probability. Thirteen of eighteen candidates sit at $P = 1$ under the history-visible event, which does not distinguish them, and x has terminal probability 0. (c) Distribution over the 230 benchmark systems of $F - T$, selected-history confirmations revoked at terminal evaluation: revocations are systematic under Local-Hull (mean 1.32) and rare under Delta-Hull (mean 0.26).

arbitrarily suboptimal in general, while weak posterior movement gives a finite-horizon loss bound. Third, we study these effects on 230 MatPES systems using controlled comparisons of the terminal event, acquisition objective, scoring rule, and planning depth. The largest empirical change comes from evaluating stability against the completed candidate pool rather than the history-visible hull. Once that event is fixed, a deterministic complete-pool score recovers essentially the same $B = 6$ utility as posterior probability integration, and the rollout yields only a small terminal-utility gain despite frequent action changes.

2 Active Search under Pool-Dependent Utility

2.1 Feedback and Terminal Utility

General formulation. Let C be a finite candidate pool and let Z_C denote the latent complete state, with Z_x its candidate-level coordinate. Querying $x \in C$ returns an observation, while the latent state determines the utility label:

$$O_x = o_x(Z_x), \qquad Y_x = g_x(Z_C), \qquad R(S, Z_C) = \sum_{x \in S} Y_x. \tag{1}$$

Pool dependence means that Y_x can change even when Z_x is held fixed: changing only the latent state of other candidates can alter the reward assigned to x. This is stronger than noisy or partial local feedback. We use the deterministic notation $O_x = o_x(Z_x)$ for clarity; stochastic observations can be represented with an observation kernel. A policy observes the query history, while unqueried coordinates and their terminal labels remain latent.

Top-K, Pareto-front, and latent-graph adjudicators provide further examples of the abstraction; the experiments study the materials-hull instantiation.

Relation to active search and BAX. Standard active search attaches utility directly to the queried label, whereas BAX methods allocate evaluations to infer an algorithm-defined global output. Pool-dependent active search has a different utility object: querying x reveals a local observation O_x, but reward is attached to x through a terminal label $Y_x = g_x(Z_C)$ that depends on the shared latent state. Consequently, two policies can share the same posterior over Z_C and the same global adjudicator

while choosing different queries, because one values reconstruction of the global object and the other values the terminal utility of the selected subset. This distinction drives both the one-step rule in Eq. (6) and the lookahead trade-off in Proposition 1.

2.2 Finite-pool hull instantiation and scope

For one exact chemical system, let C be the visible finite candidate pool, $S_t \subseteq C$ the candidates queried before decision t, and $U_t = C \setminus S_t$ the available actions. Let H_t contain the visible source energies and features, the immutable record of energies revealed for S_t, the reference phases, and the resulting history-consistent hull. Thus S_t is recoverable from H_t; we write both arguments below to make the terminal objective explicit. The selected candidate is recorded before the oracle reveals its target energy; every revealed energy is appended to the archive and remains available in later rounds.

For a complete target-energy vector $e = (e_x : x \in C)$, define the terminal complete-pool label

$$Y_x(e) = \mathbf{1}\{x \in \mathrm{Stable}(\mathcal{R}^0 \cup \{(c_z, e_z) : z \in C\})\}, \tag{2}$$

where $\mathcal{R}^0$ is the fixed reference phase set. The oracle reveals e_x, not Y_x. For a queried set S, terminal utility is

$$R(S, e) = \sum_{x \in S} Y_x(e). \tag{3}$$

In the MatPES study, C is the finite candidate pool for one chemical system and $\mathcal{R}^0$ contains the elemental reference phases. Queries have equal cost, and terminal utility counts selected candidates that lie on the completed target-fidelity hull. All policies use the same candidate pool, reference set, hull tolerance, and posterior construction. Appendix A.2 gives the full evaluation protocol.

2.3 One-step Bayes action

For a history h, let $S(h)$ be the queried items, $U(h) = C \setminus S(h)$ the legal actions, and $h \oplus (x, O_x)$ the history after observing O_x. The exact finite-horizon problem has the terminal boundary

$$V_0(h) = \mathbb{E}_{Z_C|h}[R(S(h), Z_C)]\,, \tag{4}$$

$$V_b(h) = \max_{x \in U(h)} \mathbb{E}_{O_x|h}[V_{b-1}(h \oplus (x, O_x))]\,, \quad b \geq 1. \tag{5}$$

The terminal expectation is over the complete latent state; observed coordinates are fixed under the conditional distribution. Complete-pool utility evaluation enters through this terminal world, with utility realized at the terminal boundary. Exact nonmyopic planning expands an action–observation tree through this recursion and is intractable at material-pool scale. With one query remaining and equal query costs, previously selected-item terms are action-independent, so

$$Q_1(x \mid h) = \underbrace{\sum_{s \in S(h)} \Pr(Y_s = 1 \mid h)}_{\text{independent of } x} + p_h(x), \qquad \pi_1^\star(h) = \operatorname*{lex\,arg\,max}_{x \in U(h)} p_h(x), \tag{6}$$

where $p_h(x) = \Pr(Y_x = 1 \mid h)$. All action ties use the same deterministic candidate-index rule. We call this terminal-utility greedy policy *Delta-Hull* in the materials experiments.

3 Conditions for Lookahead Value

For the theory, we suppress the time index and write h for a generic observable history H_t. Throughout, the generic latent state Z_C corresponds to the complete target-energy vector E, the observation O_x to a revealed energy e_x, and the terminal label Y_x to hull membership determined by the threshold margin m_x in Eq. (10). Write $p_h(x) = \Pr(Y_x = 1 \mid h)$ for the posterior utility-label probability. Once the terminal event is specified, the planning question is whether an observation changes future utility decisions enough to justify leaving the current best action.

Exact two-step decision value. At a history with two queries remaining, suppress the contribution of previously selected items. For a first action x, let $h' = h \oplus (x, O_x)$ and define

$$D_h(x) = \mathbb{E}_{O_x|h}\left[\max_{y \in U(h)\setminus\{x\}} p_{h'}(y)\right] - \max_{y \in U(h)\setminus\{x\}} p_h(y). \tag{7}$$

By Jensen's inequality, $D_h(x) \geq 0$, and the exact two-step value is

$$Q_2^\star(x \mid h) = p_h(x) + \max_{y \neq x} p_h(y) + D_h(x). \tag{8}$$

Order candidates by $p_h(x)$ with the policy's deterministic tie-breaking rule, so $p_1 \geq p_2 \geq \cdots \geq p_m$, and write $D_j = D_h(x_{(j)})$.

Proposition 1 (Exact two-step gain over greedy)**.** *The exact gain over the current greedy action is*

$$H_2(h) = \max\left\{0,\ D_2 - D_1,\ \max_{j \geq 3}\left[D_j - D_1 - (p_2 - p_j)\right]\right\}. \tag{9}$$

Thus a lower-ranked candidate must overcome the posterior utility gap $p_2 - p_j$ in addition to its decision value. The proof is in Appendix A.14.

Proof intuition. The comparison is bookkeeping over the ordered candidates. The greedy action $x_{(1)}$ and the runner-up $x_{(2)}$ collect the same immediate top-two sum $p_1 + p_2$, so only the decision values $D_2 - D_1$ separate them. A lower-ranked candidate $x_{(j)}$, $j \geq 3$, instead collects only $p_j + p_1$ up front; its observation must supply at least the utility gap $p_2 - p_j$ of extra decision value before it can overtake the greedy action.

Under additive selected-item terminal utility and equal query costs, departure from the greedy action separates exactly into an observation-induced decision-value advantage and the posterior-utility opportunity cost.

Weak posterior movement. For candidate x, let $\ell_{-x}(c_x; e_{-x})$ be the lowest competing-hull energy attainable at c_x. With hull tolerance τ, define

$$m_x(e) = \ell_{-x}(c_x; e_{-x}) - e_x + \tau, \qquad Y_x(e) = \mathbf{1}\{m_x(e) \geq 0\}. \tag{10}$$

Lemma 1 (Hull-margin stability)**.** *If $\|e - e'\|_\infty \leq \delta$, then $|m_x(e) - m_x(e')| \leq 2\delta$; consequently, $|m_x(e)| > 2\delta$ implies $Y_x(e) = Y_x(e')$.*

Call ψ a reachable legal observation continuation from h if it has positive posterior probability and contains at most $n-1$ further observations. For a candidate legal after ψ, write $p_{h\cup\psi}(x) = \Pr(Y_x = 1 \mid h, \psi)$.

Proposition 2 (Greedy safety under weak posterior movement)**.** *Suppose that for every reachable legal observation continuation ψ from h and every candidate that remains legal after ψ,*

$$|p_{h\cup\psi}(x) - p_h(x)| \leq \epsilon. \tag{11}$$

Then with n queries remaining,

$$V_n^\star(h) - V_n^{\mathrm{greedy}}(h) \leq \min\{n, 2n\epsilon\}. \tag{12}$$

Corollary 1 (A hull-specific bound for weak posterior movement)**.** *Suppose the current and continuation posteriors admit a coupling (E, E_ψ) with $\Pr(\|E - E_\psi\|_\infty > \delta) \leq \alpha$. Then*

$$|p_{h\cup\psi}(x) - p_h(x)| \leq \alpha + \Pr_{E \sim P_h}(|m_x(E)| \leq 2\delta). \tag{13}$$

If the right-hand side is uniformly bounded by ϵ_h over legal continuations and candidates, then

$$V_n^\star(h) - V_n^{\mathrm{greedy}}(h) \leq \min\{n, 2n\epsilon_h\}. \tag{14}$$

The result is a sufficient condition: greedy search is safe when posterior continuations move the energy world little and little posterior mass lies near a membership boundary. The bound uses a general posterior coupling.

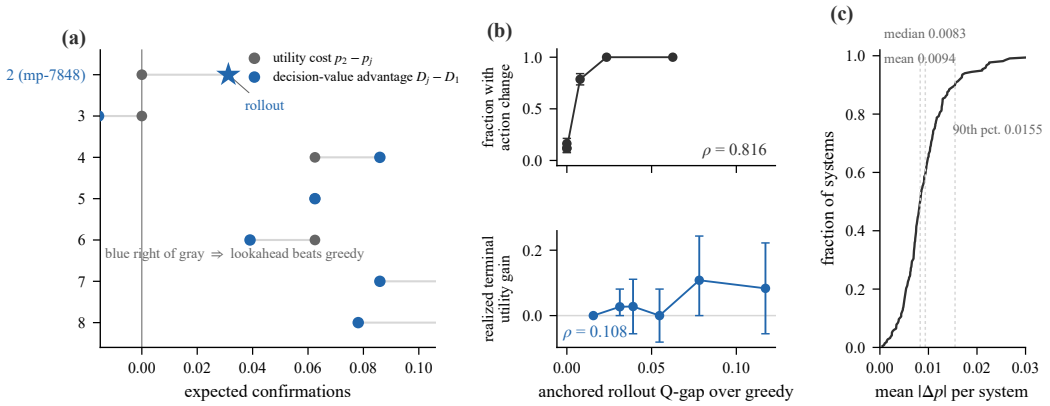


Figure 2: **Mechanisms behind limited lookahead value.** (a) At a representative state, the rollout advantage $D_j - D_1$ is compared with the utility cost $p_2 - p_j$ of leaving the current best action. (b) Across evaluated states, action disagreement rises with the rollout Q-gap, whereas the realized terminal-utility gain remains small. (c) The empirical distribution of posterior membership movement across systems, with its mean, median, and 90th percentile.

Proposition 3 (Repeated greedy has no uniform approximation guarantee)**.** *For every $\eta > 0$, there is a finite binary-composition convex-hull search instance with exact self-energy observations, equal query costs, and budget two for which*

$$V_2^{\text{greedy}} < \eta V_2^{\star}.$$

The construction uses a low-immediate-utility probe whose observation identifies which future candidate survives global competition; complete proofs are in Appendix A.14.

Together, the results characterize both sides of the planning problem. Pool coupling can make repeated greedy search fail arbitrarily badly, whereas weak posterior movement limits the value of adaptation. Proposition 1 identifies the corresponding local trade-off at horizon two: information from a lower-ranked query matters only when it compensates for the utility gap paid to obtain it. Section 5 locates the MatPES benchmark within these regimes.

4 Policies and Controls

The experiments compare the terminal membership event, acquisition objective, score functional, and evaluated two-step planning. Each controlled contrast holds the posterior, candidate pool, hull evaluator, tie-breaking rule, and query budget fixed. The joint PBE–r2SCAN energy posterior is fit on systems disjoint from each query system and conditioned on queried energies; descriptor, posterior, and hull details are in Appendix A.2. Table A.2 summarizes the matched controls across terminal-event definition, acquisition objective, score functional, and planning depth.

Terminal-label and score controls. Local-Hull Probability uses the same posterior membership statistic, posterior worlds, Monte Carlo count, numerical tolerance, and deterministic tie-breaking as Delta-Hull, but evaluates candidates against the history-visible hull. Target Margin ranks candidates by posterior-mean margin against that same hull. In the materials instantiation, posterior terminal utility is complete-pool hull membership probability, so the one-step action becomes Delta-Hull:

$$\pi_{\Delta}(H_t) = \operatorname{lex\,arg\,max}_{x \in U_t} \Pr(Y_x(E) = 1 \mid H_t)\,. \tag{15}$$

Here $\operatorname{lex\,arg\,max}$ denotes maximization with ties broken by pre-assigned candidate index; the name Delta-Hull refers to this candidate-level complete-pool hull-membership decision. At $B > 1$ we use repeated Delta-Hull as the posterior terminal-utility greedy policy. Complete-Pool Mean Margin keeps the same complete-pool evaluation with a deterministic posterior-mean plug-in score; its explicit score is given in the Appendix.

Algorithm 1 Delta-Hull execution loop

Require: pool C, references $\mathcal{R}^0$, budget B, cross-fitted posterior
1: archive $\leftarrow$ source energies; $S \leftarrow \emptyset$
2: **for** $t = 1, \ldots, B$ **do**
3: $p_{H_t}(x) \leftarrow \Pr(Y_x(E) = 1 \mid H_t)$ from posterior worlds conditioned on the archive, $x \in U_t$
4: $x_t \leftarrow \text{lex arg max}_{x \in U_t} p_{H_t}(x)$ ▷ deterministic tie-break
5: reveal e_{x_t}; append to archive; $S \leftarrow S \cup \{x_t\}$
6: **end for**
7: **return** $T(S) = \sum_{x \in S} Y_x$ on the completed pool

Reconstruction and planning controls. Matched CAL Hull Entropy and PS-BAX Target-Set are reconstruction-objective controls, not assessments of these methods on their native tasks: with the posterior, pool, and terminal evaluator fixed, they isolate the utility cost of optimizing a reconstructed global object or target set instead of selected-item terminal utility. Matched CAL Hull Entropy uses the same conditioned posterior, candidate pool, complete-pool evaluator, and deterministic tie-breaking while reconstructing the posterior whole-hull vector $\mathbf{L}$; it is a matched CAL-style reconstruction control (Novick et al., 2024). PS-BAX Target-Set draws one posterior world, forms its stable set, and ranks candidates there by posterior standard deviation (Cheng et al., 2024). The evaluated two-step rollout is anchored at repeated Delta-Hull and retains the complete-pool utility objective. All reported planning experiments use horizon two. Estimators, fantasy sampling, and execution details for these controls appear in Appendices A.3 and A.5.

Evaluation. We evaluate terminal selected-item utility

$$T(S_B) = \sum_{x \in S_B} \mathbf{1}\{x \in \text{Stable}(\mathcal{A}_{\text{full}})\}. \tag{16}$$

Query-time and selected-history diagnostics are reported in Appendix A.8.

5 Experiments on the MatPES finite-pool benchmark

5.1 Experimental setting

We evaluate the policies on 230 MatPES 2025.2 chemical systems (Kaplan et al., 2025; 2026a;b) with paired PBE (Perdew et al., 1996) and r2SCAN (Furness et al., 2020) calculations. PBE-derived quantities are available to the policy, and each query reveals the corresponding r2SCAN energy. All comparisons use the same candidate pools, source features, posterior construction, hull evaluator, query budget, and tie-breaking. Terminal utility is the number of selected candidates that remain on the completed r2SCAN hull after $B \leq 6$ queries. We interpret $B \leq 6$ as a low-budget screening regime: candidate pools contain a median of 24 candidates (IQR 19–41, maximum 64; a median pool has 4 complete-pool hull members), so six queries examine 25% of a median pool. The theory allows arbitrary finite horizons, but these results do not imply that planning stays negligible at larger budgets or pools.

The energy posterior is cross-fitted by chemical system. It combines a system-balanced ridge transport mean, a system random intercept, and a local Matérn-5/2 residual Gaussian process (Rasmussen & Williams, 2006). Of the 230 systems, 217 lie within the model's element support; the remaining 13 use the same source-based fallback for every policy. A separate 94-system panel is used for the secondary terminal-event comparison. Full posterior and execution details are given in Appendix A.2.

For system i and matched policies π, π', the area under the utility–budget curve is

$$A_i(\pi) = \sum_{b=1}^{5} \frac{T_{i,b}(\pi) + T_{i,b+1}(\pi)}{2}, \qquad \Delta A_i = A_i(\pi) - A_i(\pi'), \qquad \Delta A = \frac{1}{N} \sum_{i=1}^{N} \Delta A_i. \tag{17}$$

We refer to A as the budget-curve area and use exact chemical system as the paired unit. Figure 4 reports paired 95% intervals.

5.2 Posterior calibration and baseline checks

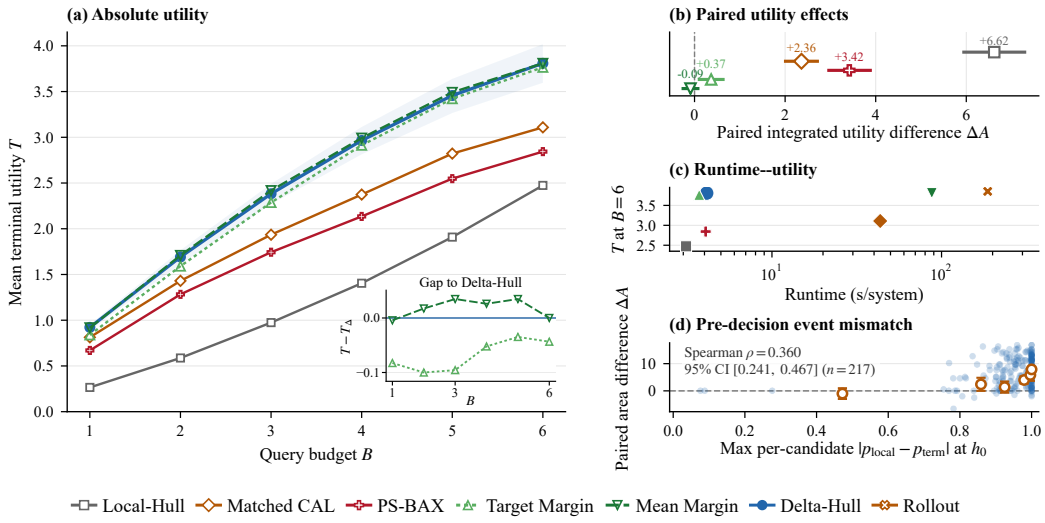


Figure 4: **Controlled comparisons on MatPES.** (a) Mean terminal utility across query budgets; the shaded interval is shown for Delta-Hull. (b) Paired budget-curve-area differences, reported as Delta-Hull minus each comparator. (c) Terminal utility at $B = 6$ versus measured per-system runtime. (d) Pre-decision event mismatch $M_{\max} = \max_x |p_{\text{local}}(x) - p_{\text{term}}(x)|$ at h_0 versus the paired Delta-Hull–Local-Hull area difference; orange markers show fixed-bin means with 95% confidence intervals. Runtime reflects the evaluated implementations.

Calibration is evaluated over all legal candidate–state predictions from the 217 modeled systems. The complete-pool membership probabilities obtain Brier score 0.0598, negative log loss 0.2095, and ROC–AUC 0.9411; the selected action is harder, with Brier score 0.1269 and ROC–AUC 0.9044. Figure 3 shows the corresponding reliability diagram against the perfect-calibration reference (dashed), with the marginal candidate count per probability bin. Because Delta-Hull ranks posterior terminal probabilities, calibration is part of validating the decision statistic; we report it together with proper probabilistic scores (Gneiting & Raftery, 2007; Popordanoska et al., 2024). As a scale reference, the three random permutations obtain areas 3.17–3.51 and Ridge Uncertainty 5.53, versus 12.88 for Delta-Hull.

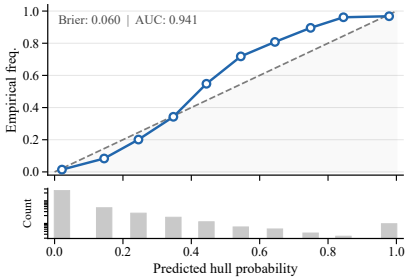


Figure 3: **Posterior terminal-label calibration.** Reliability diagram (top) and marginal candidate count across probability bins (bottom).

5.3 Terminal-label alignment before planning

Changing only the membership event has the largest effect.
Local-Hull Probability evaluates the posterior against the history-visible hull, whereas Delta-Hull evaluates the same posterior worlds against the completed candidate pool. At $B = 1$, before a future observation can affect another action, this change increases mean terminal utility by 0.657 confirmations. Across $B = 1, \ldots, 6$, it increases paired budget-curve area by 6.617 confirmations–budget per system (95% interval $[5.946, 7.289]$). The secondary 94-system panel gives a similar area difference of 6.782. Figure 1 illustrates the candidate-level mechanism. The association is also visible before any query: systems whose history-visible and complete-pool membership probabilities disagree more at the initial state tend to show larger paired utility differences (Figure 4(d); Spearman $\rho = 0.360$, bootstrap 95% CI $[0.241, 0.467]$ for the maximal per-candidate probability gap, replicated among the 91 transport-supported systems of the secondary 94-system panel with $\rho = 0.378$ $[0.159, 0.561]$).

Table 1: Main paired contrasts on the 230-system MatPES panel. Every utility contrast is Delta-Hull minus the listed comparator; ΔA is the trapezoidal budget-curve area. Intervals are paired system-level 95% intervals. Runtime reports measured seconds per system for Delta-Hull and the comparator; the planning row is evaluated at $B = 6$ only.

Decision layer	Comparator	$\Delta T_{B=1}$	$\Delta T_{B=6}$	ΔA	Paired 95% CI for ΔA	Runtime (Delta / comp.)
Terminal label Acquisition	Local-Hull Probability	$+0.657$	$+1.335$	$+6.617$	$[+5.946, +7.289]$	$4.12/3.12$
objective Acquisition	Matched CAL Hull Entropy	$+0.100$	$+0.700$	$+2.361$	$[+2.009, +2.712]$	$4.12/43.48$
objective	PS-BAX Target-Set control	$+0.252$	$+0.965$	$+3.417$	$[+2.965, +3.880]$	$4.12/4.07$
Practical score	Target Margin Complete-Pool Mean	$+0.074$	$+0.043$	$+0.367$	$[+0.107, +0.629]$	$4.12/3.73$
Score functional	Margin	-0.004	$+0.000$	-0.089	$[-0.252, +0.074]$	$4.12/87.61$
Planning depth	Anchored two-step rollout	—	-0.039	—	—	$3.99/187.13$

5.4 Decision layers after alignment

Once the terminal event is fixed, the scoring rule matters much less on this benchmark. Complete-Pool Mean Margin and Delta-Hull both reach $T = 3.809$ at $B = 6$, and their integrated difference is -0.089 (95% interval $[-0.252, 0.074]$). Thus, most of the empirical gain comes from evaluating the complete-pool terminal event rather than from Monte Carlo integration of its membership probability. The similarity in utility does not arise from identical actions: the two rules agree on the first query in 72.6% of systems and on only 54.5% of decisions across all six rounds; complete trajectories coincide in $38/230$ systems. After terminal-event alignment, substantial action-level differences therefore need not translate into utility differences on this benchmark.

Under the selected-item terminal objective, substituting reconstruction-oriented acquisitions carries a material utility cost: Matched CAL Hull Entropy has 2.361 lower budget-curve area and the PS-BAX Target-Set control 3.417 lower area than the aligned Delta-Hull rule. These contrasts quantify objective mismatch under a common terminal evaluator, not the quality of CAL or BAX on their native reconstruction objectives. Target Margin is closer, with a 0.367 area difference concentrated at early budgets.

5.5 Empirical value of lookahead and posterior movement

At $B = 6$, the anchored rollout chooses a different action from repeated Delta-Hull in 211 of 230 systems but increases mean terminal utility by only 0.039 confirmations per system, while runtime rises from 3.99 to 187.13 seconds per system, a $46.9\times$ increase. We therefore treat the rollout as a diagnostic of two-step decision value rather than as the recommended policy. It evaluates 128 common posterior worlds per decision state, shared across first-action branches, with 16 conditional continuation worlds. Terminal utility on each realized trajectory is evaluated exactly from completed-pool labels; Monte Carlo enters only through action selection. A separate MC-count diagnostic shows that the rollout argmax remains estimator-sensitive when the leading Q-values are nearly tied (Appendix A.10), so the planning comparison is interpreted as the behavior of this evaluated finite-MC planner. Two independent Monte Carlo seeds produce similarly small $B = 6$ gains of 0.004 and 0.026 confirmations. In the primary run, the rollout gain remains between 0.013 and 0.039 confirmations per system from $B = 2$ to 6, while action disagreement rises from $104/230$ to $211/230$ systems; intermediate-budget contrasts vary more across seeds (Appendix A.9). Observed membership probabilities also move little along realized trajectories: the mean per-transition change in membership probability is 0.0094, and the top action is preserved across 84.8% of consecutive states, a pattern consistent with—but not sufficient to establish—the weak-movement condition of Proposition 2. Appendix A.13 reports a gap-triggered compute diagnostic.

6 Related work

Active search maximizes discoveries among queried items, with nonmyopic methods such as ENS accounting for how observations affect future queries (Garnett et al., 2012; Jiang et al., 2017). Correlated and multifidelity variants generalize information sources, costs, and dependence structures

(Jiang et al., 2019; Nguyen & Garnett, 2023; Nguyen et al., 2021). Pool-dependent search instead decouples feedback from reward: querying x reveals O_x, while its utility label Y_x depends on the latent shared pool. Bayesian probing also couples observations and rewards under structured uncertainty (Gupta et al., 2026), but here reward is attached to queried items through terminal adjudication.

Bayesian experimental design and decision-aware active learning choose queries for an explicit downstream utility or loss (Rainforth et al., 2024; Bickford Smith et al., 2023; Huang et al., 2026). Pool-dependent active search is a particular instance in which the query observation and the event defining its eventual reward are distinct random variables; information reduction therefore need not maximize selected-item terminal utility.

Adaptive submodularity justifies greedy policies under adaptive diminishing returns, with extensions based on adaptive submodularity ratios (Golovin & Krause, 2011; Fujii & Sakaue, 2019; Weitzman & Sabato, 2024). Pool-dependent utility need not satisfy this property (Appendix A.14); our weak-movement result instead identifies a sufficient regime for controlling greedy loss. Related global-object methods include BAX, CAL, and BALPI, which target algorithm-defined outputs, reconstructed hulls, or phase-stability regions (Neiswanger et al., 2021; Cheng et al., 2024; Novick et al., 2024; Fan et al., 2026), as well as Bayesian convex-hull search (Wen et al., 2024), GP level-set estimation (Gotovos et al., 2013; Bogunovic et al., 2016), and entropy search (Hennig & Schuler, 2012; Hernández-Lobato et al., 2014). These methods may share the same posterior while optimizing a different terminal quantity, motivating our matched acquisition controls.

Materials active learning spans phase-diagram construction, autonomous discovery, and uncertainty-aware workflows (Ament et al., 2021; Nair & Foppa, 2026). Recent Bayesian methods explicitly target convex-hull or phase-stability reconstruction (Novick et al., 2024; Fan et al., 2026), while stability screening is inherently relational because a candidate's status depends on competing phases (Jain et al., 2013; Bartel et al., 2020). Our experiments focus instead on the terminal utility of queried structures after complete-pool adjudication.

7 Conclusion

We introduced active search with pool-dependent utility, where a query reveals a local observation but the queried item's reward is determined only after a shared latent pool is resolved. The one-step Bayes action ranks candidates by posterior terminal-success probability, while the value of lookahead depends on whether an observation can change future decisions enough to offset the utility lost by leaving the current best action. Repeated greedy search can be arbitrarily suboptimal in general, but its loss is bounded when posterior terminal probabilities move little after observations.

On the MatPES benchmark, terminal-event specification dominates the other decision layers: switching from history-visible to complete-pool membership produces the largest utility change, whereas posterior integration and the evaluated two-step rollout add comparatively little once the terminal objective is aligned. The broader implication is a decision hierarchy rather than a preference for a particular planner: first specify the terminal utility, then choose a score aligned with that utility, and increase planning depth only when observations can materially change future terminal decisions.

AI use statement

In this work, we used generative AI tools solely to polish the writing, as well as to assist with code implementation and script development for data processing and experimental workflows. We have not used generative AI tools for primary scientific ideation or for generating experimental data.

All AI-assisted contributions were thoroughly reviewed by the authors: all AI-polished text was carefully examined and edited to preserve original intent and ensure factual and technical accuracy, and all AI-assisted code was manually verified, debugged, and tested for correctness. The human authors take full responsibility for the final content of this work, including all claims, text, and accompanying artifacts.

REFERENCES

Sebastian Ament, Maximilian Amsler, Duncan R. Sutherland, Ming-Chiang Chang, Dan Guevarra, Aine B. Connolly, John M. Gregoire, Michael O. Thompson, Carla P. Gomes, and R. Bruce van Dover. Autonomous materials synthesis via hierarchical active learning of nonequilibrium phase diagrams. *Science Advances*, 7 (51):eabg4930, 2021. doi: 10.1126/sciadv.abg4930.

Christopher J. Bartel, Amalie Trewartha, Qi Wang, Alexander Dunn, Anubhav Jain, and Gerbrand Ceder. A critical examination of compound stability predictions from machine-learned formation energies. *npj Computational Materials*, 6:97, 2020. doi: 10.1038/s41524-020-00362-y.

Freddie Bickford Smith, Andreas Kirsch, Sebastian Farquhar, Yarin Gal, Adam Foster, and Tom Rainforth. Prediction-oriented bayesian active learning. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 7331–7348. PMLR, 2023. URL https://proceedings.mlr.press/v206/bickfordsmith23a.html.

Ilija Bogunovic, Jonathan Scarlett, Andreas Krause, and Volkan Cevher. Truncated variance reduction: A unified approach to bayesian optimization and level-set estimation. In *Advances in Neural Information Processing Systems*, volume 29, pp. 1507–1515, 2016.

Chu Xin Cheng, Raul Astudillo, Thomas Desautels, and Yisong Yue. Practical bayesian algorithm execution via posterior sampling. In *Advances in Neural Information Processing Systems*, volume 37, pp. 135186–135207, 2024. doi: 10.52202/079017-4296.

Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J. Bartel, and Gerbrand Ceder. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023. doi: 10.1038/s42256-023-00716-3.

Mingzhou Fan, Yucheng Wang, Guillermo Vazquez, Ruida Zhou, Ibrahim Karaman, Raymundo Arróyave, and Xiaoning Qian. Bayesian active learning to accelerate high throughput phase diagram exploration. *Digital Discovery*, 5(6):2478–2490, 2026. doi: 10.1039/D5DD00459D. URL https://doi.org/10.1039/D5DD00459D.

Ichiro Fujii and Shinsaku Sakaue. Beyond adaptive submodularity: Approximation guarantees of greedy policy with adaptive submodularity ratio. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, 2019. URL https://proceedings.mlr.press/v97/fujii19a.html.

James W. Furness, Aaron D. Kaplan, Jinliang Ning, John P. Perdew, and Jianwei Sun. Accurate and numerically efficient r2SCAN meta-generalized gradient approximation. *The Journal of Physical Chemistry Letters*, 11 (19):8208–8215, 2020. doi: 10.1021/acs.jpclett.0c02405.

Roman Garnett, Yamuna Krishnamurthy, Xuehan Xiong, Jeff Schneider, and Richard Mann. Bayesian optimal active search and surveying. In *Proceedings of the 29th International Conference on Machine Learning*, 2012.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011.

Alkis Gotovos, Nathalie Casati, Gregory Hitz, and Andreas Krause. Active learning for level set estimation. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, pp. 1344–1350, 2013.

Anupam Gupta, Benjamin Moseley, and Rudy Zhou. Bayesian probing on graphs. In *Integer Programming and Combinatorial Optimization*, pp. 410–425, 2026. URL https://doi.org/10.1007/978-3-032-28691-8_27.

Philipp Hennig and Christian J. Schuler. Entropy search for information-efficient global optimization. *Journal of Machine Learning Research*, 13:1809–1837, 2012.

José Miguel Hernández-Lobato, Matthew W. Hoffman, and Zoubin Ghahramani. Predictive entropy search for efficient global optimization of black-box functions. In *Advances in Neural Information Processing Systems*, volume 27, pp. 918–926, 2014.

Zhuoyue Huang, Freddie Bickford Smith, and Tom Rainforth. Loss-driven bayesian active learning. In *Proceedings of The 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 5140–5148. PMLR, 2026. URL https://proceedings.mlr.press/v300/huang26b.html.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1): 011002, 2013. doi: 10.1063/1.4812323.

Shali Jiang, Gustavo Malkomes, Geoff Converse, Alyssa Shofner, Benjamin Moseley, and Roman Garnett. Efficient nonmyopic active search. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1714–1723. PMLR, 2017.

Shali Jiang, Roman Garnett, and Benjamin Moseley. Cost effective active search. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/df0e09d6f25a15a815563df9827f48fa-Abstract.html.

Aaron D. Kaplan, Runze Liu, Ji Qi, Tsz Wai Ko, Bowen Deng, Janosh Riebesell, Gerbrand Ceder, Kristin A. Persson, and Shyue Ping Ong. A foundational potential energy surface dataset for materials. *arXiv preprint arXiv:2503.04070*, 2025.

Aaron D. Kaplan, Runze Liu, Ji Qi, Tsz Wai Ko, Bowen Deng, Janosh Riebesell, Gerbrand Ceder, Kristin A. Persson, and Shyue Ping Ong. Matpes-pbe-2025.2. ColabFit Exchange, 2026a. URL https://doi.org/10.60732/99baa6c7.

Aaron D. Kaplan, Runze Liu, Ji Qi, Tsz Wai Ko, Bowen Deng, Janosh Riebesell, Gerbrand Ceder, Kristin A. Persson, and Shyue Ping Ong. MatPES-R2SCAN-2025.2. ColabFit Exchange, 2026b. URL https://doi.org/10.60732/64a29d9d.

Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, 2023. doi: 10.1038/s41586-023-06735-9.

Akhil S. Nair and Lucas Foppa. A critical examination of active learning workflows in materials science. *Digital Discovery*, 5(6):2366–2382, 2026. doi: 10.1039/D6DD00081A. URL https://doi.org/10.1039/D6DD00081A.

Willie Neiswanger, Ke Alexander Wang, and Stefano Ermon. Bayesian algorithm execution: Estimating computable properties of black-box functions using mutual information. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8005–8015. PMLR, 2021.

Quan Nguyen and Roman Garnett. Nonmyopic multiclass active search with diminishing returns for diverse discovery. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 5231–5249. PMLR, 2023. URL https://proceedings.mlr.press/v206/nguyen23d.html.

Quan Nguyen, Arghavan Modiri, and Roman Garnett. Nonmyopic multifidelity active search. In *Proceedings of the 38th International Conference on Machine Learning*, 2021.

Andrew Novick, Diana Cai, Quan Nguyen, Roman Garnett, Ryan Adams, and Eric Toberer. Probabilistic prediction of material stability: Integrating convex hulls into active learning. *Materials Horizons*, 11(21): 5381–5393, 2024. doi: 10.1039/D4MH00432A. URL https://doi.org/10.1039/D4MH00432A.

John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, 1996. doi: 10.1103/PhysRevLett.77.3865.

Teodora Popordanoska, Sebastian Gregor Gruber, Aleksei Tiulpin, Florian Buettner, and Matthew B. Blaschko. Consistent and asymptotically unbiased estimation of proper calibration errors. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 3466–3474. PMLR, 2024. URL https://proceedings.mlr.press/v238/popordanoska24a.html.

Tom Rainforth, Adam Foster, Desi R. Ivanova, and Freddie Bickford Smith. Modern bayesian experimental design. *Statistical Science*, 39(1):100–114, 2024. doi: 10.1214/23-STS915. URL https://doi.org/10.1214/23-STS915.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, MA, 2006. ISBN 978-0-262-18253-9. URL https://gaussianprocess.org/gpml/.

Janosh Riebesell, Rhys E. A. Goodall, Philipp Benner, Yuan Chiang, Bowen Deng, Gerbrand Ceder, Mark Asta, Alpha A. Lee, Anubhav Jain, and Kristin A. Persson. A framework to evaluate machine learning crystal stability predictions. *Nature Machine Intelligence*, 7:836–847, 2025. doi: 10.1038/s42256-025-01055-1.

Shlomi Weitzman and Sivan Sabato. Adaptive combinatorial maximization: Beyond approximate greedy policies. In *Proceedings of The 35th International Conference on Algorithmic Learning Theory*, volume 237 of *Proceedings of Machine Learning Research*, pp. 1188–1207. PMLR, 2024. URL https://proceedings.mlr.press/v237/weitzman24a.html.

Dongsheng Wen, Victoria Tucker, and Michael S. Titus. Bayesian optimization acquisition functions for accelerated search of cluster expansion convex hull of multi-component alloys. *npj Computational Materials*, 10:210, 2024. doi: 10.1038/s41524-024-01391-7.

A Additional results and proofs

A.1 Notation and experimental organization

Table A.1 collects the notation used by the general formulation and its materials instantiation. Figure A.1 summarizes the distinction between query-time feedback and terminal utility: in the pool-dependent setting, a query reveals a local observation, whereas the reward of the queried item is resolved only after adjudication against the shared latent state.

The MatPES experiments instantiate this distinction through matched controls that vary one decision layer at a time while retaining a common evaluator. Table A.2 summarizes the varied component and the quantities held fixed in each comparison. Detailed posterior construction, execution, and evaluation protocols follow in Appendix A.2.

Table A.1: Notation used by the general formulation and its materials instantiation. The latent-state notation is general; E_C is the corresponding target-energy vector in the materials experiment.

Symbol	General role	Materials instantiation
Z_C / E_C	complete latent state	complete target-energy vector
O_x	immediate query observation	revealed target energy for x
Y_x	delayed complete-pool utility label	final-hull membership of x
$R(S, Z_C)$ / $T(S_B)$	selected-item terminal utility	complete-pool confirmation count
H_t, U_t	observable history and legal actions	revealed energies and remaining candidates

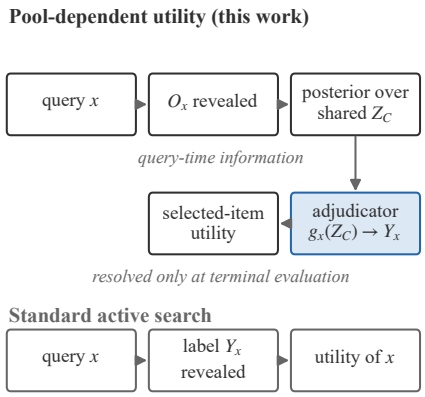


Figure A.1: **The abstraction.** (top) Pool-dependent utility: a query reveals O_x, while the reward of x is attached only after the adjudicator resolves the shared pool state. (bottom) Standard active search: the queried label is itself the reward.

Table A.2: Detailed matched contrasts used in the MatPES study. The main text reports the resulting utility comparisons; this table records the varied decision component and the shared evaluation controls.

Decision component	Reference	Matched control	Shared controls
Utility evaluation	Delta-Hull	Local-Hull Probability	posterior, pool, score, budget
Acquisition objective	Delta-Hull	Matched CAL Hull Entropy	posterior, pool, evaluator, budget
Acquisition objective	Delta-Hull	PS-BAX target set	posterior, pool, evaluator, budget
Complete-pool score	Delta-Hull	Complete-Pool Mean Margin	utility, pool, budget
Planning depth	repeated Delta-Hull	anchored rollout	posterior, utility, pool, budget

A.2 Posterior and evaluation protocol

All primary analyses use five-fold cross-fitting by chemical system; the 94-system secondary panel is held out from its transport-fit set.

Posterior and margin-policy definitions. The PBE–r2SCAN posterior is a source-to-target transport model with a system-balanced ridge mean, a system random intercept, and a local Matérn-$5/2$ residual GP. It is fit on systems disjoint from the query systems; conditioning on observed queries yields a joint Gaussian over remaining candidates. Let $\widehat{\ell}_t^{-x}(c_x)$ be the current competing-hull energy at composition c_x, excluding x and using fixed references plus legal reveals, and let $\mu_t(x) = \mathbb{E}[E_T(x) \mid H_t]$. The two margin rules are

$$\pi_{\mathrm{src}}(H_t) = \operatorname*{lex\,arg\,min}_{x \in U_t} \{E_S(x) - \widehat{\ell}_t^{-x}(c_x), \mathrm{id}(x)\},$$

$$\pi_{\mu\text{-margin}}(H_t) = \operatorname*{lex\,arg\,min}_{x \in U_t} \{\mu_t(x) - \widehat{\ell}_t^{-x}(c_x), \mathrm{id}(x)\}.$$

The target-margin rule uses the current history-consistent hull, whereas Delta-Hull uses complete-pool membership probabilities. Implementation-level feature and covariance details are documented in the reproducibility artifact.

Execution template and planner-cost accounting. At each legal state, the implementation (i) conditions the joint posterior on the immutable history of revealed energies; (ii) evaluates complete-pool membership probabilities for legal candidates; (iii) selects the policy action; (iv), when enabled, enumerates first actions under common posterior worlds and applies the continuation; and (v) reveals the selected target energy and appends it to the full archive. With $n = |U_t|$ legal candidates and M posterior worlds, Delta-Hull evaluates nM candidate–world pairs and at most nM complete-hull constructions when each pair is materialized separately. A horizon-b rollout enumerates nM first-action branch–world pairs and at most bnM branch-stage hull constructions. The generic horizon-b notation is used for implementation cost accounting only; all reported planning experiments set $b = 2$, reducing to $b = 1$ at the final query round when one query remains. Common posterior worlds are reused across first-action branches. These are planner/evaluator costs; target-query cost is a separate quantity. Finite Monte Carlo evaluation can reverse the exact rollout improvement relation in an individual realization.

Posterior, representation, and hull protocol. The MatPES transport task uses the 2025.2 release (Kaplan et al., 2026a) and pairs PBE and r2SCAN calculations for the same configuration. An exact chemical system is the group of paired candidates with the same elemental system (`chemsys`); fit and query systems are disjoint within every cross-fitting fold. The policy-visible source descriptor has 133 dimensions. Its first 15 entries are cohesive energy, transformed site count, log volume per atom, density, band gap and its missingness flag, normalized symmetry number, three force-norm summaries, log stress norm, log lattice singular-value anisotropy, and three composition-weighted atomic-number statistics. The remaining 118 entries are elemental fractions in atomic-number order. All descriptor inputs are PBE/source observables; r2SCAN outcomes and target geometries are excluded. Missing band gaps are represented by zero plus the explicit missingness indicator, and feature scales use a 10^{-8} floor.

The local structure representation is the 64-dimensional source-crystal pooled feature from CHGNet 0.3.0 (Deng et al., 2023). It uses the source structure; target structure and target outcome are absent from the representation. The target energy model is

$$E_T = E_S + \phi(x)^\top \beta + b_s + g_s(h(x)) + \varepsilon_x, \tag{A.1}$$

where the mean is a system-balanced ridge transport model with penalty 1.0, b_s is a system random intercept, and g_s is a local Matérn-$5/2$ residual GP. Source features are standardized; variance components have 10^{-8} floors, local-kernel hyperparameters use empirical-Bayes L-BFGS-B bounds, and the posterior construction uses 10^{-10} kernel jitter. At query time the model conditions on every legally revealed energy in the current system and forms one joint Gaussian over the remaining target energies. Its covariance is symmetrized and eigen-decomposed before randomized-Sobol sampling.

The hull backend is fixed-composition. The evaluated reference set contains elemental reference phases with corrected formation energy zero; lower-order compound competitors are outside this evaluated reference set. Policy scores use per-atom formation energy, while corrected total energy is the per-atom value multiplied by the exact stoichiometric atom count. Duplicate compositions are grouped by rounded reduced composition at 12 decimal places, and hull membership uses numerical tolerance 10^{-11}. Candidate, reference, duplicate, tolerance, tie-breaking, and query-observation

rules are identical across all policies and across D, F, and T evaluation. Thirteen systems fall outside the transport model's element support; they remain in complete-pool utility estimates, while membership-probability calibration covers the 217 transport-supported systems.[1]

The MatPES evidence is conditioned on this closed finite pool, elemental references, equal query costs, additive selected-item utility, and the fixed cross-fitted posterior. Open candidate pools, heterogeneous query costs, nonadditive utilities, and other posterior models define further operating regimes for the same decision structure. Known fixed external phases can be incorporated into $\mathcal{R}^0$ before policy execution without changing the formulation.

A.3 Matched CAL hull-entropy control

The Matched CAL hull-entropy control uses the same 230-system cross-fitted panel, five 46-query/184-fit cross-fitting folds, $B = 6$ trajectories, fixed-composition hull backend, and policy-facing posterior as Delta-Hull and Target Margin. It uses 200 posterior worlds and 10 conditional fantasies per candidate to estimate expected whole-hull entropy reduction; $B = 1, \ldots, 6$ are complete prefixes of the same trajectories. All 230 systems enter the primary utility analysis; the entropy acquisition and runtime summaries cover the 217 transport-supported systems.

Let $G = \{g_1, \ldots, g_c\}$ be the fixed grid of unique reduced compositions in the visible query pool; reference-only compositions remain hull inputs but are not entropy coordinates. Posterior energy world $e^{(m)}$ induces the complete-pool lower-hull vector

$$\mathbf{h}^{(m)} = \left[L(g_1; e^{(m)}), \ldots, L(g_c; e^{(m)})\right]^\top. \tag{A.2}$$

For the M sampled vectors, the implementation uses the Gaussian differential entropy approximation

$$\widehat{\mathcal{H}}(\mathbf{L}) = \frac{c}{2}\log(2\pi e) + \frac{1}{2}\log\det\left(\widehat{\Sigma}_H + \lambda I\right), \qquad \lambda = 10^{-10}\max\left\{\max_{ij}|(\widehat{\Sigma}_H)_{ij}|, \epsilon_{\text{tiny}}\right\}, \tag{A.3}$$

where $\widehat{\Sigma}_H = M^{-1}\sum_m (\mathbf{h}^{(m)} - \bar{\mathbf{h}})(\mathbf{h}^{(m)} - \bar{\mathbf{h}})^\top$ and ϵ_{tiny} is the smallest positive IEEE 754 binary64 normal number, $2.2250738585072014 \times 10^{-308}$. The relative ridge gives a finite log determinant for rank-deficient hull vectors. Each candidate score averages the conditional entropy over 10 Gaussian fantasies and subtracts it from the current-state entropy. Specifically, at decision state H_t let

$$E_{U_t} \mid H_t \sim \mathcal{N}(\mu, \Sigma). \tag{A.4}$$

For candidate x and noiseless fantasy observation $y \sim \mathcal{N}(\mu_x, \Sigma_{xx})$, the remaining coordinates follow

$$\mu_{-x|y} = \mu_{-x} + \Sigma_{-x,x}\Sigma_{xx}^{-1}(y - \mu_x), \tag{A.5}$$

$$\Sigma_{-x|x} = \Sigma_{-x,-x} - \Sigma_{-x,x}\Sigma_{xx}^{-1}\Sigma_{x,-x}. \tag{A.6}$$

Candidate x is fixed to y in each conditional world, already revealed coordinates remain fixed, and all other candidates are sampled from this conditional Gaussian. Each candidate uses 10 fantasies and each conditional entropy uses $M = 200$ posterior worlds. The largest candidate pool in this study has 64 candidates, hence the unique-composition hull-vector dimension satisfies $c \leq 64 < 200 = M$ for every system.

With exact chemical system as the paired analysis unit, the Delta-Hull-minus-Matched-CAL-Hull-Entropy difference is $+0.100$, $+0.243$, $+0.457$, $+0.630$, $+0.630$, and $+0.700$ confirmations at $B = 1, \ldots, 6$, respectively. The $B = 6$ inverted 95% paired sign-flip interval under the stated symmetry convention is $[+0.587, +0.814]$ ($p < 10^{-4}$), and the trapezoidal $B = 1$–6 effect is $+2.361$ ($[+2.009, +2.712]$, $p < 10^{-4}$). The corresponding Delta-Hull-minus-target-margin integrated effect is $+0.367$ ($[+0.107, +0.629]$, $p = 0.0065$).

[1] For these systems, all policies use the same source-observable margin rule with fixed tie-breaking, so their utility prefixes are identical and their contribution to policy contrasts is zero.

Table A.3: Matched CAL hull-entropy control on 230 systems. Absolute columns report mean complete-pool confirmations T; the contrast is Delta-Hull minus Matched CAL hull-entropy control.

B	Target Margin	Matched CAL hull-entropy control	Delta-Hull	ΔT (95% interval)	p
1	0.839	0.813	0.913	$+0.100$ $[+0.057, +0.143]$	$< 10^{-4}$
2	1.591	1.430	1.674	$+0.243$ $[+0.177, +0.310]$	$< 10^{-4}$
3	2.287	1.935	2.391	$+0.457$ $[+0.369, +0.545]$	$< 10^{-4}$
4	2.913	2.374	3.004	$+0.630$ $[+0.533, +0.728]$	$< 10^{-4}$
5	3.422	2.822	3.452	$+0.630$ $[+0.516, +0.745]$	$< 10^{-4}$
6	3.765	3.109	3.809	$+0.700$ $[+0.587, +0.814]$	$< 10^{-4}$
Trapezoidal $B = 1$–6 area	12.515	10.522	12.883	$+2.361$ $[+2.009, +2.712]$	$< 10^{-4}$

A numerical-sensitivity diagnostic varied Delta-Hull over $M \in \{64, 128, 256, 512, 1024\}$; mean $B = 6$ utility changed by 0.004 from $M = 512$ to $M = 1024$. A matched 200-world/10-fantasy versus 400-world/20-fantasy CAL diagnostic covered 58 action-prefix/candidate states (82.8% top-1 agreement; median rank correlation 0.894).

All five fold-specific Delta-Hull-minus-Matched CAL hull-entropy control integrated effects have the same sign (range $[+1.891, +2.880]$ confirmations per system in trapezoidal area). Two independent randomized-Sobol scrambles of the same 230-system, 200-world/10-fantasy control gave $B = 6$ Delta-Hull-minus-Hull-entropy effects of $+0.683$ and $+0.687$, and integrated effects of $+2.170$ and $+2.254$. Table A.4 separates the utility mechanism from acquisition cost. The Matched CAL hull-entropy control has more selected-history revocations by unqueried competitors in this run, reflecting the additional opportunities for complete-pool adjudication to change local membership. Its conditional-fantasy integration also incurs substantially greater acquisition cost. Runtime is recorded wall-clock time for each complete six-query trajectory in the same 200-world execution on a dual-socket Intel Xeon Platinum 8160 host (48 physical cores, 96 hardware threads). The Complete-Pool Mean Margin control provides a simpler complete-pool scoring comparator.

Table A.4: Mechanism and evaluated implementation wall-clock at $B = 6$. Utility columns ($F, T, F - T$) are means over all 230 systems; runtime summaries are over the 217 transport-supported systems executing the modeled acquisition. Runtime is seconds per six-query system.

Policy	F	T	$F - T$	Runtime median	IQR	95th percentile
Target Margin	4.165	3.765	0.400	3.725	3.607–3.924	5.636
Matched CAL hull-entropy control	3.674	3.109	0.565	43.484	18.145–121.589	1,792.280
Delta-Hull	4.070	3.809	0.261	4.123	3.861–4.579	5.964
Complete-Pool Mean Margin	4.400	3.809	0.591	87.609	55.895–157.803	1,275.233

A.4 Complete-Pool Mean Margin control

The Complete-Pool Mean Margin control compares two complete-pool scoring rules: posterior-world averaging and a deterministic plug-in. It uses the same five-fold 230-system panel, conditioned posterior, candidate pool, fixed-composition hull evaluator, and $B = 6$ trajectory design as the Matched CAL hull-entropy control. Its score is the deterministic leave-one-out margin of the posterior-mean complete pool: for candidate x, it subtracts the lower-hull energy at c_x formed from the reference phases and the posterior mean of every remaining candidate except x from $\mu_t(x)$, and selects the smallest margin. No posterior-world integration is used to compute this score; the matched trajectory execution retains the same outer protocol and seed. It is an objective-evaluation control. The 217 transport-supported systems provide six candidate-level margin diagnostic states each. The remaining 13 systems use the deterministic source-observable fallback and contribute their complete utility prefixes but no model-based margin diagnostics.

The utility curve is close to Delta-Hull at the terminal budget: $T = 3.809$ versus 3.809 at $B = 6$. The integrated area is 12.972 versus 12.883 confirmations–budget per system, respectively. The paired Delta-Hull-minus-margin contrast is 0.000 at $B = 6$ ($[-0.058, +0.058]$, $p = 1.000$; 20/190/20 wins/ties/losses), and -0.089 in integrated area ($[-0.252, +0.074]$, $p = 0.303$; 37/145/48). The five fold-specific integrated effects (Delta-Hull minus margin) range from -0.217 to $+0.152$ confirmations per system. First-action agreement is 72.6%, agreement over all six rounds is 54.5%,

and complete six-round trajectories agree on 38/230 systems. This control therefore supplies a direct utility-level comparison between posterior integration and a deterministic complete-pool score.

Table A.5: Complete-Pool Mean Margin control on the 230-system cross-fitted panel. Utility means use all 230 systems. The final column is the per-budget Delta-Hull-minus-margin difference; paired inference for $B = 6$ and the integrated area is reported in the text.

B	Complete-Pool Mean Margin	Delta-Hull	ΔT (Delta−margin)
1	0.917	0.913	-0.004
2	1.709	1.674	-0.035
3	2.417	2.391	-0.026
4	2.991	3.004	$+0.013$
5	3.491	3.452	-0.039
6	3.809	3.809	$+0.000$
Trapezoidal $B = 1$–6 area	12.972	12.883	-0.089

As a separate numerical robustness diagnostic, four independent randomized Sobol scramblings of the 200-world Delta-Hull trajectory were evaluated on the same five-fold panel. Across scramblings, the integrated area ranged from 12.776 to 12.850 and the $B = 6$ mean T ranged from 3.787 to 3.822; mean pairwise top-1 action agreement was 94.1% (mean agreement over all rounds: 79.4%).

A.5 PS-BAX TARGET-SET CONTROL

We use the PS-BAX target-set control for a target-set heuristic motivated by posterior-sampling BAX constructions and locally reproducing the target-set construction. At each legal state it draws one deterministic scrambled-Sobol posterior world, forms the complete-pool target set on that world, and ranks candidates in the sampled set by posterior standard deviation, with fixed candidate-index tie-breaking. If the sampled target set is empty, the same uncertainty ranking is applied to all legal candidates and the event is recorded as a fallback. The policy receives source observations, posterior state, and legal candidates; complete-pool labels are joined after the trajectory for evaluation.

Across the five folds, each of the 230 query systems contributes one $B = 6$ trajectory, with $B = 1, \ldots, 6$ obtained as prefixes of that trajectory. The policy draws one posterior world per decision, uses fixed-composition hull evaluation, and has equal query costs. It yields mean terminal $T = 2.8435$ and integrated area 9.4652, compared with 3.8087 and 12.8826 for Delta-Hull. The paired Delta-Hull-minus-target-set effect is $+3.4174$ in integrated area (paired system-bootstrap 95% CI $[+2.9652, +3.8804]$; paired sign-flip $p = 10^{-5}$; 176/45/9 wins/ties/losses). Fold-specific effects are $+2.880$, $+3.630$, $+3.185$, $+3.283$, and $+4.109$. Action agreement falls from 34.3% at $B = 1$ to 13.5% at $B = 6$.

Three additional independent runs with seeds 20260811, 20260812, and 20260813 reused the same 230-system five-fold inputs and experimental configuration. Their Delta-Hull-minus- target-set integrated effects were $+3.185$, $+3.200$, and $+3.098$, with paired-bootstrap 95% intervals $[+2.783, +3.598]$, $[+2.728, +3.685]$, and $[+2.643, +3.567]$, respectively. The PS-BAX target-set control's $B = 6$ mean T ranged from 2.991 to 3.013 and its integrated area from 9.683 to 9.785.

The two policies use the same system set and complete-pool label map. Median wall time is 4.32 seconds per system for Delta-Hull and 4.07 for the PS-BAX target-set control; these values describe the evaluated implementations.

A.6 MATCHED COMPLETE-POOL EVALUATION CONTROL

The Local-Hull Probability study compares Target Margin, the Local-Hull Probability control, and Delta-Hull on the 230-system cross-fitted panel. The primary 230-system ablation and Delta-Hull share the conditioned posterior, 1,024 randomized Sobol worlds, hull tolerance, candidate roster, and fixed tie-breaking. The utility-evaluation pool is the varied factor: the ablation tests each candidate against the history-consistent hull, whereas Delta-Hull tests it against the complete-pool hull in every posterior world. Five folds each use 46 query systems and the disjoint complementary 184-system transport fit. $B = 1, \ldots, 6$ results are prefixes of the same $B = 6$ trajectory.

For each paired contrast, the exact chemical system is the unit. We report the mean paired difference, a two-sided paired sign-flip p-value under the stated symmetry convention, and a 95% interval obtained by inverting that same test with one fixed set of 100,000 Rademacher draws. Policy assignment remains fixed; the sign-flip supplies an inferential symmetry convention for paired differences under a zero-effect null. The interval and test therefore target the same estimand and use the same randomization distribution.

Table A.6: Controlled utility-evaluation comparison on 230 cross-fitted systems. Absolute columns report mean complete-pool confirmations T; the contrast is Delta-Hull minus the Local-Hull Probability control.

B	Target Margin	Local-Hull Probability	Delta-Hull	ΔT (95% interval)	p
1	0.839	0.265	0.922	$+0.657$ $[+0.593, +0.720]$	$< 10^{-4}$
2	1.591	0.587	1.691	$+1.104$ $[+1.000, +1.211]$	$< 10^{-4}$
3	2.287	0.974	2.383	$+1.409$ $[+1.262, +1.556]$	$< 10^{-4}$
4	2.913	1.404	2.965	$+1.561$ $[+1.387, +1.735]$	$< 10^{-4}$
5	3.422	1.909	3.457	$+1.548$ $[+1.358, +1.737]$	$< 10^{-4}$
6	3.765	2.474	3.809	$+1.335$ $[+1.144, +1.525]$	$< 10^{-4}$
Trapezoidal $B = 1$–6 area	12.515	6.244	12.861	$+6.617$ $[+5.946, +7.289]$	$< 10^{-4}$

For the primary 1,024-world utility-evaluation comparison, the five fold-specific Delta-Hull-minus-local-evaluation integrated effects range from $+4.902$ to $+8.076$ confirmations–budget per system; all five are positive.

The O–Te state provides a candidate-level instance of the same mechanism shown in Figure 1: Local-Hull Probability selects 'mp-1077823', whose local membership probability is 1.0 but whose final complete-pool label is false. Target Margin selects 'mp-557', while Delta-Hull selects 'mp-12177', which has the largest complete-pool membership probability among the legal candidates in that state.

The same policies and analysis were then evaluated on a secondary held-out 94-system MatPES panel using all 230 main-panel systems for transport fitting; query and fit rosters have zero overlap. This secondary rerun uses a separate 1,024-world execution on its held-out panel.

Table A.7: Secondary 94-system within-MatPES panel using a separate 1,024-world execution. The contrast is Delta-Hull minus the Local-Hull Probability control.

B	Target Margin	Local-Hull Probability	Delta-Hull	ΔT (95% interval)	p
1	0.798	0.277	0.862	$+0.585$ $[+0.467, +0.703]$	$< 10^{-4}$
2	1.606	0.564	1.638	$+1.074$ $[+0.889, +1.260]$	$< 10^{-4}$
3	2.234	0.894	2.340	$+1.447$ $[+1.205, +1.689]$	$< 10^{-4}$
4	2.830	1.309	2.926	$+1.617$ $[+1.341, +1.891]$	$< 10^{-4}$
5	3.351	1.819	3.436	$+1.617$ $[+1.295, +1.938]$	$< 10^{-4}$
6	3.809	2.287	3.755	$+1.468$ $[+1.122, +1.814]$	$< 10^{-4}$
Trapezoidal $B = 1$–6 area	12.324	5.867	12.649	$+6.782$ $[+5.652, +7.907]$	$< 10^{-4}$

A.7 Membership calibration

Section 5.2 and Figure 3 report the headline calibration result; this section gives the complete population definition and all calibration strata. Final-hull membership probabilities are recorded before each query observation and joined to complete-pool labels after each trajectory terminates. The primary calibration population contains every legal candidate–state prediction from the complete Delta-Hull trajectories: 37,665 predictions over 1,302 states in 217 transport-supported systems. Thirteen systems use the shared source-observable margin rule and have no model-based membership probabilities. The primary summary uses equal-system weighting: each exact chemical system contributes equal total weight. Table A.8 reports point estimates for all legal candidates and the secondary selected-action and decision-top-3 populations.

Table A.8: Final-hull membership calibration on the complete Delta-Hull trajectories. Count is the number of candidate–state predictions; every metric uses equal-system weighting. Decision top-3 contains the selected action and the two highest-probability unselected candidates at each state.

Population	Count	Brier	NLL	ECE	ROC–AUC	Average precision
All legal candidates	37,665	0.0598	0.2095	0.0263	0.9411	0.7924
Selected action (top-1)	1,302	0.1269	0.4013	0.0690	0.9044	0.9455
Decision top-3	3,906	0.1407	0.4386	0.0697	0.8941	0.9141

A.8 Lookahead and revocation diagnostics

This appendix analyzes a two-step rollout anchored at Delta-Hull and revocation diagnostics on the 230-system panel. It provides theory and mechanism analysis separately from the matched materials comparison in Section 5. For each queried candidate $x \in S_B$, let $\mathcal{A}_{t_x}$ be the archive immediately after its energy is revealed at query time t_x, $\mathcal{A}_B$ the selected-history archive, and $\mathcal{A}_{\text{full}}$ the evaluator's completed pool. Define

$$\begin{aligned} D &= \sum_{x \in S_B} \mathbf{1}\{x \in \text{Stable}(\mathcal{A}_{t_x})\}, \\ F &= \sum_{x \in S_B} \mathbf{1}\{x \in \text{Stable}(\mathcal{A}_B)\}, \\ T &= \sum_{x \in S_B} \mathbf{1}\{x \in \text{Stable}(\mathcal{A}_{\text{full}})\}. \end{aligned} \tag{A.7}$$

Here D, F, and T count query-time, selected-history, and complete-pool confirmations, respectively; T is the primary utility and $F - T$ is a descriptive count of selected-history confirmations revoked by initially unqueried competitors.

Proposition 4 (Revocation decomposition)**.** *Under the same reference phases, tolerance, and duplicate-composition rule at all three evaluations, $T \le F \le D$ and $D - T = (D - F) + (F - T)$.*

Proof of revocation decomposition. The nested archives satisfy $\mathcal{A}_{t_x} \subseteq \mathcal{A}_B \subseteq \mathcal{A}_{\text{full}}$. Adding competing phases can lower a convex-hull supporting energy but cannot make a phase stable after it was unstable; summing the resulting indicator inequalities gives $T \le F \le D$, and the decomposition is algebraic. □

The descriptive system-level revocation diagnostic (Local-Hull $F - T$ against the paired Delta-Hull-minus-Local-Hull area difference, Spearman $\rho = 0.742$) shares Local-Hull outcome components between predictor and response, so it is not independent of the contrast it explains. A pre-decision event-mismatch diagnostic avoids this coupling. At the common initial state h_0, let p_{local} and p_{term} denote the history-visible and complete-pool membership probabilities. The maximal per-candidate gap $M_{\max} = \max_x |p_{\text{local}}(x) - p_{\text{term}}(x)|$ is associated with the paired area difference on the primary panel (Spearman $\rho = 0.360$, bootstrap 95% CI $[0.241, 0.467]$; $n = 217$) and replicates on the secondary 94-system panel ($\rho = 0.378$, $[0.159, 0.561]$; $n = 91$); the mean-gap and saturation variants show the same direction with smaller magnitude, and mean area differences rise monotonically across candidate-pool-size quartiles in both panels. These predictors contain no realized terminal utility; the association remains descriptive. The association is not an artifact of candidate-pool size: controlling for log pool size attenuates the primary-panel correlation to $\rho = 0.186$ (bootstrap 95% CI $[0.076, 0.305]$) but leaves the secondary-panel correlation essentially unchanged ($\rho = 0.377$, $[0.132, 0.552]$); within-quartile associations remain positive in every stratum, and stratified permutation tests reject a pure pool-size explanation ($p \le 0.001$).

For a legal first action x and posterior world $\tilde{e}$, the rollout evaluates

$$Q_b^{\Delta}(x \mid H_t) = \mathbb{E}_{\tilde{e} \sim p(\cdot \mid H_t)} \left[R\big(S_t \cup \{x\} \cup \text{Cont}_{b-1}^{\pi_{\Delta}}(\text{Reveal}(H_t, x, \tilde{e}_x); \tilde{e}), \tilde{e}\big) \right]. \tag{A.8}$$

The Delta-Hull branch is included among the first actions, so exact evaluation is baseline-improving; finite Monte Carlo evaluation need not preserve that ordering in every realization. All reported experiments use $b = 2$; the subscript only parameterizes the estimator, not a varying experimental

horizon. The evaluated executions draw 128 common posterior worlds per decision state and 16 conditional continuation worlds per first-action branch.

The mechanism trace also reports a 39.4% reduction in selected-history revocations $F - T$ at $B = 2$ for Delta-Hull relative to the target-margin heuristic, quantifying the change between selected-history and complete-pool membership along the evaluated trajectories.

Table A.9: Measured policy wall-clock for the evaluated $B = 6$ planning contrast. Target-query cost is excluded.

Policy	Mean s/system	Median	IQR
Delta-Hull	3.992	3.913	3.573–4.237
Two-step rollout anchored at Delta-Hull	187.133	100.863	56.804–234.068

These planner wall times were measured on the dual-socket Intel Xeon Platinum 8160 host described above. The measured wall-clock ratio is $46.9\times$.

Across $B = 2, \ldots, 6$, the two policies choose different actions on an increasing fraction of the panel, from $104/230$ systems at $B = 2$ to $211/230$ at $B = 6$, while the terminal contrast remains small in the primary run, between -0.0391 and -0.0130 confirmations per system under the Delta-Hull-minus-rollout convention. The independent seed checks in Appendix A.9 keep the terminal contrast near zero, with the direction varying across seeds. These results describe the evaluated finite-MC planner and are reported with paired bootstrap intervals below.

Table A.10: Direct lookahead contrasts on the 230-system panel. Utility columns are panel means and intervals are paired system-bootstrap 95% confidence intervals. Each budget uses an independently completed closed-loop trajectory at the same two-step planning horizon.

B	Delta-Hull T	Anchored rollout T	ΔT (Delta-Hull − rollout)	95% CI	Action disagreement
2	1.6696	1.6826	-0.0130	$[-0.0391, +0.0130]$	$104/230$
3	2.3739	2.4130	-0.0391	$[-0.0783, -0.0043]$	$165/230$
4	2.9696	3.0000	-0.0304	$[-0.0739, +0.0087]$	$190/230$
5	3.4609	3.5000	-0.0391	$[-0.0783, 0.0000]$	$201/230$
6	3.8130	3.8522	-0.0391	$[-0.0783, -0.0043]$	$211/230$

A.9 Finite-MC rollout seed robustness

The rollout seed sensitivity analysis repeats the evaluated two-step contrast with two independent seeds on the same 230-system panel. Each $B = 1, \ldots, 6$ value is a prefix of a completed $B = 6$ trajectory; positive differences favor repeated Delta-Hull. The integrated effects remain positive for repeated Delta-Hull, while the terminal $B = 6$ differences stay close to zero.

Table A.11: Rollout seed sensitivity for the evaluated finite-MC two-step rollout. Entries are Delta-Hull minus anchored-rollout utility differences; the final column reports the paired system-level bootstrap interval for the integrated budget-curve effect.

Seed	$B = 1$	$B = 2$	$B = 3$	$B = 4$	$B = 5$	$B = 6$	ΔA (95% CI)
20270721	$+0.039$	$+0.113$	$+0.148$	$+0.191$	$+0.230$	-0.004	$+0.700$ $[+0.513, +0.889]$
20270722	$+0.096$	$+0.117$	$+0.165$	$+0.209$	$+0.257$	-0.026	$+0.783$ $[+0.580, +0.987]$

First-action agreement with the rollout was 8.3% and 6.5%, while full-trajectory agreement was 31.9% and 26.4% for the two seeds. Mean per-system runtimes were 4.16 versus 185.48 seconds and 4.01 versus 173.88 seconds for Delta-Hull and the anchored rollout, respectively. These runs characterize the evaluated finite-MC two-step planner; arbitrary planning horizons are not evaluated.

A.10 Rollout Monte Carlo count sensitivity

We evaluate the rollout estimator at larger sampling budgets on the 25 transport-supported systems with the smallest candidate pools (12–20 candidates). At the initial state and the second state along

the corresponding Delta-Hull trajectory, we vary the number of common first-stage posterior worlds as $M \in \{128, 256, 512\}$, with continuation counts $16/16/32$. Thus, the $128 \to 256$ comparison isolates the first-stage sample count, whereas the $M = 512$ setting increases both sampling stages. The corresponding random seeds are recorded in the reproducibility artifact.

Table A.12: Monte Carlo count sensitivity over 50 decision states (25 systems $\times$ two states). Agreement is with the $M = 128$ selection at the same state; runtime is mean seconds per rollout decision on the evaluated small-pool roster.

M	Top-1 agreement (count)	Mean runtime per decision (s)
128	—	16.3
256	0.64 (32/50)	32.2
512	0.68 (34/50)	108.2

The selected action is not stable across these counts, and the mean shift of the top-two Q-gap between $M = 128$ and $M = 512$ (0.0107, 0.74 relative) is a large fraction of the mean gap itself (0.0145): rollout scores between the leading candidates are nearly tied, so the per-decision argmax is estimator-sensitive. In only 3 of the 50 states does the $M = 128$ selection deviate from the repeated-Delta-Hull action while both higher-M selections agree with it. This diagnostic characterizes estimator stability. Terminal utility is evaluated exactly on each realized trajectory; Monte Carlo affects the planning comparison through rollout action selection rather than through utility evaluation. Whether an exactly evaluated planner would select different actions is outside the scope of this diagnostic. Quadrupling M multiplies the mean per-decision cost by $6.6\times$.

A.11 Baseline validation

A baseline-validation experiment used the same 230-system five-fold panel to evaluate generic baselines. The random policy uses three independent pre-specified permutations, while Ridge Uncertainty selects the largest marginal predictive standard deviation. It is a generic uncertainty baseline. All four runs cover all 230 systems under the same paired evaluator.

Table A.13: Generic baseline validation. Entries are mean complete-pool utility T at each budget. The area is the trapezoidal $B = 1$–6 summary; the paired interval is for the Delta-Hull-minus-baseline area difference on the exact chemical system.

Policy	seed	$B = 1$	$B = 2$	$B = 3$	$B = 4$	$B = 5$	$B = 6$	Area; paired 95% CI
Random	20260824	0.1826	0.3783	0.5957	0.8174	1.0174	1.2217	3.5109; $[+8.7239, +9.9543]$
Random	20260825	0.1696	0.3783	0.5739	0.7217	0.9609	1.1522	3.2957; $[+8.9370, +10.1870]$
Random	20260826	0.2000	0.3478	0.5174	0.7043	0.9217	1.1565	3.1696; $[+9.0783, +10.2848]$
Ridge Uncertainty	20260824	0.3261	0.7000	0.9609	1.2783	1.5478	1.7609	5.5304; $[+6.8022, +7.8501]$

The mean per-system runtimes were 3.307, 3.074, 3.150, and 3.067 seconds for the rows above, respectively (medians 3.297, 3.070, 3.130, and 3.067). First-action agreement with Delta-Hull was 4.35%, 3.04%, 2.17%, and 14.78%; full-trajectory agreement was zero for all four runs.

A.12 Observed-path posterior-movement diagnostics

This diagnostic uses 220 of 230 systems with complete state-level fields, yielding 1,320 rollout states; the remaining ten lack the required fields. Of these, the 217 transport-supported systems provide model-based membership probabilities used in the movement summary below. Across the evaluated states, the association between action disagreement and the model-relative anchored-rollout Q-gap is $\rho = 0.816$, whereas the association between realized utility gain and the maximum Q-gap is $\rho = 0.108$. For each observed candidate transition from h to h', define the membership movement $\Delta_p(h, h', x) = |p_{h'}(x) - p_h(x)|$. We summarize posterior movement at the system level by the candidate-transition average

$$\overline{\Delta}_p(i) = \frac{1}{|\mathcal{E}_i|} \sum_{(h,h',x)\in\mathcal{E}_i} \Delta_p(h, h', x).$$

Across the 217 transport-supported systems, the observed-path average Δp has mean 0.00935, median 0.00830, and 90th percentile 0.01548; the top-ranked action is preserved across 84.8% of consecutive states. On the 172 systems with overlapping primary rollout records, mean observed-path movement is not associated with realized $B = 6$ rollout gain (Spearman $\rho = 0.091$, bootstrap 95% CI $[-0.089, +0.257]$), with independent rollout seeds giving correlations near zero. Thus the observed-path statistic is a descriptive measure of posterior movement rather than an empirical estimate of the uniform bound in Proposition 2. Table A.14 summarizes the system-level statistics.

Table A.14: Observed-path bridge to the weak-posterior-movement regime. Statistics are computed over the 217 transport-supported systems; rank preservation is measured across consecutive observed states.

Diagnostic	Value
Systems	217
Mean $\overline{\Delta}_p$	0.00935
Median $\overline{\Delta}_p$	0.00830
90th percentile $\overline{\Delta}_p$	0.01548
Top-action preservation	84.8%
Full-order preservation	9.9%

A.13 Gap-triggered planning diagnostic

We evaluated a posterior-world, gap-triggered wrapper around the existing Delta-Hull policy and anchored two-step rollout. The wrapper invokes the rollout only when its pre-specified decision-value proxy $\widetilde{H}$ exceeds a fixed threshold. Let $O_x^{(m)}$ and $Y_y^{(m)}$ denote the paired observation and complete-pool membership label in posterior world m, and let $B_b(x)$ be the bth equal-frequency bin obtained by sorting $\{O_x^{(m)}\}_{m=1}^M$, with $w_b(x) = |B_b(x)|/M$. For $Q = 8$, the operational estimate is

$$\widetilde{D}_h(x) = \sum_{b=1}^{Q} w_b(x) \max_{y \neq x} \widehat{\Pr}(Y_y = 1 \mid O_x \in B_b(x), h) - \max_{y \neq x} p_h(y). \tag{A.9}$$

Ordering candidates by p_h as $x_{(1)}, x_{(2)}, \ldots$, the trigger statistic is

$$\widetilde{H} = \max\left\{0, \widetilde{D}_h(x_{(2)}) - \widetilde{D}_h(x_{(1)}), \max_{j \geq 3}\left[\widetilde{D}_h(x_{(j)}) - \widetilde{D}_h(x_{(1)}) - (p_h(x_{(2)}) - p_h(x_{(j)}))\right]\right\}. \tag{A.10}$$

The estimator reuses the already sampled posterior worlds and their complete-pool labels, then performs only binwise averaging; it does not run a conditional rollout before triggering. The five thresholds were evaluated on the same 230-system, five-fold panel with posterior MC $= 128$.

Table A.15: Gap-triggered planning diagnostic. Utility and runtime summaries use all 230 systems. The first contrast uses the same MC $= 128$ gap-triggered execution and compares each threshold with its $+\infty$ row. The second contrast is paired against the independently executed anchored-rollout reference on the 184 systems shared by the two system sets; each set contains 46 additional systems.

					$A_{\mathrm{trigger}(\tau)} - A_{\mathrm{trigger}(\infty)}$	$\Delta A_{\mathrm{trigger-rollout}}$
τ	$T_{B=6}$	Area	Runtime (s)	Trigger	(same run, $N = 230$)	(paired $n = 184$; 95% CI)
0	3.822	12.863	43.30	0.118	-0.015	-1.546 $[-1.970, -1.141]$
0.005	3.822	12.863	43.21	0.118	-0.015	-1.546 $[-1.970, -1.136]$
0.01	3.822	12.891	31.77	0.077	$+0.013$	-1.533 $[-1.976, -1.125]$
0.02	3.822	12.900	22.63	0.050	$+0.022$	-1.533 $[-1.965, -1.122]$
$+\infty$	3.804	12.878	9.70	0	0.000	-1.552 $[-1.997, -1.141]$

The $+\infty$ row is the Delta-Hull action rule executed inside the same MC $= 128$ gap-triggered execution. Its finite-MC realization and measured runtime are therefore interpreted within this diagnostic rather than numerically identified with the main Delta-Hull evaluation in Table 1. The decision-value proxy has weak association with the existing rollout value proxy: Spearman correlations between system-level $\widetilde{H}$ and the mean top-two rollout-score gap are 0.032, 0.032, 0.042, and 0.047 for thresholds 0, 0.005, 0.01, and 0.02, respectively. First-action agreement with Delta-Hull is 90.8–91.3%,

while full-trajectory agreement with the existing rollout is 7.6%. The threshold provides a controllable compute–trigger trade-off; the weak association indicates that this posterior-world proxy does not reliably identify high-value lookahead states on MatPES.

As a diagnostic of the inexpensive trigger statistic, the binned proxy $\widetilde{H}$ is positive in 12.5% of the $5,208$ evaluated states (median 0, 90th percentile 0.0078). Conditional on a positive proxy, the evaluated rollout changes the action in 49.7% of states. System-level mean $\widetilde{H}$ is not associated with realized $B = 6$ rollout gain (Spearman $\rho = 0.122$, bootstrap 95% CI $[-0.093, +0.293]$). These results characterize $\widetilde{H}$ as a compute-trigger heuristic rather than an estimator of the exact decision value D_h in Proposition 1.

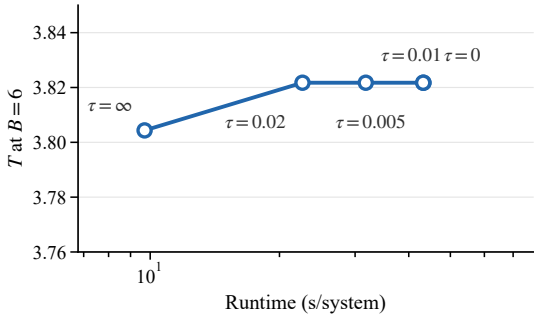


Figure A.2: Utility–runtime trade-off for the gap-triggered planning diagnostic; all points are from the separate MC=128 gap-trigger diagnostic. The $+\infty$ point denotes its no-trigger Delta-Hull action rule and is not the runtime measurement used in Table 1; labels give the trigger threshold.

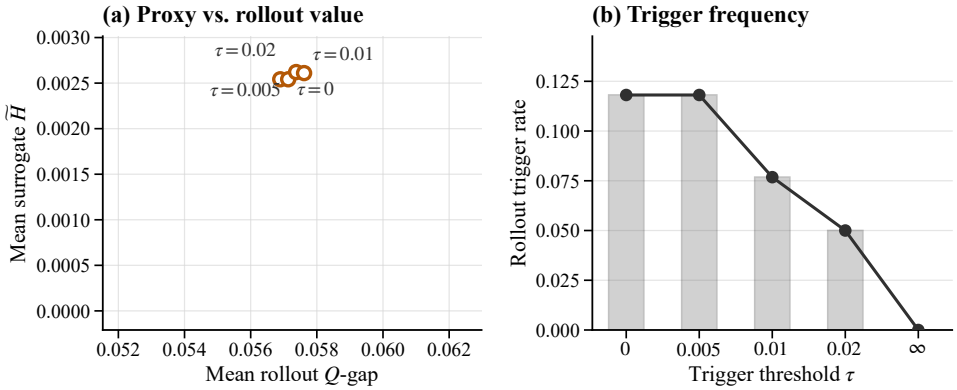


Figure A.3: Gap-triggered diagnostic summary. The two panels show the pre-specified proxy $\widetilde{H}$ versus the mean top-two rollout-score gap and rollout trigger rate across the fixed threshold grid. The duplicated utility–runtime panel is reported in Figure A.2; all threshold records remain in Table A.15.

A.14 Proofs for the main theory

We give the complete proofs of the theoretical results stated in the main text.

Proof of decision value. The first term in Eq. (8) is the posterior expected label of the queried candidate; the second is the expectation of the posterior maximum in Eq. (7). Nonnegativity follows from the tower property and convexity of the maximum, with strictness in the stated separated-maximizer case. □

Proof of exact two-step gain over greedy. The two-step value of first selecting $x_{(1)}$ is $p_1 + p_2 + D_1$. Selecting $x_{(2)}$ instead gives $p_2 + p_1 + D_2$, so the difference is $D_2 - D_1$. For $j \geq 3$, the best greedy

continuation has current value p_1, and the first-action value is $p_j + p_1 + D_j$. Relative to the greedy first action this is $D_j - D_1 - (p_2 - p_j)$; maximizing these differences and zero gives Eq. (9). □

Proof of no uniform approximation ratio. Choose an integer $K > \max\{4/\eta, 2\}$ and let Z be uniform on $\{1, \dots, K\}$. We construct a one-dimensional binary-composition hull and then apply the affine rescaling $c = u/(K+3)$, so all compositions lie in $[0, 1]$. Place fixed reference phases

$$r_i = (i, 10i^2), \qquad i = 0, \dots, K+3,$$

and write $I_v = [v-1, v]$ for the interval between consecutive anchors. The local reference chord in I_v is

$$\ell_v(t) = 10\big[(v-1)^2 + (2v-1)t\big], \qquad 0 \le t \le 1.$$

For $j = 1, \dots, K$, put a specialist at $u_{g_j} = j - 1 + 1/2$ with fixed energy $E_{g_j} = \ell_j(1/2) - 1$. For each $k \neq j$, put a distinct competitor in the same interval at $u_{y_{k,j}} = j - 1 + q_k$, where

$$q_k = \frac{1}{4} + \frac{k}{100K}, \qquad E_{y_{k,j}}(Z) = \begin{cases} \ell_j(q_k) - 2, & Z = k, \\ \ell_j(q_k) + 2, & Z \neq k. \end{cases}$$

To verify the local geometry, subtract the affine chord ℓ_j and write the four relevant points as

$$L = (0, 0), \qquad g = (1/2, -1), \qquad y = (q_k, -2), \qquad R = (1, 0).$$

Because $q_k > 1/4$, the facet yR has height

$$-\frac{1}{1 - q_k} < -\frac{4}{3} < -1 \quad \text{at } t = 1/2,$$

so g lies strictly above that facet and is revoked. Conversely, the facet Lg has height $-2q_k > -0.52 > -2$ at $t = q_k$, so y lies strictly below Lg and is itself a lower-hull vertex. If $Z = j$, every competitor in interval j is two units above the local reference chord, whereas g_j is one unit below it; hence g_j is stable. Thus g_j has a fixed self-energy and is stable exactly in world j. Each active competitor is a genuine hull phase but has label probability only $1/K$.

Use the next two intervals for decoys d_1, d_2, each at the midpoint with fixed energies $E_{d_1} = \ell_{K+1}(1/2) - 1$ and $E_{d_2} = \ell_{K+2}(1/2) - 1$. In each decoy interval add competitors y_{k,d_i} at q_k for $k = 3, \dots, K$, with the same low/high energy rule. Consequently each decoy is stable exactly in worlds 1 and 2 and has label probability $2/K$. Finally, in I_{K+3} put a probe s at the midpoint with

$$E_s(Z) = \ell_{K+3}(1/2) + 1 + \frac{Z}{K+1}.$$

The probe is always above the local reference chord, hence unstable, but its self-energy uniquely identifies Z. All observations are self-energy observations, $O_x = E_x$.

It remains to verify that the local gadgets do not destroy the fixed reference backbone. With $f(u) = 10u^2$, one has

$$\ell_v(t) - f(v - 1 + t) = 10t(1 - t).$$

The specialists and decoys are therefore at $f(u) + 3/2$, while every active competitor is at least $f(u) - 1/8$ because $q_k \in (1/4, 0.26]$. Every non-reference point is thus no more than $1/8$ below the backbone. Every chord spanning an anchor, including chords whose endpoints lie in non-adjacent intervals, has from the nearest possible candidate endpoints on the two sides at least $10(1/2)(1/4) = 1.25$ of quadratic slack at that anchor. Both endpoint perturbations are at least $-1/8$, so their linear interpolation at the anchor is also at least $-1/8$; the spanning chord therefore loses at most $1/8$ of slack, not the sum of the endpoint bounds. Hence every reference anchor remains strictly on the lower hull, and the local facet comparisons above are valid independently in each interval.

Initially $p(d_i) = 2/K > p(g_j) = p(y_{k,j}) = 1/K > p(s) = 0$. Querying d_1 reveals no information because its self-energy is fixed, so d_2 remains the strict greedy choice. The repeated-greedy policy therefore queries d_1, d_2 and receives reward 2 only in worlds $1, 2$, giving

$$V_2^{\text{greedy}} = \frac{4}{K}.$$

The policy that queries s first observes Z and then selects g_Z, giving one stable selected candidate in every world. Hence $V_2^{\text{greedy}}/V_2^\star \le 4/K < \eta$. □

Proposition 5 (Failure of adaptive submodularity)**.** *Pool-dependent convex-hull search need not be adaptively submodular, even with exact self-energy observations and equal query costs. In the $K = 2$ construction, let $\psi = \varnothing$ and let ψ' be the partial realization that queries the always-unstable probe s and observes $E_s(1)$. For the additive selected-item utility, the conditional marginal gains of querying g_1 satisfy*

$$\Delta(g_1 \mid \psi) = \frac{1}{2} \quad \textit{and} \quad \Delta(g_1 \mid \psi') = 1,$$

where adaptive submodularity would require $\Delta(g_1 \mid \psi) \geq \Delta(g_1 \mid \psi')$ for $\psi \subseteq \psi'$.

Proof of Proposition 5. In the $K = 2$ specialization above, compare the conditional marginal of g_1 before and after querying the always-unstable probe s. The probe has a self-energy observation that identifies $Z \sim \text{Bernoulli}(1/2)$, while $Y_s = 0$ and $Y_{g_1} = \mathbf{1}\{Z = 1\}$. Thus the conditional marginal of adding g_1 before any observation is $\Delta(g_1 \mid \varnothing) = 1/2$. After observing s and finding $Z = 1$, it is $\Delta(g_1 \mid \{(s, E_s(1))\}) = 1$. The marginal therefore increases after a legal self-energy observation, violating adaptive diminishing returns. The increase is driven by facet competition across compositions rather than by a change in the intrinsic energy of g_1. □

Proof of hull-margin stability. The competing-hull value is the minimum over convex combinations with nonnegative weights summing to one. Every feasible combination therefore changes by at most δ when all candidate energies change by at most δ, and the minimum also changes by at most δ. The excluded candidate energy changes by at most δ, so the signed margin changes by at most 2δ. A margin farther than 2δ from zero cannot change sign. □

Corollary 2 (Rank-stable greedy optimality)**.** *If every reachable legal observation continuation preserves the initial lexicographic ordering of legal candidates by their $p_h(x)$ values, the repeated-greedy policy is Bayes-optimal for every remaining budget.*

Proof of greedy under rank stability and weak coupling. Under rank stability, the candidate that is currently highest remains the highest after every possible continuation; backward induction therefore selects the same lexicographically ordered prefix as the repeated-greedy policy. Under Eq. (11), let $M_n(h)$ be the sum of the n largest $p_h(x)$ values. For an arbitrary adaptive policy, let $\mathcal{F}_t$ denote the history before its t-th selection and let x_t be the selected candidate. The weak-coupling condition gives

$$|\Pr(Y_{x_t} = 1 \mid \mathcal{F}_t) - p_h(x_t)| \leq \epsilon.$$

By the tower property, the expected reward at step t is the expectation of the conditional probability on the left. For every realized trajectory, the sum of the n initial scores $p_h(x_t)$ is at most $M_n(h)$; hence any adaptive policy has expected reward at most $M_n(h) + n\epsilon$. At greedy step i, let q_i be the largest initial score among the candidates still legal. At most $i - 1$ candidates have been removed, so q_i is at least the i-th largest initial score and $\sum_i q_i \geq M_n(h)$. If x_i is the greedy action and x_i^* attains q_i, then $p_{h_i}(x_i) \geq p_{h_i}(x_i^*) \geq q_i - \epsilon$. Thus the repeated-greedy policy obtains at least $M_n(h) - n\epsilon$, and the gap is at most $2n\epsilon$. □

Proof of the posterior hull-margin diagnostic. Lemma 1 gives $|m_x(E)| \leq 2\delta$ whenever $\|E - E_\psi\|_\infty \leq \delta$ and membership disagrees. Adding the coupling mismatch probability α proves Eq. (13); applying the uniform bound to Proposition 2 gives Eq. (14). □