ZERO-POOL ACTIVE LEARNING: CHOOSING EXPERIMENTS THROUGH METADATA ALONE

Anonymous authors
Paper under double-blind review

ABSTRACT

When datasets are built through prospective generation, whether via physical experiments or complex simulations, the primary cost lies in producing the samples rather than annotating an existing pool. Active learning has the right objective for this setting but the wrong assumptions: it presumes an initial labeled set, a pool of unlabeled candidates to score, and a reliable *post hoc* oracle, none of which can be assumed when data must be generated. We formalize this *zero-pool* setting and introduce ZALMBIES, an algorithm that acquires data by specifying experimental conditions rather than by selecting instances. Because such experiments are defined by a small number of categorical choices, we place a hierarchical Bayesian ANOVA model over the metadata lattice, fit it to the primary model's realized out-of-fold failure, and interpolate between additive and multiplicative geometry with a Box-Cox power selected online. Candidate bins are scored by weighing predicted failure against an effective sample size that captures the statistical strength already borrowed from the surrounding lattice, and batches are built greedily with each selection depressing the scores of overlapping bins. We show that acquisitions are D-optimal at cold start for the underlying linear model, and that under our exploration setting the resulting allocation asymptotically converges to Neyman proportions. Across four benchmarks spanning glucose simulation, quadruped robotics, wearable sensing, and sleep staging, ZALMBIES reaches the performance of random sampling using roughly half the budget. A sample here is an experiment, not an annotation, so the budget is spent in months of collection and the saving is one of schedule as much as of cost.

1 INTRODUCTION

For a growing class of applications, the cost of a dataset is dominated not by annotation but by the physical act of producing each sample. A night of polysomnography occupies a sleep laboratory and a participant until morning; a wearable study, whether of activity or of continuous glucose, must recruit, instrument, and retain a cohort and hold it to protocol; a robot must be built and operated before it can be measured; and a smart appliance must run a full cycle for every combination of load, soil, and water condition it is meant to recognize. These are all time-series problems, and in each the experiment is fixed by a handful of discrete choices before it is run: which participant, which configuration, which environment, which operating condition. In these regimes a few thousand samples can represent months of coordinated work by technicians and participants, and the budget is committed before the first sample is taken. Medical and wearable sensing are the sharpest case, as the conditions most worth measuring are often the rarest, and a passively gathered corpus underrepresents them for exactly that reason. The consequence is not simply a smaller budget. Because the collection effort is committed in advance and measured in months rather than in compute, halving the samples required also halves the interval before a design can first be evaluated, and can therefore determine whether a speculative program is undertaken at all.

Spending a data budget efficiently is the objective of Active Learning (AL) (Ren et al., 2021), improving data-efficiency by preferentially selecting informative samples, either lowering the budget needed to reach a given model performance or achieving superior performance under a fixed budget. Its methods, however, were developed under the opposite scarcity, where data are plentiful and only the labels are expensive. Traditional AL therefore fails when data collection is prospective rather than retrospective, as its method relies on three main conditions: (1) there is an initial labeled dataset

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.

on which to train a model; (2) there is a pool of unlabeled data to query; and (3) an expert oracle can reliably convert an unlabeled sample into a labeled sample, *post hoc*. This work addresses the critical regime of prospective data collection, specifically where either or both of conditions (2) and (3) are violated.

Condition (2) fails whenever data acquisition is itself the bottleneck, since a pool cannot be queried before it exists. We call this the *zero-pool* setting, where a query names the experimental conditions to run rather than an instance to label. The label is typically free in this setting, recorded by the apparatus at collection, so nothing is saved by labeling selectively and the entire expense lies in running the experiment. In these constrained scenarios, accelerating prototype development requires new techniques that operate at the interface of AL and Optimal Experimental Design (OED), where the sequential collection of data is actively designed.

Condition (3) can fail independently of condition (2). Time-series field data gathered from deployed devices or customer use arrives in volume but unlabeled, and there may be no mechanism, human expert or otherwise, able to supply reliable labels after the fact. Where *post hoc* annotation is attempted at all, as in continuous sensor readings or polysomnography-based sleep staging, it is notoriously unreliable. For instance, in sleep staging, inter-rater reliability among experts averages only $\sim 82\%$, with agreement on the most challenging stages falling as low as $\sim 40\%$ (Rosenberg & Van Hout, 2013; Cesari et al., 2021; Lee et al., 2022). Models that have been trained on the biased data have been shown to better align with expert consensus than an individual expert (Bakker et al., 2023). This is not a quirk of sleep staging. Human annotation is at its best when it can lean on the visual system, and a multichannel time-series readout gives it little to lean on. It takes an unusual image to defeat expert labeling; it takes an ordinary time series.

Lastly, even condition (1) is not a hard requirement, as the cold-start case is covered by the same machinery rather than by a separate initialization heuristic (App. F).

This work targets prospective data collection in physical, time-series environments. Our contributions are the following. **(i)** We formalize *zero-pool* active learning, the gap between pool-based AL and the continuous design spaces of Bayesian OED, where acquisition specifies the metadata of an experiment rather than selecting an instance to label (Section 3.1). **(ii)** We introduce Zero-pool Active Learning through Metadata-based Bayesian Intelligent Experiment Selection (ZALMBIES, pronounced similar to *zombies*), which fits a hierarchical Bayesian ANOVA surrogate over the metadata lattice to the primary model's realized failure, with a Box-Cox power calibrated online to bridge additive and multiplicative error, and scores bins against an effective sample size that accounts for borrowed lattice strength. **(iii)** We show that one rule spans the trajectory from zero labeled data, exactly D-optimal at cold start for the underlying linear model and asymptotically Neyman under our exploration setting. **(iv)** We assemble the first evaluation suite for the setting, on which ZALMBIES matches random collection with roughly half the samples, alongside a pool-based AL curve supplied as a reference, for which we construct a pool the setting does not otherwise provide. The common thread is that the metadata specifying an experiment carries enough signal to choose it, before any sample from that experiment exists.

2 Related Work

Active Learning and Optimal Experimental Design (OED) — Active Learning minimizes labeling cost by querying the most informative samples, typically through uncertainty or geometric diversity heuristics (Ren et al., 2021), but its dominant modern form is pool-based, requiring a pre-existing unlabeled database. Early work established the notion of a learner constructing its own queries (Angluin, 1988), but direct synthesis was found in practice to produce uninterpretable, off-manifold samples (Baum & Lang, 1992; Settles, 2009). ZALMBIES retains the premise but not the mechanism by querying metadata conditions rather than raw instances, ensuring each request is a well-posed instruction to an experimenter. Classical OED instead optimizes scalar functions of the information matrix, such as A- or D-optimality, to minimize parameter uncertainty for a predefined model (Pukelsheim, 2006). Bayesian OED makes the design sequential, directing each experiment to maximize Expected Information Gain (Chaloner & Verdinelli, 1995), and these criteria were subsequently carried into sequential active data selection (MacKay, 1992).

OED and AL in Modern Deep Learning — Recently, principles from (B)OED have been integrated into modern deep learning frameworks. Combinatorial Bayesian Optimization exploits discrete interactions for black-box minimization through sparse linear surrogates, diffusion kernels, trust regions, or smoothed categorical densities (Baptista & Poloczek, 2018; Oh et al., 2019; Papenmeier et al., 2023; Häse et al., 2021), but has no established adaptation to active learning. Information-theoretic OED principles have been adapted to pool-based active learning regimes. Methods such as BALD (Houlsby et al., 2011; Gal et al., 2017) and BatchBALD (Kirsch et al., 2019) utilize mutual information to acquire samples, while frameworks like BAIT (Ash et al., 2021) explicitly select batches that maximize Fisher information. In zero-pool environments, Deep BOED methods estimate EIG using variational lower bounds, allowing deep models to actively dictate continuous physical or computational experiments (Foster et al., 2019). Despite these successes, classical OED relies on rigid model specifications, and Deep BOED requires continuous, differentiable parameter spaces to estimate bounds. Consequently, both methodologies become computationally intractable or theoretically invalid when applied to highly combinatorial, often non-ordinal, categorical grids common in structured engineering problems.

Categorical Metadata and Bandit Frameworks — When datasets possess structural groupings or categorical metadata, the AL literature typically leverages this information to stratify sampling, mitigate bias, or enforce domain generalization (Nguyen & Smeulders, 2004; McEwen et al., 2025; van Liebergen et al., 2024). The classical form of this allocation problem is stratified survey sampling, where Neyman allocation distributes a fixed budget across strata in proportion to their within-stratum variability (Neyman, 1934), and whose adaptive descendants treat the unknown stratum variances as a bandit problem to be estimated online (Antos et al., 2010; Carpentier et al., 2011). Multi-Armed Bandit (MAB) frameworks have been applied to data collection to optimize the exploration across these discrete categories. Contextual bandits have been employed to sequentially select which stratum of an unlabeled database to query (Bouneffouf et al., 2014; Hsu & Lin, 2015). Expanding bandits to experimental design, Stochastic Contextual Linear Bandits have been utilized to mathematically optimize sequential data collection (Zanette et al., 2021). For instance, the LinUCB algorithm (Li et al., 2010) models the expected reward as a linear function of its contextual features, including categorical metadata. Yet each approach either relies on a pre-existing pool, treats strata as independent arms requiring initial sampling before scoring, or assumes linear rewards unable to capture a deep model's heteroscedastic predictive variance.

3 METHODOLOGY

In this section, we first state the problem setting, then detail the ANOVA-based categorical decomposition and the generalization to Box-Cox, before introducing the uncertainty measure and how the scoring and batch-wise acquisition functions are constructed.

3.1 PROBLEM SETTING

Zero-pool active learning. Let $\mathcal{E}$ be a space of experimental designs, each e specifying an experiment runnable at cost $\kappa(e)$, with $\mathcal{E}_{\text{feas}} \subseteq \mathcal{E}$ those actually attainable. Running e draws $(x, y) \sim P_e$ rather than returning a chosen instance, so a design fixes a distribution, not a sample. Its target y is either a coordinate of e, which lets a policy set its own class composition, or is measured at collection; either way no *post hoc* oracle is invoked. We call this *labeling at collection*, a precondition of the setting rather than a convenience, as an unlabeled outcome leaves the policy no better off for having run the design. Acquisition proceeds in rounds, a policy $\pi_t : \mathcal{D}_{t-1} \mapsto \mathcal{E}_{\text{feas}}^{B_t}$ mapping labeled history to B_t designs whose outcomes extend it, under a budget $\sum_{t,b} \kappa(e_{t,b}) \leq C$ committed in advance. There is no unlabeled pool of candidates $(x, \cdot)$ to score, so condition (2) of Section 1 fails. A policy is judged on a target $Q = \int_{\mathcal{E}} P_e \, \mathrm{d}q(e)$, whose composition q over designs is fixed by the deployment and need not resemble the allocation the policy chose. With $f_{\mathcal{D}}$ the primary model fitted on $\mathcal{D}$, one realized history $\mathcal{D}_T$ is scored by $L(f_{\mathcal{D}_T}; Q)$, a performance functional of the joint law of $(f_{\mathcal{D}_T}(x), y)$ under Q. The policy is then evaluated by an aggregating functional Ψ over the runs π induces:

$$\mathcal{R}(\pi) = \Psi_{\mathcal{D}_T \sim \pi}\big[\, L\big(f_{\mathcal{D}_T}; Q\big)\,\big]. \tag{1}$$

The categorical case. Here $\mathcal{E}$ is a *design lattice* $\mathcal{K} = \prod_{d=1}^{D} \mathcal{C}_d$, a finite product of unordered categorical dimensions, so a design is a bin $k = (c_1, \ldots, c_D)$ with $c_d \in \mathcal{C}_d$ the active category in dimension d, and writing $m_d = |\mathcal{C}_d|$ gives $|\mathcal{K}| = \prod_{d=1}^{D} m_d$. We take $\kappa \equiv 1$, so that $C = N_{\text{end}} - N_{\text{init}}$, with $\mathcal{E}_{\text{feas}} = \mathcal{K}$ and every bin queryable without limit. Deep BOED occupies the complementary corner, a continuous differentiable $\mathcal{E}$ scored by posterior entropy under a specified likelihood rather than by (1), while pool-based AL instead requires each P_e to be an enumerable finite sample. ZALMBIES does not optimize $\mathcal{R}$ directly, but targets a surrogate for the primary model's realized failure, whose relation to $\mathcal{R}$ is established empirically in Section 4.

3.2 ANOVA CATEGORICAL DECOMPOSITION AND BOX-COX INTERPOLATION

We use a classical Analysis of Variance (ANOVA) structural linear decomposition (Nelder, 1977) to parameterize these bins. Any target function $U(x)$ that assigns values to samples can then be modeled as the sum of a global mean, main effects, and higher-order interactions to give the expectation:

$$U_k \equiv \mathbb{E}[U(x)]_{x \in k} = \mu + \sum_{d=1}^{D} \omega_{c_d}^{(d)} + \sum_{d=1}^{D-1} \sum_{d'=d+1}^{D} \omega_{c_d, c_{d'}}^{(d,d')} + \ldots \tag{2}$$

where μ is the global mean, $\omega_{c_d}^{(d)}$ the main effect of category c_d in dimension d, $\omega_{c_d, c_{d'}}^{(d,d')}$ the two-way interaction between c_d and $c_{d'}$, and the terms continue through the D-way interaction. An effect indexed by t dimensions is a *tier-t* effect, so μ is tier 0 and a bin's own coefficient is the deepest tier, $t = D$.

In adopting an ANOVA coding (Gelman, 2005), we pose $\mathbf{w} = (\mu, \boldsymbol{\omega}^{(d)}, \boldsymbol{\omega}^{(d,d')}, \ldots) \in \mathbb{R}^M$ and express our bins as a fully one-hot encoded feature vector $\mathbf{x}_k \in \{0,1\}^M$, such that

$$U_k = \mathbf{x}_k^T \mathbf{w}. \tag{3}$$

A column is specified by choosing, for each dimension, either one of its categories or no involvement at all, so $M = \prod_{d=1}^{D}(m_d + 1)$, of which the $|\mathcal{K}|$ deepest-tier columns index the bins themselves and the remaining $M - |\mathcal{K}|$ form the lower-tier *parent lattice*. A bin activates exactly one column per subset of the D dimensions, so $\mathbf{x}_k$ has 2^D non-zero values, of which only the deepest is its own; each of the other $2^D - 1$ is shared with every bin agreeing with k on the sub-tuple indexing it. This redundancy is deliberate: it makes level permutations act on the columns by permutation matrices, as Theorem 4 requires, and the per-subset activation just described is what confines borrowing to the prior's scale (Proposition 2). An identified coding singles out a reference level and supplies neither.

While the above presents a strictly additive feature interaction space, metrics of model failure (e.g., entropy, class deficit, bounded losses) frequently exhibit multiplicative, highly skewed geometries where relative error is more physically meaningful than absolute error. To relax the rigid additive assumption and allow the data to dictate the interaction geometry, we apply a Box-Cox transformation (Box & Cox, 1964) to the target observations:

$$U_i^{(p)} = \begin{cases} \frac{U_i^p - 1}{p} & \text{if } p \neq 0 \\ \log(U_i) & \text{if } p = 0 \end{cases} \tag{4}$$

where we bound the power parameter $p \in [0, 1]$, motivated by the geometry of the target. For $p < 0$ the transform is unbounded below as $U \to 0$, so bins with near-zero measured failure would dominate the fit; for $p > 1$ it is convex, amplifying the upper tail against the skew reduction the transform is meant to supply. The decomposition (2) is henceforth applied to $U^{(p)}$, so the fitted estimand is $U_k^{(p)} = \mathbb{E}[U^{(p)}(x)]_{x \in k}$ in every case. Notably, setting $p = 0$ transforms the linear estimator into a log-normal regression. Fitting the additive linear model to $\mathbb{E}[\log U]$ and subsequently exponentiating the prediction returns the geometric mean, which is the bin's median under log-normality. This inverse transformation mathematically guarantees strictly non-negative predictions and translates the additive weights into multiplicative scaling factors in the original domain. Similarly, fractional p provides an interpolation between the additive and multiplicative regimes.

Let $\mathbf{X} \in \mathbb{R}^{N \times M}$ denote the design matrix of N labeled samples mapped across M categorical parameters (main effects and higher-order interactions), and let $\mathbf{U}^{(p)} \in \mathbb{R}^N$ be the corresponding vector of Box-Cox transformed observations. To prevent overfitting in sparse deep-level intersections,

we impose a Bayesian prior on the parameter weights $\mathbf{w}$, governed by a diagonal prior precision matrix $\mathbf{\Lambda} \in \mathbb{R}^{M\times M}$ (Chaloner & Verdinelli, 1995). We hierarchically scale the diagonal element λ_ℓ across the tier t of column ℓ to shield higher-order interactions from transient fluctuations, thereby requiring greater empirical support before they exert influence:

$$\lambda_0 = \epsilon = 10^{-5} \approx 0, \quad \lambda_t = \gamma^{\,t-1} \tag{5}$$

where t is the interaction order and $\gamma > 1$ is the exponential tier weighting factor.

Let $\mathbf{A} = \mathbf{X}^T\mathbf{X} + \mathbf{\Lambda} \in \mathbb{R}^{M\times M}$ be the Bayesian information matrix, whose inverse is the posterior covariance matrix. The weights $\hat{\mathbf{w}}^{(p)}$ minimize the sum of squared residuals penalized by $\mathbf{w}^T\mathbf{\Lambda}\mathbf{w}$, and are equivalently the posterior mean:

$$\hat{\mathbf{w}}^{(p)} = \mathbf{A}^{-1}\mathbf{X}^T\mathbf{U}^{(p)} \tag{6}$$

For any candidate bin represented by the one-hot encoded feature vector $\mathbf{x}_k$, the smoothed, posterior prediction is generated in the transformed space and subsequently inverted back to the original domain. To keep the inverse real-valued when inverting fractional transformations, the linear prediction is clipped at zero:

$$\hat{U}_k^{(p)} = \begin{cases} \left(\max(0, \mathbf{x}_k^T\hat{\mathbf{w}}^{(p)} \cdot p + 1)\right)^{\frac{1}{p}} & \text{if } p \neq 0 \\ \exp(\mathbf{x}_k^T\hat{\mathbf{w}}^{(p)}) & \text{if } p = 0. \end{cases} \tag{7}$$

Equation (7) defines a family of predictors indexed by p, each returning a value on the original scale of U rather than that of the transformed observations $U^{(p)}$. p is optimized dynamically over a grid on $[0, 1]$ by the penalized PRESS criterion of App. A, which selects p^*. We write $\hat{U}_k \equiv \hat{U}_k^{(p^*)}$ and call the map $k \mapsto \hat{U}_k$ the *surrogate model*, a closed-form stand-in, fitted on labeled history, for the otherwise experiment-only map from a metadata bin to the primary model's realized failure in it.

The quadratic form $\mathbf{x}_k^T\mathbf{A}^{-1}\mathbf{x}_k$ projects the model's geometric uncertainty to the specified candidate bin k. By taking the inverse of this projection, we directly calculate the Effective Sample Size (α_k) for the bin:

$$\alpha_k = \frac{1}{\mathbf{x}_k^T(\mathbf{X}^T\mathbf{X} + \mathbf{\Lambda})^{-1}\mathbf{x}_k} \tag{8}$$

The term α_k dynamically quantifies the local data density, which naturally incorporates the statistical strength borrowed from the surrounding parent lattice (see App. F.1).

3.3 Uncertainty, Score, and Batch-wise Acquisition

For an active learning or active experiment design approach, we here want to utilize a target function, U, that provides a measure of sample uncertainty. Typically, AL relies on unsupervised proxy measures like predictive entropy, which can be evaluated directly on unlabeled pool data. However, because our framework builds a surrogate model exclusively over the historical, fully labeled dataset, we are not restricted to unsupervised proxies and can explicitly model the primary network's ground-truth predictive failure. Deep AL methods have used loss prediction for acquisition before (Yoo & Kweon, 2019), but they still require a candidate pool to extract backbone features. To this end, we consider the square root of the Brier score $U_i^{rb} = \max(\sqrt{\sum_y (P(y|x_i) - \delta_{yy_i})^2}, \epsilon_{rb})$, where y_i is the sample's given label and $\epsilon_{rb} = 10^{-8}$ regularizes the transforms. Note that other target functions, including linear combinations of various measures, could be used.

To construct a batch of samples without redundant exploration, we utilize a dynamically updated acquisition function. Let $\mathbf{F} \in \mathbb{R}^{B\times M}$ represent the design matrix of the pending batch, initialized as an empty matrix. During the sequential greedy selection, the covariance matrix is updated to reflect the geometric coverage of the pending samples, updating the dynamic Effective Sample Size:

$$\alpha_k(\mathbf{F}) = \frac{1}{\mathbf{x}_k^T(\mathbf{X}^T\mathbf{X} + \mathbf{F}^T\mathbf{F} + \mathbf{\Lambda})^{-1}\mathbf{x}_k} \tag{9}$$

The acquisition score S_k balances the exploitation of the frozen posterior uncertainty $\hat{U}_k$ with the exploration of sparse regions driven by $\alpha_k(\mathbf{F})$:

$$S_k = \frac{\hat{U}_k}{(\alpha_k(\mathbf{F}))^\beta} \tag{10}$$

Algorithm 1: ZALMBIES: one acquisition round

Input: History $\mathcal{D}$; uncertainty U; lattice $\mathcal{K}$ with encodings $\{\mathbf{x}_k\}$; prior $\mathbf{\Lambda}(\gamma)$; exploration β; batch size B; folds J; grid $\mathcal{P}$

Output: Batch $(k_1, \ldots, k_B)$ of bins to run next

// Surrogate targets and Bayesian information matrix

Fine-tune the primary model on $\mathcal{D}$ in J stratified folds ;

$U_i \leftarrow U$ evaluated out-of-fold, $\forall i \in \mathcal{D}$;

$\mathbf{X} \leftarrow [\mathbf{x}_{k(i)}]_{i\in\mathcal{D}}$; $\mathbf{A} \leftarrow \mathbf{X}^T\mathbf{X} + \mathbf{\Lambda}$; $h_{ii} \leftarrow \mathbf{x}_{k(i)}^T \mathbf{A}^{-1} \mathbf{x}_{k(i)}$;

// Box--Cox power by penalized PRESS (App. A)

foreach $p \in \mathcal{P}$ **do**

  $\mathbf{z}^{(p)} \leftarrow$ Jacobian-normalized transform of $\mathbf{U}$; $\hat{\boldsymbol{\omega}}^{(p)} \leftarrow \mathbf{A}^{-1}\mathbf{X}^T\mathbf{z}^{(p)}$; $\hat{\mathbf{z}}^{(p)} \leftarrow \mathbf{X}\hat{\boldsymbol{\omega}}^{(p)}$;

  $O(p) \leftarrow \sum_i \left((z_i^{(p)} - \hat{z}_i^{(p)})/(1-h_{ii})\right)^2 + (\hat{\boldsymbol{\omega}}^{(p)})^T \mathbf{\Lambda} \hat{\boldsymbol{\omega}}^{(p)}$;

$p^* \leftarrow \arg\min_p O(p)$; $\hat{\mathbf{w}} \leftarrow \mathbf{A}^{-1}\mathbf{X}^T\mathbf{U}^{(p^*)}$; $\hat{U}_k \leftarrow$ back-transform of $\mathbf{x}_k^T\hat{\mathbf{w}}$, $\forall k \in \mathcal{K}$;

// Greedy batch; the running $\mathbf{A}^{-1}$ carries $\mathbf{F}^T\mathbf{F}$, so $\alpha_k = \alpha_k(\mathbf{F})$

$\alpha_k \leftarrow (\mathbf{x}_k^T \mathbf{A}^{-1} \mathbf{x}_k)^{-1}$, $\forall k \in \mathcal{K}$;

for $b = 1, \ldots, B$ **do**

  $k_b \leftarrow \arg\max_{k\in\mathcal{K}} \hat{U}_k \alpha_k^{-\beta}$; $\mathbf{g} \leftarrow \mathbf{A}^{-1}\mathbf{x}_{k_b}$; $s \leftarrow 1 + \alpha_{k_b}^{-1}$;

  $\alpha_k \leftarrow \left(\alpha_k^{-1} - (\mathbf{x}_k^T\mathbf{g})^2/s\right)^{-1}$, $\forall k \in \mathcal{K}$; $\mathbf{A}^{-1} \leftarrow \mathbf{A}^{-1} - \mathbf{g}\mathbf{g}^T/s$;

return $(k_1, \ldots, k_B)$

where β controls the exploration-exploitation trade-off (higher $\beta \Rightarrow$ higher exploration). The penalty $\alpha_k^{-\beta}$ imposes diminishing marginal returns, moving budget away from structurally dense regions, where residual error is more likely irreducible or capacity-bound, and toward sparse, high-error ones. Upon selecting the candidate $k_b = \arg\max_k S_k$, the vector $\mathbf{x}_{k_b}$ is appended to $\mathbf{F}$, iteratively depressing the acquisition scores of geometrically correlated bins. The choice to use a ratio is deliberate, as an additive confidence bonus would suppress exploration to zero against a fixed value gap, which is appropriate when minimizing cumulative regret (Li et al., 2010), but not when the budget must also estimate the bins it is not exploiting. In practice, multiple matrix inversions can be avoided by a Sherman–Morrison rank-1 update (Sherman & Morrison, 1949). The rule is closed-form and runs under five milliseconds per acquired sample on a single CPU core. Against fitting the base model, let alone running the experiments the batch commissions, acquisition does not enter the accounting. The sequential acquisition for one batch is outlined in Alg. 1.

The free parameter β interpolates between pure exploitation at $\beta = 0$ and, as β grows, a classical optimal-design limit. Theorem 3 shows that whenever the value weight is uninformative, which is exactly the cold-start case (i.e., the introduction's condition (1) not satisfied), every $\beta > 0$ makes the same selection, and that selection is exactly D-optimal for the linear model (2) the surrogate is fitted from. On a symmetric categorical lattice that optimum is the balanced allocation over bins. Intermediate $\beta = 1$ is further motivated by near D-optimality and Neyman allocation, while $\beta = 2$ has a nearly V-optimal interpretation. Once the uncertainty measure is fixed, ZALMBIES has only two free parameters, β, the explore-exploit knob, and γ, the prior weight. γ governs the lattice inheritance, and because $\mathbf{\Lambda}$ is fixed in scale while $\mathbf{X}^T\mathbf{X}$ grows, a longer history needs a higher γ to prevent deep-tier artifacts from driving the selection. In our primary experiments, we set $\beta = 1$ and $\gamma = 3$, see App. D for ablations. Both parameters, together with the fine-tuning schedule, were fixed *a priori* and held identical across every method, dataset, and budget; nothing was tuned per method or per benchmark.

4 RESULTS

Because our problem setup is novel, identifying adequate benchmark datasets is challenging. We require public datasets with (1) time-series data, (2) rich categorical metadata, and (3) massive data volume to ensure deep-tier bins are sufficiently populated. The last condition is particularly limiting. Any dataset dense enough to reliably evaluate the acquisition policy is one where data collection was not the bottleneck. This leaves two realistic options open for evaluation: synthetic data *factories*, which naturally align with our primary use-case of experiment selection and guarantee the data volume requirement is satisfied, or real datasets whose acquisition mechanism is not faithful to

Table 1: ZALMBIES savings are determined by using random IQM@N_{end}, and determining the savings by identifying the point, N_Z, where ZALMBIES IQM matches, logarithmically interpolated as plotted, and discounting N_{init}.

Benchmark	Random F_1 IQM@N_{end}	N_{end}	N_{Z}	Savings
SimGlucose	$86.80^{+0.71}_{-0.71}\%$	360	172.3	69.5%
MuJoCo	$88.22^{+0.45}_{-0.39}\%$	4455	2513.6	53.3%
MotionSense	$95.00^{+1.20}_{-1.51}\%$	96	55.2	56.7%
Sleep-EDF	$57.95^{+0.63}_{-0.63}\%$	990	522.9	57.7%

their experimental design. The SimGlucose (Man et al., 2014; Xie, 2018) simulator is the closest instantiation of the setting available, a third-party factory in which a query is a protocol to simulate, so acquisition is a native mechanism of the environment. A MuJoCo (Todorov et al., 2012; Towers et al., 2023) simulation of quadruped robot motion with varied chassis and environments, where the task is to classify surface friction from a subset of the state variables, is a second factory built for this work. Among real datasets, MotionSense (Malekzadeh et al., 2019) provides ample metadata and samples fairly well distributed in the categorical grid but limited overall volume, restricting our evaluation to the early (low-data) regime before the method reaches the dataset ceiling. The Sleep-EDF dataset (Kemp et al., 2000) contains 78 participants, demographic and auxiliary metadata, and over 100k samples once broken into 30s snippets, a treatment that is itself the unfaithful acquisition mechanism of the second option and that also makes the sleep stage a targetable design category. Fig. 3 of App. B considers participant-night sampling on a lattice of only 12 bins, testing whether the result depends on either. Listed in that order, the four benchmarks cross acquisition fidelity with scale, factories before real corpora and the smaller budget before the larger within each pair, and give the zero-pool setting its first evaluation suite. Our objective in all experiments is to isolate and quantify the sample efficiency gains of the acquisition policy, not to achieve state-of-the-art predictive performance on these specific benchmarks. Comprehensive details on the datasets are provided in App. B.

We use the MantisV2 (Feofanov et al., 2026) time-series foundation model, with a fine-tuning of a fresh pretrained model for all evaluation and baselines. We use a learning rate of 2×10^{-4}, weight decay of 0.05, and train for a fixed 30 epochs (model performance was not very sensitive to this choice). For all random seeds, we begin at 0 and increment by 1 for different trials. We additionally use a 3-fold stratified split of the data evaluated out-of-fold across all experiments to measure U for ZALMBIES and the LinUCB baseline. We score performance each run by macro-F_1 on held-out data and aggregate via the inter-quartile mean (IQM) (Agarwal et al., 2021) over 24 trials with 10000 bootstraps, setting L and Ψ of Section 3.1, respectively. The IQM reflects the robust expected performance of the active learning, insulating against noise from (un)lucky trials and allowing markedly smoother learning trajectories without qualitatively altering conclusions.

For the zero-pool environment, we compare these results against three baseline data collection mechanisms: (1) random, (2) metadata equal (equidistributed or ZALMBIES $|_{\hat{U}\equiv 1}$), (3) a LinUCB adaptation for AL; and a reference line, (4) pool-based active learning. Every method draws from a metadata-balanced design, imposed by reducing the corpus or by the draw itself (App. B). Metadata equal inherits whatever class composition that balance induces, and since it tracks random outside the most data-starved budgets, that composition carries no advantage on its own. The margin of ZALMBIES over metadata equal isolates what the value weight $\hat{U}$ contributes at our fixed β (App. D). Where the label is itself a design category (MuJoCo and Sleep-EDF), balance additionally makes every class equally available, so no macro-F_1 gain can come from rescuing a rare class. Pool-based AL solves a different problem, under assumptions this setting modifies (including that we use the label itself as a category for data selection), so its curve is the most informative available reference rather than a direct comparison. We supply a large pool many times the size of the budget, and the AL model greedily claims the samples of highest predictive entropy. In practice, we observe only minor differences from BAIT and BatchBALD (Fig. 4). The LinUCB adaptation uses $S = U_{\texttt{main}} + c_0\, \alpha_{\texttt{main}}^{-1/2}$ where only the main effects (tier-0 and 1) are used in setting U and α. We select $c_0 = 2\beta\, \bar{\hat{U}}\, \bar{\alpha}^{1/2}$, where $\bar{\hat{U}}$ and $\bar{\alpha}$ are the means of $\hat{U}_k$ and α_k over the lattice. The constant is fixed *a priori* so that both rules trade value against coverage identically at the operating point.

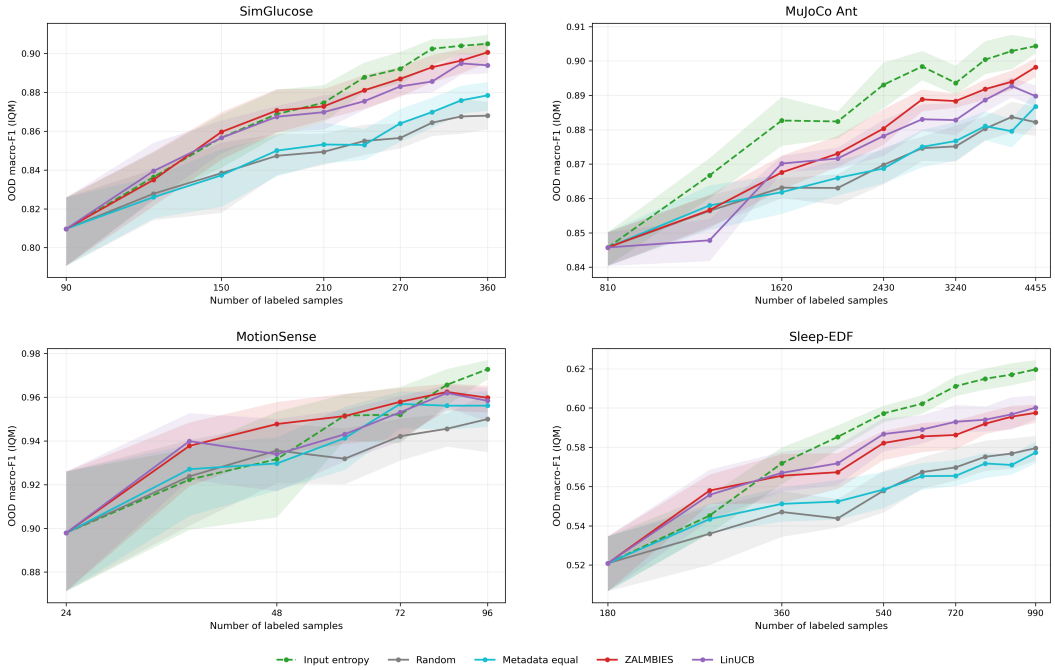


Figure 1: The performance of ZALMBIES in the zero-pool environment for all four benchmark datasets. Due to the data being balanced, random and metadata equal are expected to align closely. ZALMBIES consistently provides a 50–70% sample budget savings relative to random (Table 1). The LinUCB baseline created for this work by adapting a black-box optimization method beats ZALMBIES on Sleep-EDF. The pool-based AL reference line (dashed) requires a pool this setting precludes. We provide a favorable pool for it with many samples in each bin.

Across all four benchmarks in different regimes, ZALMBIES outperforms random and metadata equal, which tracks random in all but the most data-starved cases. Savings to the budget where central performance matches random's are 50–70%, half of runs falling short there. It does so using no information about the samples to be collected. The LinUCB baseline, created in this work for comparative purposes, exceeds random and outperforms ZALMBIES on Sleep-EDF, but is generally weaker. ZALMBIES can also be extended to standard pool-based AL by ranking the bins and labeling the highest-entropy sample within the top-ranked occupied bins. App. C develops the extension and finds it competitive with BADGE, BatchBALD, BAIT, and greedy entropy.

5 DISCUSSION

A 50–70% reduction in samples is a monetary saving, but in prospective collection it is more often a time saving. The method's underlying power comes from the ANOVA design and lattice support. The surrogate is fit to the primary model's realized out-of-fold failure on the labeled history. As bins that agree on a sub-tuple of metadata share the lower-tier columns encoding it, an unsampled bin still inherits a prediction from its parents in the lattice. Acquisition therefore requires neither a candidate instance to inspect nor even a prior visit to the bin itself, with α_k recording how much of the resulting estimate was measured in the bin and how much was borrowed around it.

The pool-based AL curve invites a direct comparison with ZALMBIES. It is not, however, a measurement. We granted it a pool many times the size of the budget, with many samples in every bin, a favorable position, and its curve moves directly with that choice: cutting it by an order of magnitude would flatter ZALMBIES, while raising it would put the reference out of reach, and neither would say anything about either method. What a pool buys is selection freedom, and a pool small relative to the budget, or large but never visiting parts of the lattice, supplies little of it, while ZALMBIES is indifferent to both. A pool can therefore be poor enough that commissioning an experiment is worth

more than labeling anything on hand, and where that point falls is a question of the relative cost of creating a sample against labeling one already paid for, not of accuracy.

The rigid structure of the categorical binning is what buys the design guarantees of Theorem 4, but real-world data-collection considerations can introduce tension. A key challenge for practitioners is constructing a strong metadata lattice, and App. E collects guidance on dimension count, category cardinality, and handling high-cardinality or rarely co-occurring fields. The encoding also assumes the lattice is a product, so a dimension whose categories exist only under certain settings of another is not simply a lattice with holes. A category confined to one parent setting has a main-effect column collinear with its interaction column, so the tiers stop separating and the problem is better split into several lattices than padded into one. Unattainable bins, safety-restricted bins, or rare categories (as may often naturally arise with human participants) break the symmetry that Theorem 4 rests on, so the optimality is no longer promised. App. F.2 identifies where the argument fails but does not bound the shortfall. The algorithm still runs, greedily moving past the bins it cannot fill, and the D-optimality of each individual acquisition is unaffected. Studying these cases properly would require data factories built to model them.

The state ZALMBIES starts from matters in both directions. Starting from nothing is free, as Theorem 3 shows that every acquisition is exactly D-optimal even before any failure has been observed. Starting from a large arbitrary allocation is not. Proposition 6 gives the limiting allocation exactly, which at $\beta = 1$ is approximately Neyman (Remark 5), but restoring equal scores from an arbitrary state costs acquisitions in proportion to the effective counts already accumulated. That cost grows as $(\hat{U}_{\max}/\hat{U}_{\min})^{1/\beta}$, so raising β mitigates it by flattening the target allocation, and does so cheaply, since an under-explored batch can be re-selected under a larger β before any experiment is run. It does not eliminate the penalty, as β must rise in step with the accumulated counts to keep the correction within one batch. Our experiments operate entirely within the sparse-data regime where this penalty remains manageable, and we leave the large-data regime to future work.

A primary challenge in expanding this approach is identifying suitable data sources. SimGlucose and MuJoCo Ant are true data factories, the latter built for this work, so their contents are dictated by the experiments run and carry none of the confounders a real setting would introduce. Sleep-EDF and MotionSense are real datasets we treat as data-on-demand factories, an adaptation that is faithful only if the samples within a bin are exchangeable with what an experiment would produce.

ZALMBIES is the first method for zero-pool active learning over categorical metadata, and the setting leaves many directions open. First, while our lattice was entirely categorical, and continuous and ordinal variables mostly admit a natural conversion, there are likely benefits to tailoring the design matrix and information matrix to handle them directly, rather than targeting the categorical simplification. The assumptions here were tailored toward equally viable experiments, but many active learning approaches consider cost, which could also arise in some prospective data scenarios with rare categories. Very deep lattices are limited by the size of $\mathbf{A}$ rather than by the acquisition rule, and one untested remedy is to let a batch probe a subset of dimensions while randomizing the rest, pruning the least valuable dimensions as it goes. There are also interesting applications in partial-information cases, for instance a setting where there is a sizable unlabeled pool of customer data but *post hoc* labeling is not viable, while specific sections of the customer base may be incentivized to provide labels for new incoming data. Lastly, the pool-based variant of this algorithm (App. C) was not optimized, and further tuning to better target those scenarios remains an opportunity.

6 CONCLUSIONS

In this work, we formalized *zero-pool* active learning, where acquisition specifies the metadata of an experiment to run rather than an instance to label, and introduced ZALMBIES, which uses historical labeled data to choose that metadata. Scoring bins by predicted failure against an effective sample size gives a single rule that makes every acquisition exactly D-optimal before any data have been seen, and that drives the allocation to Neyman proportions at $\beta = 1$ once failure can be estimated. Across glucose simulation, quadruped robotics, wearable sensing, and sleep staging, ZALMBIES reaches the performance of random collection with roughly half the samples, without inspecting a single candidate. Metadata alone is enough to decide which experiment to run next.

AI USE STATEMENT

In this work, we used generative AI tools to help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, and provide feedback on research methodology or experiments. We have not used generative AI tools to generate synthetic data sets, propose or refine hypotheses, clean or reformat datasets, support qualitative or thematic data analysis, or interpret results. Assistance with translation is not applicable to this work. Additionally, we used generative AI tools for editing software code, brainstorming, identifying relevant literature, summarizing and analyzing existing literature, editing the research paper to improve readability, and formatting references. We have reviewed all AI-assisted work. We checked LLM-generated research ideas through a manual literature survey. We refined and improved all proofs substantively from the LLMs' proposals. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Several steps have been taken to ensure reproducibility. Alg. 1 is closed-form and self-contained, requiring no training loop, no architecture choices, and no tuned schedule beyond the two parameters β and γ fixed *a priori* in Section 3.3. We have documented all training rules and random seeds. App. B specifies each benchmark completely, including the metadata lattice, the balancing and splitting procedure, and the budget schedule, and for the MuJoCo Ant data factory built for this work gives the full simulator configuration. We will release the code for all four benchmark constructions, including that data factory, the ZALMBIES implementation, and the evaluation harness before the conference.

REFERENCES

Rishabh Agarwal, Max Schwarzer, Pablo Samuel Castro, Aaron C Courville, and Marc Bellemare. Deep reinforcement learning at the edge of the statistical precipice. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 29304–29320, 2021.

David M Allen. The relationship between variable selection and data agumentation and a method for prediction. *Technometrics*, 16(1):125–127, 1974.

Dana Angluin. Queries and concept learning. *Machine Learning*, 2(4):319–342, 1988.

András Antos, Varun Grover, and Csaba Szepesvári. Active learning in heteroscedastic noise. *Theoretical Computer Science*, 411(29-30):2712–2728, 2010.

Jordan T. Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. Deep batch active learning by diverse, uncertain gradient lower bounds. In *International Conference on Learning Representations (ICLR)*, 2020.

Jordan T. Ash, Surbhi Goel, Akshay Krishnamurthy, and Sham Kakade. Gone fishing: Neural active learning with fisher embeddings. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 8927–8939, 2021.

Jessie P. Bakker, Marco Ross, Andreas Cerny, Ray Vasko, Edmund Shaw, Samuel Kuna, Ulysses J. Magalang, Naresh M. Punjabi, and Peter Anderer. Scoring sleep with artificial intelligence enables quantification of sleep stage ambiguity: hypnodensity based on multiple expert scorers and auto-scoring. *Sleep*, 46(2):zsac154, 2023.

Ricardo Baptista and Matthias Poloczek. Bayesian optimization of combinatorial structures. In *International Conference on Machine Learning (ICML)*, pp. 462–471. PMLR, 2018.

Eric B. Baum and Kenneth Lang. Query learning can work poorly when a human oracle is used. In *International Joint Conference on Neural Networks*, 1992.

Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, 1997.

Andrew An Bian, Joachim M. Buhmann, Andreas Krause, and Sebastian Tschiatschek. Guarantees for greedy maximization of non-submodular functions with applications. In *International Conference on Machine Learning (ICML)*, pp. 498–507. PMLR, 2017.

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 2006.

Djallel Bouneffouf, Romain Laroche, Tanguy Urvoy, Raphaël Féraud, and Robin Allesiardo. Contextual bandit for active learning: Active thompson sampling. In *International Conference on Neural Information Processing (ICONIP)*, pp. 405–412. Springer, 2014.

George EP Box and David R Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964.

Alexandra Carpentier, Alessandro Lazaric, Mohammad Ghavamzadeh, Rémi Munos, and Peter Auer. Upper-confidence-bound algorithms for active learning in multi-armed bandits. In *Algorithmic Learning Theory*, pp. 189–203, 2011.

Matteo Cesari, Ambra Stefani, Thomas Penzel, Abubaker Ibrahim, Heinz Hackner, Anna Heidbreder, András Szentkirályi, Beate Stubbe, Henry Völzke, Klaus Berger, and Birgit Högl. Interrater sleep stage scoring reliability between manual scoring from two european sleep centers and automatic scoring performed by the artificial intelligence–based stanford-stages algorithm. *Journal of Clinical Sleep Medicine*, 17(6):1237–1247, 2021.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2 edition, 2006.

Abhimanyu Das and David Kempe. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. In *International Conference on Machine Learning (ICML)*, 2011.

Valerii V. Fedorov. *Theory of Optimal Experiments*. Academic Press, New York, 1972.

Uriel Feige. A threshold of ln n for approximating set cover. *Journal of the ACM*, 45(4):634–652, 1998.

Vasilii Feofanov, Songkang Wen, Jianfeng Zhang, Lujia Pan, and Ievgen Redko. Mantisv2: Closing the zero-shot gap in time series classification with synthetic data and test-time strategies, 2026. arXiv:2602.17868.

Adam Foster, Martin Jankowiak, Eli Bingham, Paul Horsfall, Yee Whye Teh, Tom Rainforth, and Noah Goodman. Variational bayesian optimal experimental design. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *International Conference on Machine Learning (ICML)*, pp. 1183–1192, 2017.

Andrew Gelman. Analysis of variance - why it is more important than ever. *The Annals of Statistics*, 33(1):1–53, 2005.

David A. Harville. *Matrix Algebra From a Statistician's Perspective*. Springer, New York, 1997.

Florian Häse, Matteo Aldeghi, Riley J. Hickman, Loïc M. Roch, and Alán Aspuru-Guzik. Gryffin: An algorithm for Bayesian optimization of categorical variables informed by expert knowledge. *Applied Physics Reviews*, 8(3):031406, 2021.

Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2013.

Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning, 2011. arXiv:1112.5745.

Wei-Ning Hsu and Hsuan-Tien Lin. Active learning by learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

Bob Kemp, Arie H. Zwinderman, Bert Tuk, Hilbert A. C. Kamphuisen, and Gerard J. L. Oberyé. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the eeg. *IEEE Transactions on Biomedical Engineering*, 47(9):1185–1194, 2000.

Alex Kendall and Yarin Gal. What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017.

Jack Kiefer. Construction and optimality of generalized Youden designs. In Jagdish N. Srivastava (ed.), *A Survey of Statistical Design and Linear Models*, pp. 333–353. North-Holland, 1975.

Andreas Kirsch, Joost Van Amersfoort, and Yarin Gal. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Yun Ji Lee, Jae Yong Lee, Jae Hoon Cho, and Ji Ho Choi. Interrater reliability of sleep stage scoring: a meta-analysis. *Journal of Clinical Sleep Medicine*, 18(1):193–202, 2022.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *International Conference on World Wide Web (WWW)*, pp. 661–670, 2010.

David JC MacKay. Information-based objective functions for active data selection. *Neural Computation*, 4(4):590–604, 1992.

Mohammad Malekzadeh, Richard G Clegg, Andrea Cavallaro, and Hamed Haddadi. Mobile sensor data anonymization. In *Proceedings of the International Conference on Internet of Things Design and Implementation*, IoTDI '19, pp. 49–58, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-6283-2.

Chiara Dalla Man, Francesco Micheletto, Dayu Lv, Marc Breton, Boris Kovatchev, and Claudio Cobelli. The UVA/PADOVA type 1 diabetes simulator: New features. *Journal of Diabetes Science and Technology*, 8(1):26–34, Jan 2014.

Ben McEwen, Corentin Bernard, and Dan Stowell. Stratified active learning for spatiotemporal generalisation in large-scale bioacoustic monitoring. *bioRxiv*, 2025.

John A. Nelder. A reformulation of linear models. *Journal of the Royal Statistical Society: Series A (General)*, 140(1):48–63, 1977.

George L. Nemhauser, Laurence A. Wolsey, and Marshall L. Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical Programming*, 14(1):265–294, 1978.

Jerzy Neyman. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97(4): 558–606, 1934.

Hieu T Nguyen and Arnold Smeulders. Active learning using pre-clustering. In *International Conference on Machine Learning (ICML)*, ICML '04, pp. 79, 2004.

Changyong Oh, Jakub Tomczak, Efstratios Gavves, and Max Welling. Combinatorial Bayesian optimization using the graph Cartesian product. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Leonard Papenmeier, Luigi Nardi, and Matthias Poloczek. Bounce: Reliable high-dimensional Bayesian optimization for combinatorial and mixed spaces. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023.

Friedrich Pukelsheim. Information increasing orderings in experimental design theory. *International Statistical Review/Revue Internationale de Statistique*, 55(3):203–219, 1987.

Friedrich Pukelsheim. *Optimal Design of Experiments*. SIAM, 2006.

Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B. Gupta, Xiaojiang Chen, and Xin Wang. A survey of deep active learning, 2021. arXiv:2009.00236.

Richard S Rosenberg and Steven Van Hout. The american academy of sleep medicine inter-scorer reliability program: Sleep stage scoring. *Journal of Clinical Sleep Medicine*, 9(1):81–87, 2013.

Burr Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

Jack Sherman and Winifred J Morrison. Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix. *The Annals of Mathematical Statistics*, 20(4):621–621, 1949.

Michael H Silber, Sonia Ancoli-Israel, Michael H Bonnet, Sudhansu Chokroverty, Madeleine M Grigg-Damberger, Max Hirshkowitz, Sheldon Kapen, Sharon A Keenan, Meir H Kryger, Thomas Penzel, Mark R Pressman, and Conrad Iber. The visual scoring of sleep in adults. *Journal of Clinical Sleep Medicine*, 3(2):121–131, 2007.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033. IEEE, 2012.

Mark Towers, J. K. Terry, Ariel Kwiatkowski, John U. Balis, Gianluca Cola, Tristan Deleu, Manuel Goulão, Andreas Kallinteris, Arjun KG, Markus Krimmel, Rodrigo Perez-Vicente, Andrea Pierré, Sander Schulhoff, Jun Jet Tai, Andrew Tan Jin Shen, and Omar G. Younis. Gymnasium. https://zenodo.org/record/8127025, 2023.

Nina van Liebergen, Marianne Schaaphok, and Giovanni Sileno. Resgral: Fairness-sensitive active learning. In *Proceedings of the First Workshop on Implementing AI Ethics through a Behavioural Lens (AIEB 2024)*, volume 3948. Aachen: CEUR-WS, 2024.

Henry P. Wynn. The sequential generation of D-optimum experimental designs. *The Annals of Mathematical Statistics*, 41(5):1655–1664, 1970.

Jinyu Xie. Simglucose v0.2.1. https://github.com/jxx123/simglucose, 2018. Accessed 2026-06-03.

Donggeun Yoo and In So Kweon. Learning loss for active learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 93–102, 2019.

Andrea Zanette, Kefan Dong, Jonathan Lee, and Emma Brunskill. Design of experiments for stochastic contextual linear bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 25021–25032, 2021.

A Box-Cox Power Parameter Optimization

The Box-Cox power parameter p can be optimized dynamically from the historical data, allowing the model to better capture nonlinear relations. However, comparing residual errors across different values of p is inherently biased because the standard transformation alters the variance of the data. To allow for an unbiased assessment, the transformation must be Jacobian-normalized to preserve the geometric volume of the target space (Box & Cox, 1964). First, we compute the geometric mean of the target values, $\chi = \exp(\frac{1}{N}\sum_{i=1}^{N}\log U_i)$, which is well defined because the floor ϵ_{rb} imposed on U_i^{rb} in Section 3.3 keeps every target strictly positive, as does the $p = 0$ branch below. We then apply the Jacobian-normalized Box-Cox transform:

$$z_i^{(p)} = \begin{cases} \frac{U_i^p - 1}{p\chi^{p-1}} & \text{if } p \neq 0 \\ \chi \log(U_i) & \text{if } p = 0. \end{cases} \tag{11}$$

This normalized target vector $\mathbf{z}^{(p)}$ is used to derive the weights $\hat{\boldsymbol{\omega}}^{(p)} = \mathbf{A}^{-1}\mathbf{X}^T\mathbf{z}^{(p)}$ and the predictions $\hat{\mathbf{z}}^{(p)} = \mathbf{X}\hat{\boldsymbol{\omega}}^{(p)}$. To efficiently evaluate generalization error via leave-one-out cross-validation

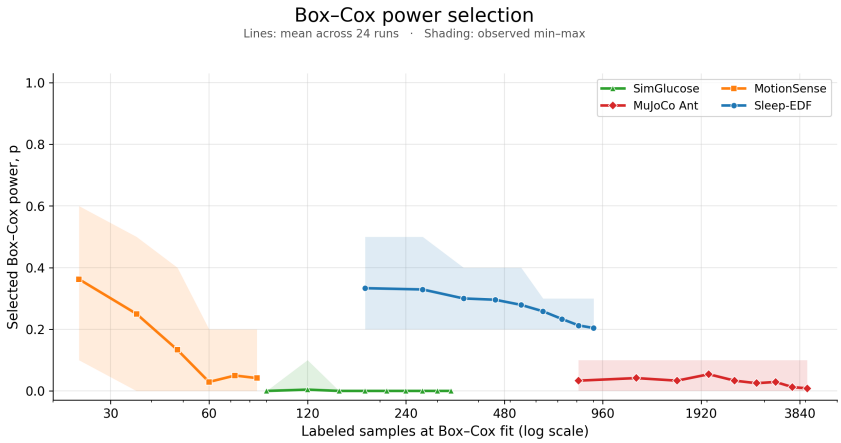


Figure 2: The mean selected p for each dataset shown with bands spanning the range of minimum and maximum selected for any trial versus the number of labeled samples prior to selection. SimGlucose and MuJoCo cleave close to 0. Sleep-EDF declines from 0.4 to 0.2. MotionSense begins near 0.4 at low data, but settles near 0, and its 0.6 is the largest p any trial selects anywhere.

(LOOCV) without retraining the linear model N times, we utilize the PRESS statistic (Allen, 1974). The leverage of each sample is $h_{ii} = (\mathbf{X}\mathbf{A}^{-1}\mathbf{X}^T)_{ii}$. For sample i belonging to bin k, this leverage is the bin's inverse effective sample size, $h_{ii} = \alpha_k^{-1}$. The optimal value of p can then be determined by evaluating the objective function:

$$O(p) = \sum_{i=1}^{N} \left(\frac{z_i^{(p)} - \hat{z}_i^{(p)}}{1 - h_{ii}} \right)^2 + \left(\hat{\boldsymbol{\omega}}^{(p)}\right)^T \boldsymbol{\Lambda} \hat{\boldsymbol{\omega}}^{(p)}. \tag{12}$$

The first term is the LOOCV error, and the second is a parsimony penalty on model capacity, explicitly favoring solutions driven by lower-tier effects. The penalty is weighted by the prior precision $\boldsymbol{\Lambda}$, which is not a tuned quantity but exactly the weight the marginal likelihood of the conjugate model in App. F.3 places on the same quadratic form. In practice, we evaluate this objective over a grid of values ($p \in [0, 1]$) and select the minimum to determine p dynamically. Fig. 2 shows the result on the four benchmarks. No trial on any benchmark selects the additive $p = 1$, and the means sit nearer the multiplicative end throughout.

This choice of objective function has two additional motivations:

- **Equivalence to Maximum Likelihood:** The Jacobian normalization ensures that $\sum_i \log |dz_i^{(p)}/dU_i|$ vanishes, thus the residuals at every p constitute a valid likelihood for the observed $\mathbf{U}$, rather than just for the transformed targets. Applying $-\frac{N}{2}\log$ to the LOOCV term yields the cross-validated Gaussian log-likelihood up to constants independent of p, so minimizing that term is cross-validated maximum likelihood. The parsimony penalty is a preference to favor lower-tier-based explanations, a preference not expressed by using PRESS alone. Because $\mathbf{X}$ and $\boldsymbol{\Lambda}$ are independent of p, the leverages h_{ii} and the leave-one-out correction they define are as well. The candidates on the grid are therefore models of identical complexity differing only in their targets, and no value of p can be favored by a more lenient accounting of its own degrees of freedom.
- **Sample-Level vs. Bin-Level Holdout:** The transform parameter p does two things at once: it stabilizes the dispersion of the observations, which shows up in the within-bin scatter, and it sets the scale on which the bin effects combine, running from additive at $p = 1$ to multiplicative at $p = 0$, which shows up between bins. A criterion for p must therefore respond to both, and the sample-level residual does, carrying the within-bin deviation together with the bin-mean discrepancy at the weight b_k/α_k the fitted model assigns it, where b_k is the borrowed strength (14). The leave-one-out weight $(1 - h_{ii})^{-2} = (\alpha_k/(\alpha_k - 1))^2$ depends on the bin solely through α_k and stays strictly below four once a bin holds two samples. Hold-

ing out entire bins leaves the within-bin term untouched, but takes the bin-mean discrepancy as the raw parent-only error, inflating its contribution by $(\alpha_k/b_k)^2 = (1 + n_k/b_k)^2$. Because b_k is capped at $b_{\max}$ (17), this weight grows without bound as the bin population n_k increases. The objective is thus dominated by heavily populated yet weakly supported bins, precisely where the parent-only prediction is least reliable. Neither criterion measures downstream acquisition quality directly; App. D compares the selected p against the fixed additive choice $p = 1$ on that basis.

B Details of datasets

We expand on the detail of the evaluation for our datasets. The dataset construction and train/OOD splits are fixed across all trials and methods. MotionSense is the exception, retaining every window rather than being downsampled to balance. We note that the pretrained MantisV2 did not use any of these datasets for training (Feofanov et al., 2026). Importantly, including the target label as a design dimension is typical of experimental environments, so we replicate that here for all but SimGlucose. Particularly because of the failure of the introduction's condition (3) — the absence of a reliable *post hoc* oracle — the class would have to be determined at collection time for reliable labeling. Were SimGlucose a physical study, the label would have to be established by continuous glucose monitoring at collection time. MotionSense and MuJoCo can place the participant or robot in the demanded situation. Sleep-EDF was produced with *post hoc* labeling with expected noise (Rosenberg & Van Hout, 2013; Cesari et al., 2021; Lee et al., 2022).

SimGlucose This simulator (Man et al., 2014; Xie, 2018) is a synthetic data factory for Type-I diabetes, with 30 virtual patients distributed equally across 3 patient groups. Episodes are generated on demand by the simulator, so the zero-pool evaluation is never limited by a dataset ceiling. Metadata are 3 patient groups (adolescent, adult, child), 5 meal patterns (regular, late dinner, snacky, skip breakfast, high variability), 3 carbohydrate levels (low, medium, and high, corresponding to 0.52, 1.00, and 1.75 times the meal-pattern-specific base carbohydrate amounts, respectively), and 2 sensor-noise levels (clean and noisy; the latter adds Gaussian CGM noise with $\sigma = 7.00/7.56/8.05$ mg/dL for adults/adolescents/children, respectively), giving 90 bins. The simulator uses one-minute internal steps, a three-minute Dexcom interval, and a basal-bolus controller targeting 140 mg/dL. Each episode has 512 points and 5 channels (CGM, BG, CHO, insulin, and ΔCGM). The label categories are safe, hypoglycemic, hyperglycemic, or unstable, constructed post hoc using predefined rules applied to the final CGM time series. For this benchmark, the label is not part of the design. All five methods share a pre-generated, balanced training pool of 2,160 episodes (24/bin), and the OOD test contains 1,620 episodes from patients #008, #009, and #010 in each of the child, adolescent, and adult groups. The sampling begins at 90, and undergoes 9 rounds of increasing by 30 samples up to 360.

MuJoCo This quadruped locomotion simulator is built on the Gymnasium `Ant` model (Towers et al., 2023) in MuJoCo (Todorov et al., 2012), and the task is to classify the friction of the ground surface from the robot's motion alone. An experiment is a single 5-second rollout of a fixed, friction-agnostic open-loop gait (a 1.5 Hz sinusoid with fixed per-actuator phase offsets, identical in every bin), integrated at a 10 ms timestep with a control decimation of 5, giving 100 control steps at 20 Hz. Metadata are leg length (all leg segments scaled by 0.70, 1.00, or 1.30), torso density (5, 40, 150 kg m^{-3}), surface gravity (9.81, 3.72, 1.62, 0.82, 0.65, 0.28 m s^{-2}, named for Earth, Mars, Moon, Eris, Pluto, and Ceres), and ramp inclination (-5, -2.5, 0, 2.5, 5 degrees), giving 270 distinct chassis-environment configurations. The label is the surface friction itself — ice, tile, or rubber, with sliding coefficients 0.05, 0.20, and 1.50 — so that treating it as a design category yields 810 bins. For the AL baseline, we produce a pool of 19,440 samples (24/bin). Because a ramp on an infinite plane is dynamically equivalent to a tilt of the gravity vector, inclination is implemented as a rotation of gravity rather than a change of contact geometry, and the sliding coefficient is set on both the foot and floor geoms, as MuJoCo combines contact friction as the elementwise maximum of the two contacting surfaces. Every rollout is nuisance-randomized, with hinge angles drawn uniformly within their limits and initial velocities from $\mathcal{N}(0, 0.03^2)$, so that repeated queries to a bin return fresh, non-repeating experiments that remain reproducible from a single master seed. This benchmark is therefore a true zero-pool data factory: any bin may be queried arbitrarily often and the dataset ceiling is never reached. Classification uses the torso linear and angular velocities and the eight hinge joint angles. Contact-force channels are deliberately withheld, as the normal and

tangential foot forces are a near-direct readout of the friction coefficient and would reduce the task to a trivial one. The sampling begins at 810, and undergoes 9 rounds of increasing by 405 samples up to 4455.

MotionSense This dataset (Malekzadeh et al., 2019) contains 24 participants with documented age, gender, height and weight, performing six activities (walking, jogging, climbing upstairs, climbing downstairs, sitting, and standing) in different environments with a smartphone in their front pocket recording accelerometer and gyroscope readings at 50 Hz. The dataset is deconstructed into 10.24 s windows with an 11.24 s stride (a 1 s gap), after trimming 3 s from each end of each recording. At most 12 windows are retained per recording, selected at evenly spaced candidate indices when more are available, yielding 1,995 windows in total. This framework is almost faithful to Section 3.1's premise, as the participant could be instructed to do the task that results in the sample. On the other hand, the data are very limited and a bit imbalanced $\approx 2:1$ for some classes. The MotionSense experiment uses 24 initial samples followed by 6 acquisitions of 12 up to 96. Metadata are gender, weight-to-height-ratio category (weight in kg divided by height in cm, grouped into < 0.38, $[0.38, 0.45]$, and > 0.45), age category (< 28 or ≥ 28 years), and activity, which is the target label. Due to the small number of participants, all 1,995 retained windows from 24 subjects are used in repeated subject-level four-fold evaluation. Each outer fold has 18 acquisition subjects and 6 held-out subjects. Predictions from all four held-out folds are pooled before calculating macro-F1, giving one score per seed, and not four independent runs. While in this construction all bins contain more than one participant, no effort was made to ensure the training set contains at least one participant per bin. Classes are only a little imbalanced, but we explicitly do a balanced random hierarchical sampling (uniform draw of class, then uniform draw of sample) to unbias the ZALMBIES comparison with the random baseline. Input entropy scores the entire remaining pool, so the mild imbalance falls on the AL baseline alone, and it is unclear whether that helps or hurts it, as balancing would instead cost it roughly a third of its pool. As noted in Section 5, this is only illustrative, so we accept the imbalance for simplicity.

Sleep-EDF This dataset (Kemp et al., 2000) contains 78 participants recorded for 2 nights each, 153 full nights in total (3 participants had only 1 night due to collection errors). A subsequent medicated telemetry study with 22 participants (Kemp et al., 2000) was not included. These full nights are broken into 30s staging clips (with N3 and N4 combined (Silber et al., 2007)). In our evaluation we view the individual clips as independently acquirable units, as this allows for an expansion of our dataset ceiling. Our classification label is the sleep stage itself, which we include here as a targetable design category. To allow for random sampling and AL benchmarks to have a fair comparison with ZALMBIES, we explicitly downsample our data to be equally stratified across all bins. The trimmed source contains 195,479 samples. The training set contains 9,990 samples, balanced across 90 bins (111/bin) defined by age, sex, sleep stage, and relative time since lights-off. The untouched OOD test contains 43,212 samples from 33 nights (17 participants) for test evaluation. The held-out participants are those with IDs SC62-SC67, SC70-SC77, and SC80-SC82. Metadata are age (40 or younger, 41 to 60, older than 60), gender, the 5 sleep stages, and 3 relative-time bins measured from lights-off ($0 \leq t < 2$, $2 \leq t < 4$, and $t \geq 4$ hours). The sampling begins at 180, and undergoes 9 rounds of increasing by 90 samples up to 990.

Sleep-EDF is by far the least aligned benchmark in the suite. The sleep stage has no reliable labeling instrument at all: expert scoring is notoriously unreliable, which is the failure of condition (3) of Section 1, and the labeling-at-collection precondition of Section 3.1 offers no escape, since a collection-time label would come from the same scoring of the same recording. All sampling methods cope with the same labels, so comparison remains valid. The noise doubles as a stress test of a specific vulnerability in that $\hat{U}$ is fit to realized failure. By Proposition 6, ZALMBIES sends budget toward high-$\hat{U}$ bins, and those with error that no sampling can reduce will demand more. The clip-level treatment is a well-posed instance of the setting and the dense but unfaithful branch of the choice laid out in Section 4, where the sleep stage becomes a design dimension that no real study could control. Fig. 3 takes the other branch, restricting the design to what a study does control and leaving 12 bins rather than 90, so the query becomes an entire participant-night of recording as the original study ran it. That changes the design, not the labeling, and the savings achieve 25%. The two branches are not on the same footing. A night is indivisible: one acquisition commits several hundred clips of one participant to one bin, so the policy cannot spread across bins or trim a bin it has overshot. After initialization the clip-level run makes 810 steerable acquisitions over 90 bins while the night-level run makes 18 over 12, and no acquisition rule can express a fine allocation through

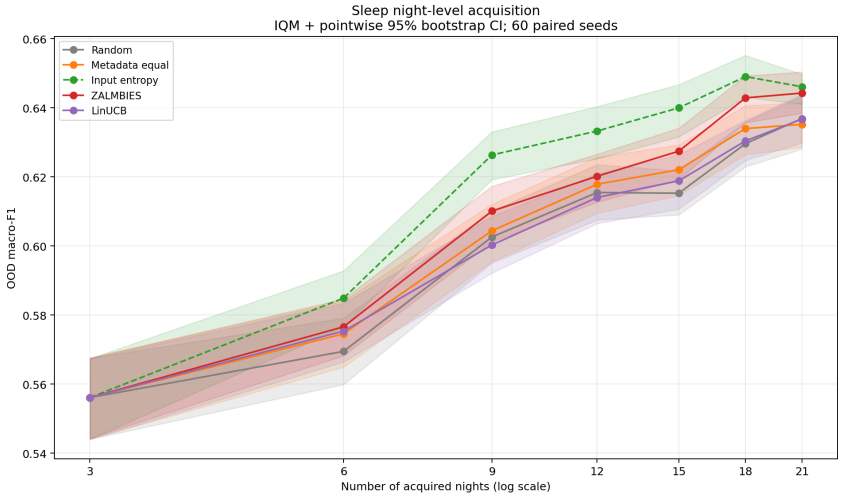


Figure 3: The performance of ZALMBIES on Sleep-EDF using participant-night level sampling. Here, the algorithm starts with 3 participant-nights, and undergoes 6 rounds of increasing by 3 additional nights up to 21. For this examination, the following were used: age (40 or younger, 41 to 60, older than 60), gender, and two bedtime bins ($<$ / $\geq$ 11:30 pm). Despite a corpus of only 120 participant-nights across 12 bins, ZALMBIES provides a 25% sample budget savings relative to random. The design matrix is constructed at the night-level, and U_j assigned the mean of root-Brier for the samples in each night. For input entropy, we sample based on the mean of uncertainties of all clips for the night. As before, the pool-based AL baseline (dashed) is not solving the same problem. LinUCB performs comparably with random.

18 all-or-nothing choices. This is a real experiment, but it is not the zero-pool setting of Section 3.1, in which a query returns a single draw the policy may place freely. We report it because it is the honest test of restricting the design to what the sleep study controls, and a 25% savings under that much coarseness indicates ZALMBIES degrades gracefully outside the setting rather than telling us what the setting delivers.

C POOL-BASED ACTIVE LEARNING

While our focus is on the zero-pool scenario, there is a simple extension to the method that enables standard pool-based active learning. In short, bin selection proceeds as above, but an extant pool of unlabeled data is assigned to the bins. Especially in environments with sparsely populated metadata bins, the scoring can be viewed as a bin ranking, allowing for, e.g., the algorithm to greedily select a sample from the top scored bin that is not empty. The top-scoring bin could contain just a single, low-entropy sample, so a more valuable usage selects several bins, until either the number of bins or the volume of data exceeds a threshold, for instance $N_{\text{samples,selected}} > N_{\text{samples,total}}^{1/2}$, and labels the highest-entropy sample in their union.

In our experiment, we apply the algorithm as designed, but rather than drawing from the top bin as the zero-pool rule does, we grab the top bins until the number of samples exceeds the square root of the pool size $N_{\text{samples,total}}^{1/2}$, and select the highest entropy sample from within their union. For the pool-based collection, we compare ZALMBIES modified for pool-based AL against random sampling, an equal allocation over metadata bins, and four AL baselines: (1) input entropy, (2) BADGE (Ash et al., 2020), (3) BatchBALD (Kirsch et al., 2019), (4) BAIT (Ash et al., 2021). We use the same SimGlucose benchmark as the zero-pool case, with a pool of 24 samples per bin that is not class balanced. All other settings are identical to that experiment (App. B): 24 seeds, the same budget schedule from 90 to 360 in steps of 30, and the same base model and evaluation protocol.

Table 2 reports the comparison and Fig. 4 shows the curves. The extension is indistinguishable from BADGE, BatchBALD, and greedy entropy at 24 seeds, every paired interval covering zero with a

Table 2: Pool-based AL on SimGlucose. The area ratio, savings, and ΔF_1 columns are defined exactly as in Table 3, but normalized to the pool-based ZALMBIES extension rather than to the zero-pool method. ΔA is the paired seedwise difference in area against that extension, with a 95% bootstrap interval over 10000 resamples. The metadata-equal comparator is omitted from Fig. 4 for legibility; its run is the one shown for SimGlucose in Fig. 1.

Method	Area ratio	Savings	ΔF_1	ΔA
ZALMBIES (pool)	1.000	70.6%	$--$	$--$
BAIT	1.236	74.9%	$+0.50$	$+0.008^{+0.007}_{-0.007}$
BADGE	1.135	78.7%	$+0.08$	$+0.004^{+0.010}_{-0.009}$
BatchBALD	0.905	64.2%	-0.50	$-0.003^{+0.006}_{-0.006}$
Input entropy	0.938	67.5%	$+0.30$	$-0.002^{+0.008}_{-0.007}$
Metadata equal	0.083	25.9%	-2.62	$-0.029^{+0.008}_{-0.008}$

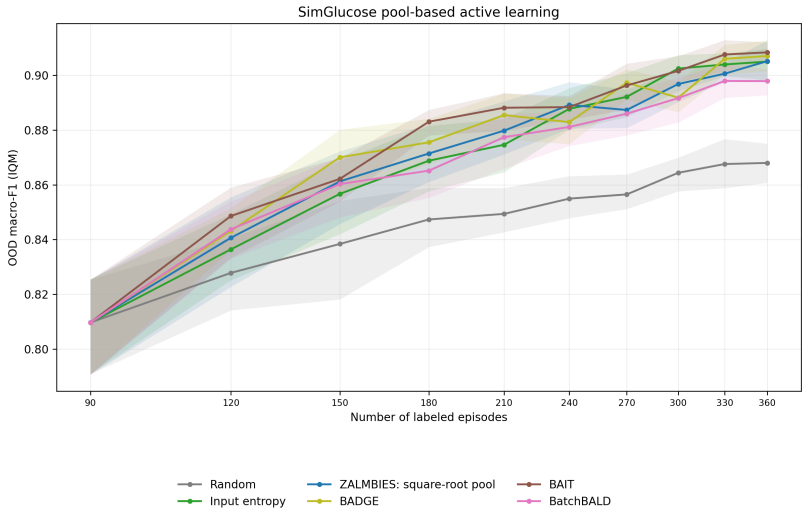


Figure 4: The performance of ZALMBIES modified for pool-based AL for the SimGlucose dataset. We compare to several powerful AL methods: BADGE, BAIT, BatchBALD, and input entropy, as well as random sampling.

width well inside the spacing of the methods. BAIT is ahead of it by a resolvable but small margin, roughly half a point of macro-F_1 averaged across the schedule, and BatchBALD is the weakest of the four. Replacing the bin scores by an equal allocation over bins collapses the result, so it is the ranking and not the assignment of a pool to bins that carries the method. We stress that this is only intended as a secondary result of this work. Optimizing ZALMBIES for pool-based AL is a promising research direction beyond the scope of this work.

D ABLATIONS

In Table 3, we present different ablations on the method in terms of three numbers. The first is $A_{\text{ablation}}/A_{\text{standard}}$, where A_{ablation} is the area between the ablated curve and random and A_{standard} is the area between the ZALMBIES curve and random, both on the central IQM curve in log sample count. The second is the savings, computed exactly as in Table 1 but with the ablated curve in place of ZALMBIES, taking the smallest budget from which the curve remains above random's terminal IQM; an ablation that never reaches it is reported as < 0. The third, ΔF_1, is the absolute change in central IQM between the ablation and ZALMBIES, averaged over the final three budgets in Fig. 1. The area and the savings summarize integrated and horizontal gaps respectively,

Table 3: Different modifications to the selection and how they impact overall performance. Benchmarks omitted from a row were not run for that ablation.

Ablation	Benchmark	$A_{\text{ablation}}/A_{\text{standard}}$	Savings	ΔF_1
$\beta = 0$	SimGlucose	0.357	36.9%	−1.73
	MotionSense	−1.751	< 0	−2.62
	Sleep-EDF	−0.436	< 0	−2.87
$\beta = 0.5$	SimGlucose	0.894	55.4%	−0.55
	MotionSense	0.496	< 0	−1.09
	Sleep-EDF	0.844	50.0%	−0.36
$\beta = 2$	SimGlucose	0.942	64.2%	−0.12
	MotionSense	1.091	63.5%	+0.23
	Sleep-EDF	1.118	57.8%	+0.01
γ_{dynamic}	SimGlucose	0.973	66.9%	+0.15
	MotionSense	1.025	59.9%	+0.42
5→3 metadata	MuJoCo	0.240	38.2%	−0.68
BC $p = 1$	SimGlucose	1.019	66.6%	+0.15
	MotionSense	1.107	71.7%	−0.15
	Sleep-EDF	0.930	51.3%	−0.20
$\hat{U}_{\text{class deficit}}$	SimGlucose	1.065	71.3%	+0.12
	MotionSense	0.844	66.7%	−0.17
	Sleep-EDF	0.939	50.8%	+0.11
$\hat{U}_{\text{entropy}}$	SimGlucose	0.969	62.8%	+0.05
	MotionSense	0.340	32.6%	−0.55
	Sleep-EDF	0.986	59.7%	−0.25
LinUCB	SimGlucose	0.899	64.2%	−0.52
	MuJoCo	0.538	46.5%	−0.43
	MotionSense	0.692	38.8%	−0.22
	Sleep-EDF	1.102	61.2%	+0.20
$S = \hat{U}_{\text{main}}/\alpha$	SimGlucose	0.960	68.1%	−0.06
	MuJoCo	0.031	32.7%	−0.52
	MotionSense	1.045	68.6%	−0.08
	Sleep-EDF	1.301	64.6%	+0.96
$S = \hat{U} + c_0\,\alpha^{-1/2}$	SimGlucose	0.607	50.2%	−0.27
	MotionSense	0.488	49.4%	−0.24
	Sleep-EDF	0.752	49.7%	−0.31
$S = \hat{U}_{\text{main}}/\alpha_{\text{main}}$	SimGlucose	0.854	68.4%	−0.43
	MotionSense	0.512	27.1%	−0.59
	Sleep-EDF	1.066	62.1%	+0.57

and so need not order the ablations identically. All ablations are evaluated with 24 trials using only that modification.

Table 3 groups these modifications into three blocks that do not carry the same implication. The first varies choices the paper hands to the practitioner, the second substitutes a component of the scoring rule, and the third varies how that rule combines its two ingredients and over which sublattice. No configuration is best on every benchmark, and nothing before experiments are run would identify the best choice, so the defaults were fixed *a priori* and held throughout. The table is evidence that this choice does consistently well, but not that it is optimal everywhere.

The first block varies the quantities that Section 3.3 and App. E leave to the user: the exponent β, the prior weight γ, and the lattice itself. We examine $\beta = 0, 0.5, 2$ (pure exploitation, predictive-standard-deviation, and V-optimally motivated). For γ_{dynamic}, we grow γ with increasing N, $\gamma_{\text{dynamic}} = (cN/L_{\text{eff}}^{t^*})^{1/(t^*-1)}$, where L_{eff} is the geometric mean of dimension cardinalities, and t^* is the target maximum tier taken here to be 4, which slows the progress of tier 2, 3, and 4 while killing higher order effects. We also reduce the MuJoCo design from 5 metadata dimensions to 3 by dropping control over incline and body mass (which are then randomly selected). These measure

sensitivity to a setting rather than to the rule. The collapse at $\beta = 0$ is not surprising as this greedily samples a batch from one bin, and it falls below random on two of the three benchmarks where we run it. $\beta = 0.5$ fares much better, but is not clearly competitive with $\beta = 1$. $\beta = 2$ is the better setting on MotionSense and Sleep-EDF, which is a choice between named points of the family rather than a badly chosen default. $\beta = 1$ has two theoretical motivations, as the first-order form of exact information gain (Remark 3) and as Neyman allocation (Remark 5), while $\beta = 2$ has one, as a greedy step on the $\hat{U}$-weighted predictive variance (Remark 4). We therefore present $\beta = 2$ as a very reasonable alternative, especially when more exploration is motivated, rather than defend $\beta = 1$, our default, as optimal. γ_{dynamic} helps mildly on MotionSense and is neutral on SimGlucose, and we do not adopt it because it adds a schedule without a consistent gain. Dropping two of MuJoCo's five dimensions is the most damaging change in the table on that benchmark, which is the empirical counterpart of App. E: the lattice design is a critical choice.

The second block substitutes a core component of the scoring rule. Setting BC $p = 1$ reduces to the linear model, removing the Box–Cox geometric adaptations, and we also examine $\hat{U}_{\text{class deficit}} = 1 - P(y = y_{\text{true}})$ and $\hat{U}_{\text{entropy}}$ in place of square-root Brier. BC $p = 1$ is not clearly separated from the default on any benchmark; selecting p online removes a tuning choice rather than beating a fixed one. Neither alternative uncertainty measure separates from square-root Brier except on MotionSense, where entropy is far worse.

The third block varies two choices at once: whether $\hat{U}$ and α combine as a ratio or as an additive confidence bonus, and whether the design is the full lattice or main effects alone. Adding $S = \hat{U} + c_0\,\alpha^{-1/2}$, with $c_0 = 2\beta\,\bar{\hat{U}}\,\bar{\alpha}^{1/2}$, and $S = \hat{U}_{\text{main}}/\alpha_{\text{main}}$ completes a 2×2 grid with ZALMBIES in one corner and our LinUCB-motivated baseline (see Fig. 1 for the full curve) in the opposite one. These variants beat the default by at most a modest amount, but can miss sizably. There is a further generalization construction, $S = \hat{U}_{\text{main}}/\alpha$, that uses a main-effects only $\hat{U}$ against the full-lattice α. This variant decisively wins somewhere and loses somewhere, exceeding ZALMBIES on Sleep-EDF by more than LinUCB does while on MuJoCo, the largest design in the suite at 810 bins, its first batch falls significantly below random before recovering to roughly split the difference between random and ZALMBIES. Its area ratio there is accordingly set almost entirely by that one budget, and is not a level of performance. Sleep-EDF is also where $\beta = 2$ is most effective, and this alignment may be explained by Proposition 6. Because the asymptotic allocation $n_k \propto \hat{U}_k^{1/\beta}$ responds just as strongly to the part of $\hat{U}$ that no sampling can remove as to the part it can, both increasing β and restricting $\hat{U}$ to main effects should dampen this sensitivity. We keep the full-lattice $\hat{U}$ as the default, but the main-effects surrogate is worth considering when higher tiers are not expected to be important.

E DESIGNING THE METADATA LATTICE

The lattice is the one part of ZALMBIES that is not supplied by the algorithm, and its shape sets both the resolution the surrogate can attain and the exploration the acquisition rule must pay for. The goal is to sample with designs that are not too broad (2–10 categories) and reasonably deep (3–7 dimensions). Very broad categorical dimensions, e.g., ZIP codes or subject ID, are not terribly useful for the lattice, and risk creating runaway exploration, as 100 categories would all need to be sampled multiple times for the method to formulate robust statistics. These are better compressed into one to three low-cardinality dimensions, e.g., subject ID into gender, age, and height. All of this presumes the dimensions are informative, as a split that carries no signal will not hold in unseen data. Scenarios with many low-frequency binary categorical settings can also create problems when the settings can co-occur but rarely do. With 20 binary settings of which no more than two are ever active at once, the design space balloons to roughly 3.5 billion columns, wasting exploration on deep combinations that add minimal value. Ideally, it would be possible to structure these into a few narrow categories to facilitate exploration. Section 5 treats the structural cases that the product encoding cannot absorb at all.

F Information-Theoretic and Optimal-Design Foundations

The acquisition rule of Section 3.3 was introduced heuristically, with denominator α_k^β. We now establish what it computes exactly. The key observation is that the quantities the rule already computes, the effective sample size α_k and its rank-one updates, carry exact optimal-design and information-theoretic meaning, with no sampling or embedding approximation. Most of what follows requires no noise model at all. The decomposition of α_k is an identity of the encoding, and the optimality of the balanced design follows from the symmetry of the categorical lattice together with the concavity of the design criteria. We establish, in order: (i) α_k splits exactly into a bin's own count plus the strength it borrows from its parent lattice, which no acquisition can diminish (Proposition 1) and which the prior caps (Proposition 2); (ii) at cold start the acquisition rule is β-independent, every selection it makes is exactly D-optimal (Theorem 3), the greedy trajectory carries a $(1 - 1/e)$ guarantee for D-optimality (Remark 2), and the balanced allocation is optimal under many classical design criteria (Theorem 4); (iii) that same determinant gain is exactly the batch mutual information under the conjugate model, with the aleatoric scale cancelling, and the sequential update avoids the redundancy of independent scoring (Theorem 5, (32)); and (iv) β is a dial between exploitation and balance whose large-sample allocation Proposition 6 gives exactly, although its interior values recover familiar criteria only under stated idealizations (App. F.4). We distinguish throughout between statements that hold exactly for the rule as implemented and identifications with named criteria from the literature that hold only approximately. Every identification of the second kind is recorded as a remark.

Throughout, $\mathbf{x}_k \in \{0,1\}^M$ is the one-hot ANOVA encoding of bin k, with 2^D nonzeros of which $2^D - 1$ lie on parent-lattice columns shared with other bins, and $\mathbf{\Lambda}$ is the hierarchical prior precision of Section 3.2. We write $\mathcal{K}$ for the acquisition lattice, which is the bin set of Section 3.2 except where Remark 1 states otherwise. For a pending batch with design rows $\mathbf{F}$ we write

$$\mathbf{A}(\mathbf{F}) = \mathbf{X}^T\mathbf{X} + \mathbf{F}^T\mathbf{F} + \mathbf{\Lambda}, \qquad \alpha_k(\mathbf{F}) = \left(\mathbf{x}_k^T \mathbf{A}(\mathbf{F})^{-1}\mathbf{x}_k\right)^{-1}, \tag{13}$$

abbreviating $\mathbf{A} = \mathbf{A}(\mathbf{0})$ and $\alpha_k = \alpha_k(\mathbf{0})$, so that $\mathbf{A}(\mathbf{F})$ is the Bayesian information matrix the rule maintains and $\alpha_k(\mathbf{F})^{-1}$ the associated unscaled predictive variance at bin k. Every such matrix is positive definite, since $\mathbf{X}^T\mathbf{X}$ and $\mathbf{F}^T\mathbf{F}$ are positive semidefinite and $\mathbf{\Lambda} \succ \mathbf{0}$ by $\lambda_0 = \epsilon > 0$ and $\lambda_t = \gamma^{t-1} > 0$, as is the reduced matrix $\mathbf{R}_k$ introduced in Proposition 1, so every inverse written below is well defined.

F.1 What the effective sample size measures

Proposition 1 (Exact decomposition of the effective sample size)**.** *Suppose every sample acquired in bin $k \in \mathcal{K}$ contributes the same design row $\mathbf{x}_k$ to the information matrix against which α_k is evaluated. Let n_k be the number of acquired samples in bin k, and let $\mathbf{R}_k = \mathbf{X}_{-k}^T\mathbf{X}_{-k} + \mathbf{\Lambda}$ be the Bayesian information matrix formed from all data* except *for the samples in bin k. Then*

$$\alpha_k = n_k + b_k, \qquad b_k = \left(\mathbf{x}_k^T\mathbf{R}_k^{-1}\mathbf{x}_k\right)^{-1} > 0, \tag{14}$$

where b_k is the statistical strength borrowed from the surrounding parent lattice. Moreover, acquiring one additional sample in bin k increases α_k by exactly one, while acquiring a sample in any other bin cannot decrease it.

Proof. By hypothesis every sample in bin k contributes the identical row $\mathbf{x}_k$, so $\mathbf{A} = n_k\mathbf{x}_k\mathbf{x}_k^T + \mathbf{R}_k$. Writing $\tau_k = \mathbf{x}_k^T\mathbf{R}_k^{-1}\mathbf{x}_k$ and applying the Sherman–Morrison identity (Sherman & Morrison, 1949),

$$\mathbf{x}_k^T\mathbf{A}^{-1}\mathbf{x}_k = \tau_k - \frac{n_k\tau_k^2}{1 + n_k\tau_k} = \frac{\tau_k}{1 + n_k\tau_k} = \frac{1}{n_k + \tau_k^{-1}}, \tag{15}$$

so $\alpha_k = n_k + \tau_k^{-1}$. Since $\mathbf{R}_k$ excludes bin k's own samples, $b_k = \tau_k^{-1}$ is invariant to n_k, and $n_k \mapsto n_k + 1$ gives $\alpha_k \mapsto \alpha_k + 1$. For an acquisition in any bin i, write $\mathbf{A}_{+i} = \mathbf{A} + \mathbf{x}_i\mathbf{x}_i^T$. An application of Sherman–Morrison gives

$$\mathbf{x}_k^T\mathbf{A}_{+i}^{-1}\mathbf{x}_k = \mathbf{x}_k^T\mathbf{A}^{-1}\mathbf{x}_k - \frac{\left(\mathbf{x}_k^T\mathbf{A}^{-1}\mathbf{x}_i\right)^2}{1 + \mathbf{x}_i^T\mathbf{A}^{-1}\mathbf{x}_i}, \tag{16}$$

whose subtracted term is non-negative because $\mathbf{A} \succ \mathbf{0}$, so α_k cannot decrease. As n_k is unchanged when $i \neq k$, and rises by one alongside α_k when $i = k$, the borrowed strength b_k will never decrease as bins fill. □

Borrowing is not unbounded, but capping it takes more than the identity above. Where (14) constrains only the rows of bin k and holds for any positive prior, the cap below constrains the whole design, which must be drawn from the bins of a product lattice, and the prior, which must be the hierarchical $\mathbf{\Lambda}$ of Section 3.2 with $\epsilon \leq \gamma^{-1}$.

Proposition 2 (Cap on borrowed strength)**.** *Let $\mathbf{R}_k$ and b_k be as in Proposition 1, let the acquisition lattice be $\mathcal{K} \subseteq \prod_{d=1}^{D}\{1,\dots,m_d\}$, and suppose every row of $\mathbf{X}_{-k}$ is the encoding $\mathbf{x}_j$ of some bin $j \in \mathcal{K} \setminus \{k\}$. Let $\mathbf{\Lambda}$ be the hierarchical prior of Section 3.2, and suppose its two parameters satisfy $\epsilon \leq \gamma^{-1}$. Then borrowing is bounded above by a constant independent of the data,*

$$0 \;<\; b_k \;\leq\; b_{\max} \;:=\; \gamma^{\,D-1} \prod_{d=1}^{D} \frac{m_d - 1 + \gamma^{-1}}{m_d + \gamma^{-1}} \;<\; \gamma^{\,D-1} = \lambda_D. \tag{17}$$

Proof. Write $\tau_k = b_k^{-1}$. Cauchy–Schwarz in the $\mathbf{R}_k$ inner product gives, for any $\mathbf{u}$,

$$\tau_k \;=\; \mathbf{x}_k^T \mathbf{R}_k^{-1} \mathbf{x}_k \;\geq\; \frac{(\mathbf{x}_k^T \mathbf{u})^2}{\mathbf{u}^T \mathbf{R}_k \mathbf{u}}, \tag{18}$$

and we choose a $\mathbf{u}$ orthogonal to every row of $\mathbf{X}_{-k}$, so that the data term vanishes from the denominator and leaves $\mathbf{u}^T \mathbf{\Lambda} \mathbf{u}$. Each column of the encoding is one interaction term, specified by naming, for every dimension d, either a level $z_d \in \{1,\dots,m_d\}$ or the symbol $z_d = *$ meaning that the dimension does not appear in the term. Columns are therefore indexed by $z \in \prod_d \{*, 1, \dots, m_d\}$, the tier of z is the number of dimensions it names, and bin j activates column z exactly when every level z names agrees with j, that is when $z_d \in \{*, j_d\}$ for all d. Take

$$u_z = \prod_{d=1}^{D} a_{z_d}^{(d)}, \qquad a_*^{(d)} = 1, \quad a_{k_d}^{(d)} = m_d - 1 + \gamma^{-1}, \quad a_i^{(d)} = -1 \quad (i \neq k_d). \tag{19}$$

For any bin j, $\mathbf{x}_j^T \mathbf{u} = \prod_d (a_*^{(d)} + a_{j_d}^{(d)})$, whose dth factor vanishes whenever $j_d \neq k_d$, so $\mathbf{u}$ annihilates every bin but k. No row of $\mathbf{X}_{-k}$ contributes, $\mathbf{u}^T \mathbf{R}_k \mathbf{u} = \mathbf{u}^T \mathbf{\Lambda} \mathbf{u}$, and $\mathbf{x}_k^T \mathbf{u} = \prod_d (m_d + \gamma^{-1})$. For the denominator, $\mathbf{\Lambda}$ weights a tier-t column by $\gamma^{\,t-1}$, so it is γ^{-1} times the Kronecker product of the per-dimension diagonals $\mathrm{diag}(1, \gamma, \dots, \gamma)$, except where $\lambda_0 = \epsilon$ stands in place of γ^{-1}. Since $\epsilon \leq \gamma^{-1}$, replacing ϵ by the larger γ^{-1} can only raise the quadratic form, so using γ^{-1} throughout and applying $1 + \gamma(m-1) + \gamma(m - 1 + \gamma^{-1})^2 = \gamma\,(m - 1 + \gamma^{-1})(m + \gamma^{-1})$ to each factor,

$$\mathbf{u}^T \mathbf{R}_k \mathbf{u} \;\leq\; \gamma^{-1} \prod_{d=1}^{D} \Big[1 + \gamma(m_d - 1) + \gamma\big(m_d - 1 + \gamma^{-1}\big)^2\Big] \;=\; \gamma^{\,D-1} \prod_{d=1}^{D} \big(m_d - 1 + \gamma^{-1}\big)\big(m_d + \gamma^{-1}\big). \tag{20}$$

Substituting both into (18) gives $\tau_k \geq \gamma^{\,1-D} \prod_d (m_d + \gamma^{-1}) / (m_d - 1 + \gamma^{-1})$, which is (17). □

Proposition 1 quantifies the phrase "borrowed statistical strength," and Proposition 2 bounds it. b_k is measured in units of samples and is exactly the excess of the effective sample size over the bin's own count. With the further fact that each acquired sample raises α_k by exactly one, the penalty is an explicit function of the bin's own count, $\alpha_k^{-\beta} = (n_k + b_k)^{-\beta}$, where diminishing returns in n_k are offset by borrowing of at most $b_{\max}$ effective samples (17). This places the prior on the scale of the data: $b_{\max}$ is the occupancy beyond which a bin's own samples necessarily outweigh everything its parents can lend it, so γ is set by the count at which one is willing to resolve a deepest-tier effect rather than defer to its parents. That occupancy is approached only as the rest of the lattice is sampled without bound, so $b_{\max}$ does not mark a budget at which the lattice is exhausted. App. D applies the same idea one tier shallower, equating the prior precision λ_{t^*} to the occupancy of a tier-t^* cell and solving for the implied γ.

Remark 1 (Acquisition depth). Proposition 1 requires that every sample acquired in bin k enter the information matrix through the same row $\mathbf{x}_k$. This holds whenever the acquisition lattice and

the surrogate sit at the same depth, i.e., the full lattice and full model, or a marginal sub-lattice of bounded interaction order, provided α_k is evaluated against that sub-lattice's own information matrix, in which case n_k is the count aggregated over the bin's sub-cells and b_k the strength borrowed from its lower-tier parents. It fails when the depths are mismatched, that is when a marginal bin is scored against the full-depth information matrix. Samples assigned to marginal bin k then occupy different cells c and contribute distinct rows $\mathbf{x}_c = \mathbf{x}_k + \mathbf{z}_c$, with $\mathbf{z}_c$ on the complement columns, so the accumulated update has rank at most n_k instead and the proof does not hold. Even so, acquisition still raises α_k, but no longer by exactly one per sample, and (14) degrades from an identity to an approximation. Proposition 2 rests on the same matching, since its hypothesis that every row of $\mathbf{X}_{-k}$ encodes a bin of $\mathcal{K}$ fails once cells and bins sit at different depths, and we claim no cap in that case. Theorem 4 is unaffected either way, as it constrains the full design matrix, and a design balanced over cells is balanced over every marginal.

F.2 D-OPTIMALITY AT COLD START

In (10), setting $\hat{U} \equiv 1$ makes the score depend only on the effective sample size. This corresponds precisely to the *cold-start* regime of condition (1) of the Introduction, where no labeled data have been collected, no surrogate has been fitted, and $\hat{U}$ carries no information.

Two theorems are proved below, and they rest on different hypotheses. The first concerns the acquisition rule itself, and requires nothing beyond positive definiteness of $\mathbf{A}$. The second concerns the design that rule accumulates, and requires the symmetry of the lattice. For the first we recall the matrix determinant lemma (Harville, 1997), which states that for $\mathbf{A} \succ \mathbf{0}$ and any $\mathbf{F}$ with M columns,

$$\det\big(\mathbf{A} + \mathbf{F}^T\mathbf{F}\big) = \det(\mathbf{A})\, \det\big(\mathbf{I} + \mathbf{F}\mathbf{A}^{-1}\mathbf{F}^T\big)\,. \tag{21}$$

We treat $\mathbf{A}$ in this subsection purely as the Bayesian information matrix of Section 3.2. D-optimality is a criterion on that matrix and requires no observation model. Its identification with an *information* gain does require one, and is deferred to App. F.3.

Theorem 3 (β-invariance and D-optimal acquisition at constant $\hat{U}$)**.** *Fix $\hat{U}$ constant across bins, without loss of generality $\hat{U} \equiv 1$. For every $\beta > 0$ the acquisition rule (10) selects*

$$k^* \;=\; \arg\max_k \alpha_k(\mathbf{F})^{-\beta} \;=\; \arg\min_k \alpha_k(\mathbf{F}) \;=\; \arg\max_k \mathbf{x}_k^T \mathbf{A}(\mathbf{F})^{-1}\mathbf{x}_k, \tag{22}$$

so the entire greedy trajectory is independent of β. This selection also maximizes the one-step increase in $\log\det \mathbf{A}$, so each acquisition is the exact D-optimal augmentation of the design accumulated so far. Both claims hold from any state $\mathbf{F}$, on any lattice, and under any tie-break that does not depend on β.

Proof. For $\beta > 0$ the map $t \mapsto t^{-\beta}$ is strictly decreasing on $t > 0$, carrying the $\arg\min$ of α_k directly to an $\arg\max$. The two $\arg$ sets therefore coincide for every $\beta > 0$, including when they are not singletons, as at cold start where all α_k are equal by symmetry. Fixing any tie-break rule that does not depend on β, the selected bin is identical at the first step for all $\beta > 0$, yielding identical updates to $\mathbf{F}$. The claim follows by induction.

For the second claim, apply (21) with the single row $\mathbf{F} = \mathbf{x}_k^T$. The factor $\mathbf{I} + \mathbf{F}\mathbf{A}^{-1}\mathbf{F}^T$ is then the scalar $1 + \mathbf{x}_k^T\mathbf{A}^{-1}\mathbf{x}_k = 1 + \alpha_k^{-1}$, giving

$$\log\det\big(\mathbf{A}(\mathbf{F}) + \mathbf{x}_k\mathbf{x}_k^T\big) - \log\det \mathbf{A}(\mathbf{F}) \;=\; \log\big(1 + \alpha_k(\mathbf{F})^{-1}\big)\,. \tag{23}$$

The map $t \mapsto \log(1 + t^{-1})$ is strictly decreasing on $t > 0$, so the $\arg\min$ of α_k is exactly the $\arg\max$ of (23), the tie sets again coinciding. Neither claim uses the symmetry of the lattice as a hypothesis, nor the absence of previously acquired data, nor any property of $\mathbf{\Lambda}$ beyond $\mathbf{\Lambda} \succ \mathbf{0}$. □

Remark 2 (When the greedy guarantee applies). Write $f(\mathcal{S}) = \log\det \mathbf{A}(\mathbf{F}_\mathcal{S}) - \log\det \mathbf{A}$ for the determinant gain of a batch $\mathcal{S}$, with $f(\varnothing) = 0$. Since a batch may repeat a bin, the ground set is the $\mathcal{B}$ labelled copies of each bin rather than $\mathcal{K}$ itself, on which f is an ordinary set function. This set function is monotone, because adding a positive semi-definite rank-one update strictly increases the determinant. It is also submodular. No acquisition can decrease any α_j (Proposition 1, applied to $\mathbf{A}(\mathbf{F}_\mathcal{S})$), so iterating over $\mathcal{T} \setminus \mathcal{S}$ gives $\alpha_j(\mathbf{F}_\mathcal{S})^{-1} \geq \alpha_j(\mathbf{F}_\mathcal{T})^{-1}$ for $\mathcal{S} \subseteq \mathcal{T}$, which by

the marginal gain (23) is exactly the diminishing-returns condition. Greedy selection under a cardinality constraint $|\mathcal{S}| \leq \mathcal{B}$ therefore achieves at least a $(1 - 1/e)$ approximation of the optimal batch (Nemhauser et al., 1978), which no polynomial-time algorithm can improve under standard complexity assumptions (Feige, 1998). The bound holds at every budget $\mathcal{B}$, since it compares the greedy set against the best augmentation of the same size. At cold start $\mathbf{A} = \mathbf{\Lambda}$, so the comparison is against the best exact design of size $\mathcal{B}$. This argument uses only the determinant identity (23) and the positivity of $\mathbf{\Lambda}$. It needs no observation model, and f acquires its reading as an information gain only later, in App. F.3. The practical scope of the guarantee is narrow. It justifies a sequential rule in place of an exhaustive batch search, nothing more. $(1 - 1/e)$ is a worst-case bound and is often loose for $\log\det$, and it constrains only the accumulated determinant gain, making no statement about downstream task loss. The bound also requires $\hat{U} \equiv$ constant, where the greedy step is $\arg\min_j \alpha_j(\mathbf{F}_\mathcal{S})$. By Theorem 3 the acquisition rule (10) makes that selection for every $\beta > 0$, so the bound is inherited exactly. For informative weights the rule departs from the pure $\log\det$ objective and D-optimality is lost, so we claim no bound in that regime.

The lattice possesses a further symmetry, which determines where this rule comes to rest. Let $\mathcal{G} = \mathfrak{S}_{m_1} \times \cdots \times \mathfrak{S}_{m_D}$ be the group that independently permutes the choices within each metadata dimension. This group is represented on $\mathbb{R}^M$ by permutation matrices $\mathbf{P}_g$. Two critical properties of our construction follow:

1. **Prior Invariance:** The hierarchical prior depends only on the interaction order, not the specific choices. Thus, it is $\mathcal{G}$-invariant: $\mathbf{P}_g \mathbf{\Lambda} \mathbf{P}_g^T = \mathbf{\Lambda}$.
2. **Feature Equivariance:** Permuting choices permutes the one-hot columns of every interaction tier. The map $k \mapsto \mathbf{x}_k$ is therefore $\mathcal{G}$-equivariant: $\mathbf{P}_g \mathbf{x}_k = \mathbf{x}_{g(k)}$, where $g(k)$ is the relabeled bin.

Beyond these two, $\mathcal{G}$ permutes the bins among themselves transitively, since any choice tuple can be carried to any other, which forces the optimum to be uniform rather than merely constant on orbits.

To analyze this, we define the simplex of approximate designs as weight vectors over the bins:

$$\Delta = \Big\{ \xi \in \mathbb{R}^{\mathcal{K}}_{\geq 0} : \textstyle\sum_{k \in \mathcal{K}} \xi_k = 1 \Big\}, \qquad \mathcal{A}(\xi) = N \textstyle\sum_{k \in \mathcal{K}} \xi_k \, \mathbf{x}_k \mathbf{x}_k^T + \mathbf{\Lambda}, \tag{24}$$

making $\mathcal{A}$ affine in ξ, and the exact designs of size N are the lattice points $\Delta \cap N^{-1}\mathbb{Z}^{\mathcal{K}}$. The group acts on Δ by permuting coordinates, $(g \cdot \xi)_k = \xi_{g^{-1}(k)}$. Equivariance and prior invariance together imply $\mathcal{A}(g \cdot \xi) = \mathbf{P}_g \, \mathcal{A}(\xi) \, \mathbf{P}_g^T$. These are the ingredients of the group-averaging (centering) argument for invariant design problems (Pukelsheim, 1987; Kiefer, 1975), which Theorem 4 below applies.

Theorem 4 (Optimality of the balanced design under lattice symmetry)**.** (i) *Let $\Phi(\mathbf{A})$ be a $\mathcal{G}$-invariant design criterion, meaning $\Phi(\mathbf{P}_g \mathbf{A} \mathbf{P}_g^T) = \Phi(\mathbf{A})$ for all $g \in \mathcal{G}$, which is concave (to be maximized) or convex (to be minimized) on the positive-definite cone. The perfectly balanced design $\bar{\xi} \equiv 1/|\mathcal{K}|$ maximizes (minimizes) the concave (convex) criterion $\Phi(\mathcal{A}(\xi))$ over all approximate designs $\xi \in \Delta$.* (ii) *In particular, $\bar{\xi}$ is simultaneously D-, A-, E-, G-, and V-optimal.*

Proof. *(i)* Let $\psi = \Phi \circ \mathcal{A}$ on Δ. Assuming Φ is concave (the convex case simply reverses the inequalities), ψ is concave because $\mathcal{A}$ is affine. It is also $\mathcal{G}$-invariant because $\psi(g \cdot \xi) = \Phi(\mathbf{P}_g \mathcal{A}(\xi) \mathbf{P}_g^T) = \psi(\xi)$. For any $\xi \in \Delta$, construct the group average $\tilde{\xi} := |\mathcal{G}|^{-1} \sum_{g \in \mathcal{G}} g \cdot \xi$, which again lies in Δ because Δ is convex and $\mathcal{G}$-stable. Jensen's inequality and $\mathcal{G}$-invariance give:

$$\psi(\tilde{\xi}) \;\geq\; \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \psi(g \cdot \xi) \;=\; \psi(\xi). \tag{25}$$

Because $\tilde{\xi}$ is $\mathcal{G}$-invariant by construction, it is constant on the orbits of $\mathcal{G}$ in $\mathcal{K}$. Since the group action is transitive, there is only a single orbit, forcing $\tilde{\xi}$ to be the uniform weight $1/|\mathcal{K}|$, that is $\tilde{\xi} = \bar{\xi}$. As ξ was arbitrary, this proves (i).

(ii) It remains to check that each named criterion is $\mathcal{G}$-invariant and concave or convex as appropriate. The Kiefer matrix means

$$\phi_q(\mathbf{A}) = \left(\tfrac{1}{M} \operatorname{tr} \mathbf{A}^q \right)^{1/q}, \qquad -\infty \leq q \leq 1, \tag{26}$$

are concave on the positive-definite cone (Pukelsheim, 1987), the family being cut at $q = 1$ because concavity fails beyond it. They depend on $\mathbf{A}$ only through its eigenvalues, so they are invariant under conjugation by any orthogonal matrix, and in particular by $\mathbf{P}_g$. D-, A- and E-optimality are the cases $q = 0$, $q = -1$ and $q = -\infty$, the first and last by continuous limit. Moving $\phi_0 = (\det \mathbf{A})^{1/M}$ and ϕ_{-1} to the familiar $\log\det \mathbf{A}$ and $\mathrm{tr}(\mathbf{A}^{-1})$ is a strictly monotone reparametrization that does not move the optimizer. V- and G-optimality are not spectral, and are handled separately. Both are convex: $\mathrm{tr}(\mathbf{A}^{-1}\mathbf{M})$ is convex for every $\mathbf{M} \succeq \mathbf{0}$ by matrix convexity of the inverse (Bhatia, 1997), and $\max_k \mathbf{x}_k^T \mathbf{A}^{-1} \mathbf{x}_k$ is a point-wise maximum of convex functions. For V-optimality the relevant matrix is $\mathbf{M} = \sum_k \mathbf{x}_k \mathbf{x}_k^T$, for which cyclicity of the trace gives $\mathrm{tr}(\mathbf{A}^{-1}\mathbf{M}) = \sum_k \mathbf{x}_k^T \mathbf{A}^{-1} \mathbf{x}_k$. Both criteria are $\mathcal{G}$-invariant because $\mathbf{P}_g$ only permutes the vectors $\{\mathbf{x}_k\}$ among themselves: this gives $\mathbf{P}_g^T \mathbf{M} \mathbf{P}_g = \mathbf{M}$ for V-optimality, and for G-optimality the maximum is taken over a $\mathcal{G}$-stable index set. Each of the five criteria therefore satisfies the hypotheses of (i), which gives (ii). □

Jensen's inequality requires a convex feasible set, and Δ is the set of *approximate* designs. The exact designs form a non-convex subset of it. The conclusion is therefore a statement about approximate designs, attained exactly by an exact design only when $|\mathcal{K}|$ divides N, and up to integer rounding otherwise. Theorem 4 also treats the design de novo. With an already-acquired history, $\mathbf{X}^T\mathbf{X}$ need not be $\mathcal{G}$-invariant. The centering argument then applies with $\mathcal{G}$ replaced by the subgroup stabilizing it, and the conclusion weakens from uniformity to constancy on the orbits of that subgroup. The balanced design is in this sense an optimum of the criterion rather than an attractor of the rule. There are several consequences for the cold-start phase of active learning:

- **The greedy bound holds at every budget.** Remark 2 bounds the accumulated design against the best exact design of the same size at *every* budget, not only at the sizes where the balanced design of Theorem 4 is attainable. At those sizes the design the greedy is measured against is itself an exact design, optimal under every criterion of Theorem 4(ii).
- **No initialization heuristic is needed.** The selection of Theorem 3 is the classical sequential design algorithm of Wynn and Fedorov (Wynn, 1970; Fedorov, 1972). The prior $\mathbf{\Lambda}$ contributes at most $b_{\max}$ effective samples to any bin (Proposition 2) while n_k grows without bound, so its relative contribution to α_k vanishes as the budget increases. ZALMBIES therefore needs no separate initialization stage, and warms smoothly into value-weighted acquisition only as the surrogate becomes informative.
- **The role of β.** It is tempting to view the exponent β as a knob that tunes between different optimality criteria, but in the unweighted case it has no effect on the selection at all. By Theorem 3 every $\beta > 0$ produces the same trajectory, and by Theorem 4(ii) the named criteria are optimized by the same design, so the distinctions between them are invisible here. Whatever β controls, it controls only through its interaction with $\hat{U}$ (App. F.4).
- **When the guarantee breaks.** The proof of Theorem 4 requires the full *breadth* of the acquisition lattice, i.e., every category of every retained dimension must be attainable and equally costly. It does not require full *depth*, since a marginal sub-lattice of bounded interaction order is itself $\mathcal{G}$-symmetric with a $\mathcal{G}$-invariant prior, and Theorem 4 applies to it verbatim (cf. Remark 1). Missing combinations, safety-restricted cells, ragged designs, and rare categories break the symmetry the centering argument uses, and with it the simultaneous optimality. The centering argument applies only on the full support, so the balanced design then optimizes the symmetrized problem rather than the one actually faced, and our analysis does not bound the difference. Theorem 3 is unaffected. The D-optimality of each individual acquisition never invoked that symmetry, and survives all of these.

F.3 INFORMATION GAIN AND BATCH CONSTRUCTION

We now identify what the exploration term measures, and what the sequential rank-one update provides over scoring bins independently. Everything to this point has been a statement about the matrix $\mathbf{A}$, established without reference to how observations arise. An observation model buys two things the determinant alone cannot supply. It identifies $\log\det \mathbf{A}(\mathbf{F}) - \log\det \mathbf{A}$ with an information gain rather than a change of volume in parameter space, and shows that the exploration term $\alpha_k^{-\beta}$ tracks only reducible uncertainty rather than the noise level. Nothing so far has needed

such a model, so we now adopt the conjugate model,

$$y = \mathbf{x}^T\mathbf{w} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2), \quad \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{\Lambda}^{-1}). \tag{27}$$

Under this model, the posterior after acquiring data $\mathcal{D}$ with design $\mathbf{X}$ is $\mathbf{w} \mid \mathcal{D} \sim \mathcal{N}(\hat{\mathbf{w}}, \sigma^2\mathbf{A}^{-1})$ (Bishop, 2006), meaning the matrix $\mathbf{A}$ directly provides the posterior precision. The prior here is not a free choice. Under a zero-mean Gaussian prior of precision $\mathbf{\Pi}$, the posterior mean is $(\mathbf{X}^T\mathbf{X} + \sigma^2\mathbf{\Pi})^{-1}\mathbf{X}^T\mathbf{U}^{(p)}$. This exactly matches the estimator of Section 3.2 when $\sigma^2\mathbf{\Pi} = \mathbf{\Lambda}$, making the prior above the only choice consistent with it. This conjugate model is therefore fixed by the regularizer already in use rather than adopted for convenience, but making it explicit provides the posterior covariance $\sigma^2\mathbf{A}^{-1}$. We condition on a fixed noise scale σ^2 rather than marginalizing it, ensuring the predictive distributions remain exactly Gaussian. (A fully Bayesian treatment integrating out σ^2 yields Student-t predictives and lies outside our scope.) The scalar case is a special case of the batch one, so we state the batch result first and specialize.

Theorem 5 (Information-gain identification)**.** *For a batch $\mathcal{S}$ with stacked design rows $\mathbf{F}_\mathcal{S}$, the mutual information between the batch outcomes and the parameters is*

$$I(y_\mathcal{S}; \mathbf{w} \mid \mathcal{D}) = \tfrac{1}{2}\log\det\big(\mathbf{I} + \mathbf{F}_\mathcal{S}\mathbf{A}^{-1}\mathbf{F}_\mathcal{S}^T\big) = \tfrac{1}{2}\big[\log\det\mathbf{A}(\mathbf{F}_\mathcal{S}) - \log\det\mathbf{A}\big], \tag{28}$$

and the aleatoric scale σ^2 cancels identically. For a single observation in bin k this reads

$$I(y_k; \mathbf{w} \mid \mathcal{D}) = \tfrac{1}{2}\log\left(1 + \mathbf{x}_k^T\mathbf{A}^{-1}\mathbf{x}_k\right) = \tfrac{1}{2}\log\left(1 + \alpha_k^{-1}\right). \tag{29}$$

Proof. Write $y_\mathcal{S} \in \mathbb{R}^{|\mathcal{S}|}$ for the not-yet-observed outcomes at the batch rows $\mathbf{F}_\mathcal{S} \in \mathbb{R}^{|\mathcal{S}|\times M}$. Stacking the conjugate model over the batch gives $y_\mathcal{S} = \mathbf{F}_\mathcal{S}\mathbf{w} + \boldsymbol{\varepsilon}$ with independent noise $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. Conditioning on $\mathbf{w}$ isolates the observation noise, $y_\mathcal{S} \mid \mathbf{w} \sim \mathcal{N}(\mathbf{F}_\mathcal{S}\mathbf{w}, \sigma^2\mathbf{I})$. Conditioning instead on $\mathcal{D}$ leaves $\mathbf{w}$ as a random variable with law $\mathcal{N}(\hat{\mathbf{w}}, \sigma^2\mathbf{A}^{-1})$. Because $y_\mathcal{S}$ is a sum of two independent Gaussians, their covariances add directly:

$$y_\mathcal{S} \mid \mathcal{D} \;\sim\; \mathcal{N}\big(\mathbf{F}_\mathcal{S}\hat{\mathbf{w}},\; \mathbf{F}_\mathcal{S}(\sigma^2\mathbf{A}^{-1})\mathbf{F}_\mathcal{S}^T + \sigma^2\mathbf{I}\big) = \mathcal{N}\big(\mathbf{F}_\mathcal{S}\hat{\mathbf{w}},\; \sigma^2(\mathbf{I} + \mathbf{F}_\mathcal{S}\mathbf{A}^{-1}\mathbf{F}_\mathcal{S}^T)\big). \tag{30}$$

Note the first term (posterior covariance) also carries σ^2, as the chosen prior ties the parameter variance to the noise scale.

A multivariate Gaussian on $\mathbb{R}^{|\mathcal{S}|}$ with covariance $\mathbf{\Sigma}$ has differential entropy $H = \frac{1}{2}\log\det(2\pi e\mathbf{\Sigma})$ (Cover & Thomas, 2006). Because the covariance of $y_\mathcal{S} \mid \mathbf{w}$ is simply $\sigma^2\mathbf{I}$ (which is free of $\mathbf{w}$), the expectation $\mathbb{E}_{\mathbf{w}|\mathcal{D}}[H(y_\mathcal{S} \mid \mathbf{w})]$ evaluates to a constant $\frac{1}{2}\log\det(2\pi e\sigma^2\mathbf{I})$. The mutual information is the difference in entropies, $I = H(y_\mathcal{S} \mid \mathcal{D}) - \mathbb{E}_{\mathbf{w}|\mathcal{D}}[H(y_\mathcal{S} \mid \mathbf{w})]$. Subtracting these yields

$$\begin{aligned} I &= \tfrac{1}{2}\log\det\left(2\pi e\sigma^2(\mathbf{I} + \mathbf{F}_\mathcal{S}\mathbf{A}^{-1}\mathbf{F}_\mathcal{S}^T)\right) - \tfrac{1}{2}\log\det\left(2\pi e\sigma^2\mathbf{I}\right) \\ &= \tfrac{1}{2}\log\det(\mathbf{I} + \mathbf{F}_\mathcal{S}\mathbf{A}^{-1}\mathbf{F}_\mathcal{S}^T), \end{aligned}$$

showing that the common factor $(2\pi e\sigma^2)^{|\mathcal{S}|}$ cancels exactly. Applying (21) with $\mathbf{F} = \mathbf{F}_\mathcal{S}$ gives the second equality of (28). Taking $|\mathcal{S}| = 1$ and $\mathbf{F}_\mathcal{S} = \mathbf{x}_k^T$ reduces the determinant to $1 + \mathbf{x}_k^T\mathbf{A}^{-1}\mathbf{x}_k = 1 + \alpha_k^{-1}$, yielding (29). □

The determinant difference in (28) is the quantity Theorem 3 already maximizes one step at a time. What is added here is its identification with a mutual information, and the cancellation of σ^2 that comes with it. Both (28) and (29) are identities of the conjugate model, holding for every value weight $\hat{U}$ and exponent β in (10).

The cancellation of σ^2 in (29) is what makes the exploration term epistemic. Maximum-entropy (predictive-entropy) sampling maximizes $H(y_k \mid \mathcal{D})$ and therefore rewards bins with large *irreducible* noise, whereas the mutual information rewards only the reducible, parameter-borne (epistemic) uncertainty (Kendall & Gal, 2017). The exactness of this cancellation is a property of the conjugate coupling $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{\Lambda}^{-1})$, in which the prior scales with the noise. For an uncoupled prior the noise would instead damp the gain rather than vanish. Exactness therefore requires the prior scale to match the noise scale of the bin being scored. A single prior cannot match every bin's noise scale at once, so comparing *across* bins additionally uses the approximately homoscedastic scale

secured by the transform of App. A, without which each $I(y_k; \mathbf{w} \mid \mathcal{D})$ would carry a factor σ^2/σ_k^2 on α_k^{-1} inside the logarithm. The exact-gain form of the score, written $\tilde{S}_k$, factorizes accordingly,

$$\tilde{S}_k = \underbrace{\hat{U}_k}_{\text{value weight}} \quad \underbrace{\log\bigl(1 + \alpha_k(\mathbf{F})^{-1}\bigr)}_{\text{epistemic gain}}, \tag{31}$$

which Remark 3 relates at $\beta = 1$ to the power score (10) that ZALMBIES implements. In either form, any dependence on aleatoric noise enters solely through the deliberate value weight $\hat{U}_k$, never through the exploration term. Because (28) is exact and its submodularity follows algebraically rather than from an estimate, the guarantee of Remark 2 needs no sampling approximation, only the restriction on $\hat{U}$ recorded there.

At $\hat{U} \equiv 1$, scoring bins *independently* against a posterior held fixed across the batch optimizes an upper bound on the joint information (28) rather than the information itself. Let $\mathbf{H} = \mathbf{F}_{\mathcal{S}}\mathbf{A}^{-1}\mathbf{F}_{\mathcal{S}}^T \succeq \mathbf{0}$, whose diagonal entries are $h_{ii} = \mathbf{x}_i^T\mathbf{A}^{-1}\mathbf{x}_i = \alpha_i^{-1}$. Hadamard's inequality (Cover & Thomas, 2006; Horn & Johnson, 2013) applied to $\mathbf{I} + \mathbf{H}$, with (28), gives

$$\sum_{i \in \mathcal{S}} \tfrac{1}{2}\log\bigl(1 + \alpha_i^{-1}\bigr) \;\geq\; I(y_{\mathcal{S}}; \mathbf{w} \mid \mathcal{D}) = \tfrac{1}{2}\log\det(\mathbf{I} + \mathbf{H}), \tag{32}$$

with equality if and only if $\mathbf{I} + \mathbf{H}$ is diagonal, that is, if the selected rows are mutually $\mathbf{A}^{-1}$-orthogonal. The left-hand side is additive over the batch, so the B highest-scoring selections are its exact maximizer. The naive top-B strategy optimizes precisely this upper bound, overshooting the true joint information by the redundancy $R(\mathcal{S}) = \sum_{i \in \mathcal{S}} \frac{1}{2}\log(1 + h_{ii}) - \frac{1}{2}\log\det(\mathbf{I} + \mathbf{H}) \geq 0$. The rank-one $\alpha_k(\mathbf{F})$ update of Section 3.3 instead evaluates each marginal gain against the matrix already *updated* with the rows selected so far, so the accumulated gains telescope to $\log\det\mathbf{A}(\mathbf{F}_{\mathcal{S}}) - \log\det\mathbf{A} = 2\,I(y_{\mathcal{S}}; \mathbf{w} \mid \mathcal{D})$ and the sequential update never incurs $R(\mathcal{S})$.

Remark 3 ($\beta = 1$ as an approximate information gain). Since $\log(1 + \alpha^{-1}) = \alpha^{-1} + O(\alpha^{-2})$, the power score (10) at $\beta = 1$ is the first-order ($\alpha_k \gg 1$) expansion of (31). The term it expands is the exact information gain (29) of Theorem 5, which is the BALD score (Houlsby et al., 2011); ZALMBIES at $\beta = 1$ is therefore a first-order BALD with a value weight attached. The identification of $\beta = 1$ with greedy Bayesian D-optimal design is likewise asymptotic rather than exact once $\hat{U}$ varies across bins. The expansion overshoots by $\frac{1}{2}\alpha_k^{-2}$ to leading order, and overshoots relatively most where α_k is smallest, so the power score is marginally more exploratory in sparse bins. At $\hat{U} \equiv 1$ the two forms are related by a strictly increasing map and select identically (Theorem 3), so they can diverge only where the value weight varies across bins.

F.4 WHAT THE EXPONENT β CONTROLS

Whenever $\hat{U}$ is constant and $\beta > 0$, by Theorem 3, pure exploration makes an exactly D-optimal selection at every step. The role of β is therefore entirely to mediate between two well-defined endpoints, value exploitation and pure coverage. At $\beta = 0$, the score is $S_k = \hat{U}_k$, and acquisition is pure exploitation, with no reference to design coverage. Conversely, as $\beta \to \infty$, the log-score $\log S_k = \log\hat{U}_k - \beta\log\alpha_k(\mathbf{F})$ becomes entirely dominated by the second term. Provided $\hat{U}$ is bounded and strictly positive, and α_k has a unique minimizer, the selection converges exactly to $\arg\min_k \alpha_k(\mathbf{F})$, the pure-coverage step of Theorem 3, whose optimum is the balanced design of Theorem 4. Thus, the exponent provides a dial controlling how far the design is permitted to depart from balanced allocation in pursuit of predicted failure. The metadata equal baseline of Section 4 is accordingly not an arbitrary comparator, but a direct empirical reading of this $\beta \to \infty$ limit.

Interior values of β retain their classical optimal-design interpretations, subject to the idealizations discussed in Remarks 3 and 4, for $\beta = 1$ and $\beta = 2$ respectively. Specifically, $\beta = \frac{1}{2}$ is a predictive-standard-deviation (confidence-width) rule, $\beta = 1$ is the first-order form of exact information gain (greedy D-optimality), and $\beta = 2$ is the well-sampled limit of a $\hat{U}$-weighted V-optimal step. Only $\beta = \frac{1}{2}$ requires no idealization. Under the conjugate model of App. F.3 the bin mean has posterior variance $\sigma^2\mathbf{x}_k^T\mathbf{A}^{-1}\mathbf{x}_k = \sigma^2\alpha_k^{-1}$, so $\alpha_k^{-1/2}$ is exactly the epistemic predictive standard deviation up to the constant σ, and $\beta = \frac{1}{2}$ therefore scales the value weight by a confidence width rather than by a variance. The score applies this width multiplicatively rather than as an additive bonus, for the

reason given in Section 3.3. Crucially, these are interpretations of points along a single continuous family rather than entirely separate algorithms. They are distinguishable from one another only because an informative $\hat{U}$ breaks the β-invariance that Theorem 3 establishes. Our ablations over $\beta \in \{0, \frac{1}{2}, 1, 2\}$ (App. D) therefore systematically measure how much value weighting should be allowed to override balance.

Remark 4 ($\beta = 2$ and weighted V-optimal design). Let $\mathbf{W} = \sum_k \hat{U}_k \, \mathbf{x}_k \mathbf{x}_k^T$. Define the uncertainty-weighted total predictive variance as $\Phi_{\hat{U}}(\mathbf{F}) = \mathrm{tr}(\mathbf{A}(\mathbf{F})^{-1}\mathbf{W}) = \sum_k \hat{U}_k \, \alpha_k(\mathbf{F})^{-1}$. By the Sherman–Morrison formula (Sherman & Morrison, 1949), the exact variance reduction from acquiring in bin j is:

$$\Delta_j = \frac{\mathbf{x}_j^T \mathbf{A}^{-1} \mathbf{W} \mathbf{A}^{-1} \mathbf{x}_j}{1 + \alpha_j^{-1}}. \tag{33}$$

If the acquisition directions are mutually $\mathbf{A}^{-1}$-orthogonal (the disjoint-support limit), this reduces to $\hat{U}_j/(\alpha_j^2 + \alpha_j)$. Sibling bins share columns, so the cross terms this discards are non-negative and the exact decrement (33) is larger. Once the bin is well-sampled ($\alpha_j \gg 1$), $\hat{U}_j/(\alpha_j^2 + \alpha_j)$ approaches the $\beta = 2$ score $\hat{U}_j \, \alpha_j^{-2}$, which exceeds it by a factor of $1+\alpha_j^{-1}$. The two deviations have opposite signs, so the identification with V-optimality should be viewed as suggestive rather than exact. Furthermore, because $\mathbf{W}$ is weighted, $\Phi_{\hat{U}}$ is only *weakly* submodular. Its greedy guarantee degrades from $(1 - 1/e)$ to $(1 - e^{-\nu})$, where ν is the submodularity ratio (Das & Kempe, 2011; Bian et al., 2017). At $\hat{U} \equiv 1$ the criterion becomes unweighted V-optimality, a $\mathcal{G}$-invariant Φ to which Theorem 4 applies. The degradation is a cost of the value weighting alone.

The preceding identifications are local, interpreting individual greedy steps through classical criteria that only hold in unattained limits. Beyond these local steps, the exponent β controls the design's global behavior, where an exact statement is possible. As α_k never shrinks and the borrowing term b_k is bounded (Propositions 1 and 2), the system cannot indefinitely starve any single bin, and the scores are driven towards equality. The allocation this leaves behind depends on nothing but the value weights and β; the prior, the lattice geometry, and the entire acquisition history wash out.

Proposition 6 (Limiting allocation of the acquisition rule)**.** *Fix value weights $\hat{U}_k > 0$ and an exponent $\beta > 0$, and let sequential acquisition under (10) proceed indefinitely. Then every bin is selected infinitely often, and the realized allocation converges to*

$$\frac{n_k}{n} \longrightarrow \frac{\hat{U}_k^{1/\beta}}{\sum_j \hat{U}_j^{1/\beta}}. \tag{34}$$

In particular, if the weights are held fixed across a batch of B acquisitions, the batch's own composition Δn_k satisfies $\Delta n_k / B \to \hat{U}_k^{1/\beta} / \sum_j \hat{U}_j^{1/\beta}$ as $B \to \infty$, whatever the counts at the start of the batch.

Proof. Write $c = 1 + b_{\max}$ and $S_k(T) = \hat{U}_k \, \alpha_k(T)^{-\beta}$ for the score of bin k after T acquisitions. By Proposition 1 no acquisition decreases any α_k, so every S_k is non-increasing in T.

Every bin is selected infinitely often. Suppose bin k is selected only finitely often. Then n_k is eventually constant and, by (17), $\alpha_k \le n_k + b_{\max}$ is bounded, so S_k is bounded below by a positive constant. Any bin j selected infinitely often has $n_j \to \infty$, hence $\alpha_j \to \infty$ and $S_j \to 0$, so eventually $S_k > S_j$, contradicting the greedy selection of j. Therefore $n_k \to \infty$, and hence $\alpha_k \to \infty$, for every bin.

The scores equalize in ratio. Fix T and a bin k, and let $t_k \le T$ be the last time k was selected. Since n_k has increased by exactly one since t_k, and b_k is non-decreasing and bounded by $b_{\max}$ (Propositions 1 and 2), we have $\alpha_k(T) \le \alpha_k(t_k) + c$. Greedy selection at t_k together with the monotonicity of the scores gives $\max_j S_j(T) \le \max_j S_j(t_k) = S_k(t_k)$, so

$$1 \le \frac{\max_j S_j(T)}{S_k(T)} \le \left(\frac{\alpha_k(T)}{\alpha_k(t_k)}\right)^{\beta} \le \left(1 + \frac{c}{\alpha_k(t_k)}\right)^{\beta}. \tag{35}$$

As $T \to \infty$ the last selection time t_k tends to infinity as well, since k is selected infinitely often, so $\alpha_k(t_k) \to \infty$ and the right side tends to one.

Conclusion. The bound holds for every k, so $\hat{U}_k \alpha_k^{-\beta}$ and $\hat{U}_j \alpha_j^{-\beta}$ agree asymptotically and $\alpha_k/\alpha_j \to (\hat{U}_k/\hat{U}_j)^{1/\beta}$. Since $\alpha_k = n_k + b_k$ with b_k bounded and $n_k \to \infty$, the same ratio holds for the counts, and normalizing over $\mathcal{K}$ gives (34). Acquisition in batches of fixed size changes only the constant, c being enlarged by the batch size. The batch composition claim is the same conclusion read with the pre-batch counts as the initial design; being finite, they vanish under normalization by B. □

Remark 5 ($\beta = 1$ and Neyman allocation). When $\beta = 1$, the limit (34) matches Neyman allocation, the classical rule minimizing the variance of a stratified estimate by setting $n_k \propto \sigma_k$ (Neyman, 1934). The identification requires $\hat{U}$ on the scale of a standard deviation rather than a variance, matching the square root taken in Section 3.3. The rule is thus pinned at both ends: Theorem 3 supplies exact D-optimality at zero data, and Proposition 6 a classical stratified allocation once acquisitions are plentiful.

The comparison is not exact because our bins are not independent strata. Sharing parameters across the lattice means a bin mean carries a variance of σ^2/α_k rather than σ^2/n_k, so the weighted criterion of Remark 4 is exactly $\sum_k \hat{U}_k/\alpha_k$. Minimizing it under a budget constraint yields $\alpha_k \propto \hat{U}_k^{1/2}$, the exponent of (34) at $\beta = 2$ on the variance scale and at $\beta = 1$ on the standard-deviation scale, which is Neyman allocation targeting the *effective* sample size. The algorithm pursues this distribution from the very first acquisition. The proof of Proposition 6 directly establishes $\alpha_k/\alpha_j \to (\hat{U}_k/\hat{U}_j)^{1/\beta}$. The classical count statement emerges only as $b_k/\alpha_k \to 0$. Treating b_k as fixed while n_k varies makes the Lagrange step a first-order heuristic rather than a theorem.

Within a single batch the frozen weights assumed in Proposition 6 are not an idealization but the implementation. No labels arrive during a batch, so $\hat{U}$ cannot move while the rank-one updates advance α. Unlike Remarks 3 and 4, whose hypotheses ZALMBIES violates, this one ZALMBIES enforces. Extending the statement across an entire run is another matter, requiring the value estimates to settle faster than the sampled pool grows. Because acquisition only adds points, any initial surplus can only be diluted. We make no claim either way.

We stress that none of this is the source of ZALMBIES's advantage. Classical Neyman allocation requires an estimate for every stratum and breaks down if any bin is empty, as must happen whenever the budget is smaller than $|\mathcal{K}|$. The asymptotic limit merely indicates that ZALMBIES suffers no long-term penalty for its sparse-data efficiency. The parameter borrowing that makes the small-sample regime tractable seamlessly fades away exactly as the classical estimates gain enough data to become reliable.

Relation to existing acquisition functions. The closed forms derived in the preceding subsections are the conjugate linear–Gaussian specializations of acquisition functions that are usually approximated in deep active learning. The single-point information gain (29) is exactly the BALD score (Houlsby et al., 2011; Gal et al., 2017). The joint objective (28) together with the redundancy gap (32) is the counterpart of BatchBALD (Kirsch et al., 2019), which must estimate the joint mutual information by Monte-Carlo over posterior samples and rely on approximate submodularity. The V-optimal setup (33) is a BAIT-style Fisher design (Ash et al., 2021), obtained here in closed form rather than through last-layer embedding approximations.

These connections are separated from our practical setting by two layers of idealization. First, the model gap: we derive them exactly on the linear–Gaussian surrogate, not on the network whose failures it models. Second, the structural gap: recovering these classical criteria requires limits that do not hold during actual execution. Specifically, BALD/BatchBALD requires flat value weights ($\hat{U} \equiv 1$), and the BAIT-style variance reduction requires mutually orthogonal design rows, a condition explicitly violated by the feature sharing of our categorical lattice and the hierarchical prior. They are conceptual touchstones rather than guarantees, and the utility of the β parameter, together with the ideal trade-off between design balance and value-weighting, remains an empirical question that our ablations are designed to answer.