Multi-fidelity Bayesian experimental design using power posteriors

Andrew Jones* **Diana Cai***
Princeton University

Barbara E. Engelhardt
Gladstone Institutes & Stanford University

1 Introduction

As experimental tools in the physical and life sciences become increasingly sophisticated and costly, there is a need to optimize the choice of experimental parameters to maximize the informativeness of the data and minimize cost. When designing an experiment, there is often a choice among a suite of data collection modalities or instruments with varying *fidelities* that trade off between accuracy and cost. For example, biologists have access to a range of tools to study living tissues, from state-of-the-art yet expensive technologies for mapping the molecular makeup of individual cells to more dated yet cheap tools to measure the aggregate molecular composition of the tissue's cells. Analyzing the tradeoff between high-fidelity, high-cost measurements and low-fidelity, low-cost measurements is often difficult due to complex data collection procedures and budget constraints.

Multi-fidelity methods combine multiple models of varying fidelity to accelerate computational algorithms. Multi-fidelity methods have been proposed for applications such as Bayesian optimization [3, 13, 17, 20, 24, 28], Bayesian quadrature [10, 29], Bayesian inference algorithms [4–6, 8]. Notably, several multi-fidelity extensions of Gaussian process (GP) models have been proposed. Most of these approaches take the form of a multi-output GP, where each output corresponds to a fidelity, and the relationship between fidelities is captured in the kernel function. In particular, common multi-fidelity GP approaches include using the linear model of regionalization (LMC, [19]), autoregressive models [18], and deep GPs [7] (see Brevault et al. [3] for an overview). In the field of experimental design, a handful of multi-fidelity models have been proposed to select among a panel of data sources with varying fidelities [11, 12].

Despite these developments, most existing approaches for multi-fidelity modeling and experimental design are tied to specific data modeling choices and therefore may constrain the analyst to a specific choice of model. Moreover, for GP models, choosing among existing multi-fidelity approaches induces a tradeoff between computational tractability and flexibility in modeling the fidelities' relationships. In particular, while the LMC-based approach to multi-fidelity GPs is computationally simple, it assumes a symmetric relationship between fidelities, which is unrealistic in most settings where there is a natural preference for high-fidelity measurements. However, using multi-fidelity methods that allow for more complex fidelity relationships require more demanding computation.

We propose an approach for designing experiments using a Bayesian power posterior [14, 16, 22, 27], which naturally accounts for varying fidelities by using a fractional power to discount the likelihood of low-fidelity observations. This approach—termed a *multi-fidelity posterior*—was recently proposed in the context of Bayesian online change point detection [15]. We apply this in the context of information-based objectives for design problems using *expected information gain* (EIG) (a.k.a. mutual information). Concretely, we propose a multi-fidelity expected information rate (MF-EIR) objective, which measures the expected information gain with respect to the multi-fidelity posterior, while accounting for the cost of information. Our approach using power posteriors is applicable to a large class of probabilistic models, allowing for multi-fidelity experimental design beyond GP-based modeling approaches. The implementation of the power posterior is conceptually straightforward, and algorithms for approximate Bayesian inference can be leveraged. We demonstrate our approach through a simulation study with a GP regression model and an application to a genomics experiment.

*First authorship is shared jointly by A. Jones and D. Cai.

2022 NeurIPS Workshop on Gaussian Processes, Spatiotemporal Modeling, and Decision-making Systems.

2 Multi-fidelity Bayesian experimental design

We consider a sequential experimental design setting with T iterations. Let $\mathcal{X}$ denote the design space. The task on iteration t is to choose a design $x_t \in \mathcal{X}$ given data from the first $t-1$ iterations, which will yield an observation y_t. Intuitively, the goal in choosing the design is to maximize the chances of receiving an informative experimental result. To do this, a utility function must be chosen that will be optimized; in this work, we will explicitly consider the expected information gain (EIG) objective, but our approach can be applied more generally. Finally, the cost of gathering an observation is often expensive. Thus, the goal is to use observations of varying fidelities to reduce the computational cost of experimental design while maintaining informativeness.

In the multi-fidelity setting, each design x_t has an associated fidelity ζ_t, which is a known measure of the expected quality or precision of the observation to be yielded from the design. For instance, the fidelities could correspond to different collection instruments with varying noise levels and costs. Each fidelity ζ_t takes values in the set $\mathcal{Z} := \{\zeta^{(1)}, \dots, \zeta^{(K)}\}$, where for all k, $\zeta^{(k)} \in (0, 1]$ and K is the number of fidelities. Larger $\zeta_t \in \mathcal{Z}$ correspond to higher fidelity observations. Let $c : \mathcal{Z} \times \mathcal{X} \to \mathbb{R}$ denote a cost function, where observing a design x with fidelity ζ incurs cost $c(\zeta, x)$; we assume the cost for a fixed design increases with fidelity. In most applications, a higher fidelity design induces a larger cost.

2.1 Multi-fidelity expected information rate

Before introducing the proposed multi-fidelity utility function, we first review the idea of a multi-fidelity posterior [15], which uses a tempered likelihood to represent how noisy a particular data source is. Let $\mathcal{D}_t := \{(x_n, y_n)\}_{n=1}^t$ be the observations with fidelities $\{\zeta_n\}_{n=1}^t$. Consider a Bayesian model with parameter of interest θ and prior $\pi(\theta)$, and let $p_\theta(y) := p(y|\theta, x)$ denote the likelihood of an observation y with design x. Given a set of observations $\mathcal{D}_t$ measured at fidelities $\zeta_{1:t}$, the *multi-fidelity posterior* is

$$\pi(\theta|\mathcal{D}_t, \zeta_{1:t}) \propto \pi(\theta) \prod_{n=1}^{t} p_\theta(y_n)^{\zeta_n}. \tag{1}$$

When $\zeta_n < 1$, the nth likelihood term is downweighted, and the posterior becomes less concentrated. Thus, a smaller ζ_n has the effect of a higher noise level for that sample. Re-weighting terms in the likelihood has been considered in many other contexts in the Bayesian literature [2, 14, 16, 22, 26, 27].

Incorporating the fidelity generalizes the design task described above: a fidelity must also be chosen on each experimental iteration in addition to the design. This naturally yields the general multi-fidelity form of the EIG for a design (x_t, ζ_t) using the multi-fidelity posterior defined in Equation 1:

$$\text{MF-EIG}(x_t, \zeta_t) = \mathbb{H}[\pi(\theta|\mathcal{D}_{t-1}, \zeta_{1:t-1})] - \mathbb{E}_{y_t}\left[\mathbb{H}[\pi(\theta|\mathcal{D}_{t-1}, x_t, y_t, \zeta_{1:t})]\right]. \tag{2}$$

Finally, we incorporate costs into the design criterion by penalizing high-cost fidelities. The final objective, which we call the *multi-fidelity expected information rate* (MF-EIR), is given by

$$(x_t^\star, \zeta_t^\star) = \text{argmax}_{x_t, \zeta_t} \text{MF-EIG}(x_t, \zeta_t)/c(\zeta_t, x_t), \tag{3}$$

where Equation 2 is divided by the cost of observing design x_t at fidelity ζ_t. Thus, the multi-fidelity criterion (Equation 3) can be interpreted as the expected information gain per unit of cost.

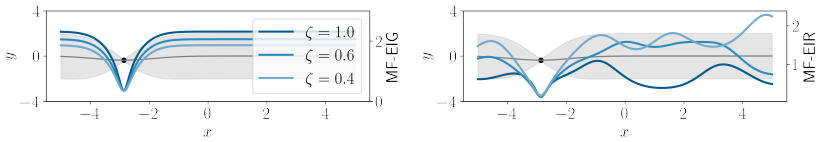
2.2 Examples

Bayesian linear regression. Consider the Bayesian linear regression model with $\mathcal{X} \subseteq \mathbb{R}^p$,

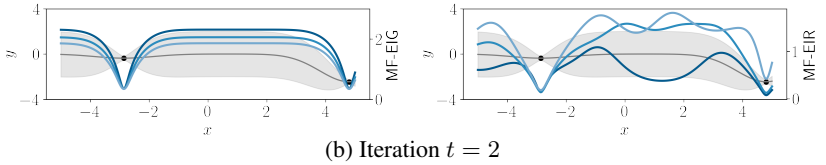
$$y = \mathbf{x}^\top \boldsymbol{\beta} + \epsilon, \;\; \boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_0), \;\; \epsilon \sim \mathcal{N}(0, \sigma^2). \tag{4}$$

When σ^2 is known, the multi-fidelity posterior for $\boldsymbol{\beta}$ takes the form $\boldsymbol{\beta}|\mathcal{D}_t \sim \mathcal{N}(\mathbf{m}_t, \mathbf{S}_t)$, where $\mathbf{m}_t = (\mathbf{X}^\top \mathbf{D}_t \mathbf{X} + \mathbf{S}_0^{-1})^{-1} \mathbf{X}^\top \mathbf{D}_t y$, $\mathbf{S}_t = \sigma^2 (\mathbf{X}^\top \mathbf{D}_t \mathbf{X} + \mathbf{S}_0^{-1})^{-1}$, $\mathbf{D}_t = \text{diag}(\zeta_1, \cdots, \zeta_t)$, and $\mathbf{X}$ is the design matrix. Suppose we have observed t samples $\{(\mathbf{x}_n, \mathbf{y}_n, \zeta_n)\}_{n=1}^{t-1}$ and are tasked with choosing the next design $(\mathbf{x}_t, \zeta_t)$. To do so, we maximize the MF-EIG, which is given by

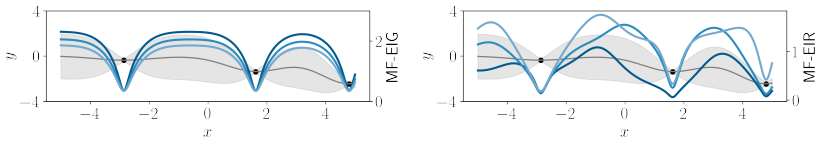
$$\text{MF-EIG}(\mathbf{x}_t, \zeta_t) = \frac{1}{2} \log(\zeta_t \sigma^{-2} \mathbf{x}_t^\top \mathbf{S}_0 \mathbf{x}_t + 1 - \zeta_t \sigma^{-2} \mathbf{x}_t \mathbf{S}_0 \mathbf{X}^\top (\mathbf{X} \mathbf{S}_0 \mathbf{X}^\top + \mathbf{D}_{t-1})^{-1} \mathbf{X} \mathbf{S}_0 \mathbf{x}_t).$$



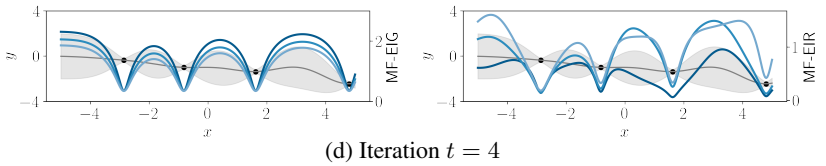
(a) Iteration $t = 1$



(b) Iteration $t = 2$



(c) Iteration $t = 3$



(d) Iteration $t = 4$

Figure 1: **Multi-fidelity EIG and EIR for GP regression.** The black dots denote observations, and the gray line and band denote the predictive mean and twice the predictive standard deviation, respectively. The left column shows the MF-EIG and the right column shows the MF-EIR.

Gaussian processes. Consider the GP regression model,

$$y = f(x) + \epsilon, \quad f \sim \mathrm{GP}(0, k(\cdot, \cdot)), \quad \epsilon \sim \mathcal{N}(0, \sigma^2). \tag{5}$$

In this case, raising the Gaussian likelihood to a power ζ simple scales the variance as σ^2/ζ. This provides a straightforward route to computing the MF-EIG for GP regression. Let $\mathbf{D} = \mathrm{diag}(\sigma^2/\zeta_1, \cdots, \sigma^2/\zeta_{t-1})$, let $\mathbf{k}_\star = [k(x_1, x_t), \cdots, k(x_{t-1}, x_t)]^\top$, and let $\mathbf{K}$ be the matrix whose ij'th element is $[\mathbf{K}]_{ij} = k(x_i, x_j)$ for $i, j = 1, \ldots, t-1$. The MF-EIG is then

$$\text{MF-EIG}(x_t, \zeta_t) = \frac{1}{2} \log(\zeta_t/\sigma^2 k(x_t, x_t) + 1 - \zeta_t/\sigma^2 \mathbf{k}_\star^\top (\mathbf{K} + \mathbf{D})^{-1} \mathbf{k}_\star).$$

3 Experiments

Simulation studies. We first demonstrate our approach on a one-dimensional synthetic example, where the outcome y is a noisy draw from a GP (Equation 5). The design space is $\mathcal{X} = [-5, 5]$, and we consider fidelities $\mathcal{Z} = \{0.4, 0.6, 1.0\}$. On each step of experimental design, we choose the fidelity and design that maximize the MF-EIR (Equation 3). For this experiment, we fix the noise variance to $\sigma^2 = 0.01$ and construct cost functions as $c(\zeta, x) = \exp\{g_\zeta(x)\}$, where $g_\zeta \sim \mathrm{GP}(\zeta, k(\cdot, \cdot))$ is a fidelity-specific cost function. We ran the experiment forward for $T = 4$ iterations and examined the MF-EIG and MF-EIR on each iteration (Figure 1). In Figure A.1 we compare the MF-EIG with a single-fidelity EIG. We find that our proposed MF-EIR provides an informative view of the trade off between the fidelity-specific information gain and the cost of running each experiment. Moreover, we find that our approach to multi-fidelity experimental design yields substantially different behavior than the analogous approach to single-fidelity experimental design using EIG.

Application to spatial transcriptomics. We next apply our approach to an experimental design problem in genomics in which a scientist has extracted a biological tissue of interest (brain, heart, lung, etc.) and is now tasked with profiling the spatial organization of gene expression across this tissue. Most current technologies for spatially-resolved transcriptomic profiling [9, 21, 23, 25] require taking cross-sections of the tissue and then performing data collection on each section separately.

We consider the problem of deciding which tissue cross-sections to collect and analyze in order to maximally understand the spatial organization of its gene expression. For simplicity, we consider

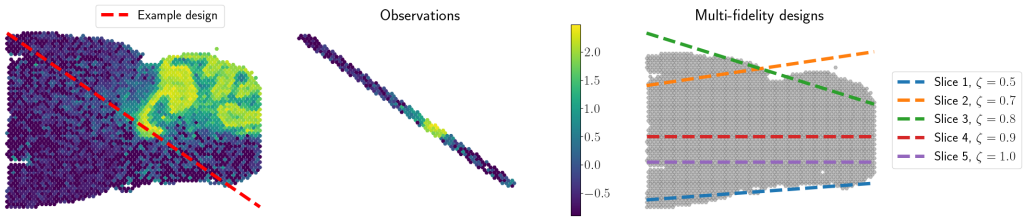


Figure 2: **Application to genomics experiment.** *Left*: Cross-section of mouse brain tissue. Each point is colored by the expression of the gene *PCP2*. The red dashed line shows one candidate design. *Middle*: The observations that would be revealed after taking the example design in the left panel. *Right*: The designs selected by our multi-fidelity experimental design approach.

the expression of a single gene. We model an outcome $y \in \mathbb{R}$ (the level of gene expression) across a spatial domain $\mathcal{X} \subset \mathbb{R}^3$ (the biological tissue) with a GP regression model. The observations consist of pairs $(\mathbf{x}, y)$, where $\mathbf{x} \in \mathbb{R}^3$ is a spatial location and y is the gene expression at that location. Several technologies exist to collect such data, each offering a different spatial resolution and precision of gene expression measurements. For each cross-section collected, we choose a technology to use to collect the data, each with its own cost and fidelity.

We leverage publicly-available data from the 10x Genomics Visium platform [1], which contains spatially-resolved gene expression data from mouse cortex. We make several simplifying assumptions. First, due to the lack of available data for all possible tissue cross-sections, we consider one-dimensional cross-section designs through a two-dimensional tissue (see Figure 2), rather than two-dimensional cross-sections through a three-dimensional tissue. Second, we assume the availability of six gene expression profiling technologies, $\mathcal{Z} = \{0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$, where the fidelity corresponds to the spatial resolution that is artificially induced by subsampling the respective fraction of points from a cross-section. Finally, we assume that chosen cross-section designs can intersect one another; in reality this would not be possible due to the fragmenting of the tissue.

We specify the cost of a slice to be equal to the number of points observed on that slice (after fidelity-specific subsampling, see Appendix A.2 for details). We run experimental design for $T = 5$ iterations, choosing the design with maximum MF-EIR on each step, and we visualize the selected designs. We find that our approach initially selects low-fidelity, low-cost cross-sections of the tissue near the boundary for the first three iterations. On the last two iterations, it selects high-fidelity, high-cost cross-sections. This behavior suggests that the MF-EIR is effectively trading off fidelity, cost, and information gain across the design process. This experiment demonstrates the flexibility of our proposed approach and serves as a proof of concept for the application of our method to real-world experimental design settings.

4 Discussion

In this work, we propose an approach to designing experiments in which multiple fidelities are available. Our approach, based on tempering the likelihood of each observation with its fidelity, is conceptually straightforward and extensible to a myriad of modeling choices. We demonstrated our approach through a simulation study of GP regression and an application to a genomics experimental design setting.

There are a number of limitations and opportunities for future work with our proposed approach. While the multi-fidelity posterior is generally applicable across probabilistic models, it only allows one way for the fidelity to interact with the model (via tempering the likelihood). In many experimental design settings, we may have more information about the data collection process that could be incorporated into the model. For example, in the genomics setting, the level of spatial resolution of a measurement could be encoded directly into the model. In future work, we will apply the approach to more complex models that require the use of approximate Bayesian inference algorithms. Finally, it would be fruitful to explore connections to other applications, such as Bayesian optimization and Bayesian quadrature. The multi-fidelity approach explored here for EIG can be likewise applied to the acquisition functions used commonly in these other applications.

Acknowledgments and Disclosure of Funding

This work was funded by Helmsley Trust grant AWD1006624, NIH NCI 5U2CCA233195, and NSF CAREER AWD1005627. D. Cai was supported in part by a Google Ph.D. Fellowship in Machine Learning.

A Additional experimental details

A.1 GP example: comparison with single-fidelity EIG

In Figure A.1, we compare the behavior of the single-fidelity EIG (top row) with the multi-fidelity EIG (bottom two rows) on $T = 3$ iterations. We considered a similar setup to the figures in the main document, but with different cost functions and fidelities $\mathcal{Z} = \{0.1, 0.5, 1.0\}$. For all experiments, a squared exponential kernel was used.

A.2 Genomics application

This section provides a more detailed formalization of the experimental design problem for the genomics application. We consider a spatial genomics dataset consisting of pairs $(\mathbf{x}, y)$, where $\mathbf{x} \in \mathbb{R}^3$ is a spatial location and y is a scalar outcome at this location (i.e., the level of expression of a gene at this location). We denote a dataset of t such pairs as $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^t$.

Suppose our goal is to profile a tissue, organ, or entire organism whose cells lie in the spatial domain $\mathcal{X} \subset \mathbb{R}^3$. A spatial genomics experiment requires collecting a slice from the tissue of interest. This slice is a two-dimensional cross-section of $\mathcal{X}$. We represent a two-dimensional plane with its normal vector $\boldsymbol{\beta} \in \mathbb{R}^4$ and denote the set of points on a cross section as

$$\mathcal{X}_\beta = \{\mathbf{x} \in \mathcal{X} : (1, \mathbf{x}^\top)^\top \boldsymbol{\beta} = 0\}.$$

The design space is thus the set of cross-sections parameterized by $\boldsymbol{\beta}$.

To incorporate varying fidelities into the experiment, we artificially create varying spatial resolutions of measurements. Specifically, when a fidelity ζ and design $\boldsymbol{\beta}$ is chosen, we subsample the set $\mathcal{X}_{\boldsymbol{\beta}}$ (without replacement) to have $\lceil \zeta |\mathcal{X}_{\boldsymbol{\beta}}| \rceil$, where $|\mathcal{X}_{\boldsymbol{\beta}}|$ is the cardinality of the set of points observed when taking the cross section defined by $\boldsymbol{\beta}$. We then define the cost function in these experiments as the number of points observed after subsampling, $c(\zeta, \boldsymbol{\beta}) = \lceil \zeta |\mathcal{X}_{\boldsymbol{\beta}}| \rceil$.

In our experiments, due to the lack of available data for three-dimensional data, we consider a two-dimensional spatial domain where the design space is the set of one-dimensional cross-sections through that domain.

A.3 Code availability

A Colab notebook for the simulation experiment is available at `https://tinyurl.com/mfed-colab`.

References

[1] x. Genomics. *Mouse Brain Serial Sections (Sagittal-Posterior)*, spatial gene expression dataset by space ranger 1.1.0, 10x genomics, (2020, june 23)., 2020.

[2] P. G. Bissiri, C. C. Holmes, and S. G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 78(5): 1103, 2016.

[3] L. Brevault, M. Balesdent, and A. Hebbal. Overview of Gaussian process based multi-fidelity techniques with variable relationship between fidelities. *arXiv preprint arXiv:2006.16728*, 2020.

[4] D. Cai and R. P. Adams. Multi-fidelity Monte Carlo: a pseudo-marginal approach. *Advances in Neural Information Processing Systems*, 35, 2022.

[5] J. A. Christen and C. Fox. Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical Statistics*, 14(4):795–810, 2005.

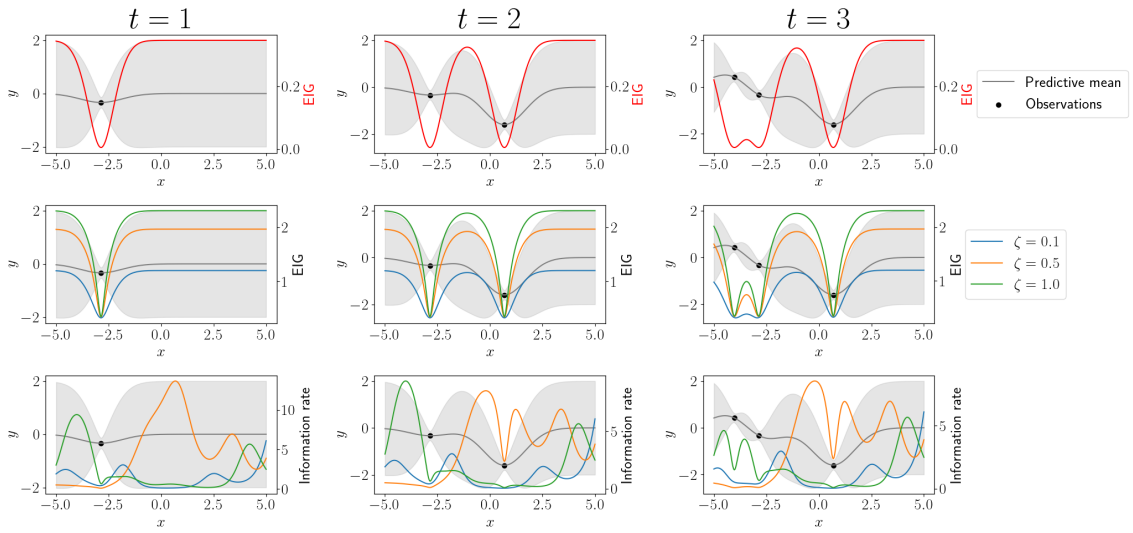


Figure A.1: **Single-fidelity EIG vs multi-fidelity EIG and EIR for GP regression with synthetic data.** Each column shows one iteration of experimental design. Each panel shows the observations (black dots), predictive mean (gray line), and twice the predictive standard deviation (gray band). The top row shows the usual (single-fidelity) EIG, the middle row shows the multi-fidelity EIG, and the bottom row shows the multi-fidelity information rate.

[6] T. Cui, Y. M. Marzouk, and K. E. Willcox. Data-driven model reduction for the Bayesian solution of inverse problems. *International Journal for Numerical Methods in Engineering*, 102 (5):966–990, 2015.

[7] K. Cutajar, M. Pullin, A. Damianou, N. Lawrence, and J. González. Deep gaussian processes for multi-fidelity modeling. *arXiv preprint arXiv:1903.07320*, 2019.

[8] Y. Efendiev, T. Hou, and W. Luo. Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. *SIAM Journal on Scientific Computing*, 28(2):776–803, 2006.

[9] C.-H. L. Eng, M. Lawson, Q. Zhu, R. Dries, N. Koulena, Y. Takei, J. Yun, C. Cronin, C. Karp, G.-C. Yuan, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*, 568(7751):235–239, 2019.

[10] A. Gessner, J. Gonzalez, and M. Mahsereci. Active multi-information source Bayesian quadrature. In *Uncertainty in Artificial Intelligence*, pages 712–721. PMLR, 2020.

[11] X. Gong and Y. Pan. Multi-fidelity Bayesian experimental design to quantify extreme-event statistics. *arXiv preprint arXiv:2201.00222*, 2022.

[12] A. A. Gorodetsky, J. D. Jakeman, G. Geraci, and M. S. Eldred. MFNets: multi-fidelity data-driven networks for Bayesian learning and prediction. *International Journal for Uncertainty Quantification*, 10(6), 2020.

[13] R. B. Gramacy and H. K. Lee. Adaptive design and analysis of supercomputer experiments. *Technometrics*, 51(2):130–145, 2009.

[14] P. Grünwald, T. Van Ommen, et al. Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Analysis*, 12(4):1069–1103, 2017.

[15] G. W. Gundersen, D. Cai, C. Zhou, B. E. Engelhardt, and R. P. Adams. Active multi-fidelity Bayesian online changepoint detection. In *Uncertainty in Artificial Intelligence*, pages 1916–1926. PMLR, 2021.

[16] R. Heide, A. Kirichenko, P. Grunwald, and N. Mehta. Safe-Bayesian generalized linear regression. In *International Conference on Artificial Intelligence and Statistics*, pages 2623–2633. PMLR, 2020.

[17] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

[18] A. J. Keane. Cokriging for robust design optimization. *AIAA journal*, 50(11):2351–2364, 2012.

[19] M. C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000.

[20] S. Li, W. Xing, R. Kirby, and S. Zhe. Multi-fidelity Bayesian optimization via deep neural networks. *Advances in Neural Information Processing Systems*, 33:8521–8531, 2020.

[21] E. Lubeck and L. Cai. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature methods*, 9(7):743–748, 2012.

[22] J. W. Miller and D. B. Dunson. Robust Bayesian inference via coarsening. *Journal of the American Statistical Association*, 2018.

[23] S. G. Rodriques, R. R. Stickels, A. Goeva, C. A. Martin, E. Murray, C. R. Vanderburg, J. Welch, L. M. Chen, F. Chen, and E. Z. Macosko. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 363(6434):1463–1467, 2019.

[24] J. Song, Y. Chen, and Y. Yue. A general framework for multi-fidelity Bayesian optimization with gaussian processes. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3158–3167. PMLR, 2019.

[25] P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016.

[26] S. Walker and N. L. Hjort. On Bayesian consistency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):811–821, 2001.

[27] Y. Wang, A. Kucukelbir, and D. M. Blei. Robust probabilistic modeling with Bayesian data reweighting. In *International Conference on Machine Learning*, pages 3646–3655. PMLR, 2017.

[28] J. Wu, S. Toscano-Palmerin, P. I. Frazier, and A. G. Wilson. Practical multi-fidelity Bayesian optimization for hyperparameter tuning. In *Uncertainty in Artificial Intelligence*, pages 788–798. PMLR, 2020.

[29] X. Xi, F.-X. Briol, and M. Girolami. Bayesian quadrature for multiple related integrals. In *International Conference on Machine Learning*, pages 5373–5382. PMLR, 2018.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

(a) If your work uses existing assets, did you cite the creators? [Yes]
(b) Did you mention the license of the assets? [N/A]
(c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]