A. Appendix

In this section, we illustrate the additive hybrid diffusion kernel (Equation 4.10) by providing a running example.

A.1. Running example for additive hybrid diffusion kernel

We illustrate the additive hybrid diffusion kernel and its recursive computation using a 3-dimensional hybrid space, where the first two dimensions correspond to discrete subspace and the last dimension correspond to continuous subspace. Let k_1, k_2, k_3 be the base kernels for first, second, and third dimension respectively. The additive diffusion kernel can be computed recursively step-wise as shown below:

$$
\begin{aligned}
\mathcal{K}_1 &= \theta_1^2 \cdot (k_1 + k_2 + k_3), & \mathcal{S}_1 &= (k_1 + k_2 + k_3)\\
\mathcal{K}_2 &= \theta_2^2 \cdot (k_1k_2 + k_1k_3 + k_2k_3), & \mathcal{S}_2 &= (k_1^2 + k_2^2 + k_3^2)\\
\mathcal{K}_3 &= \theta_3^2 \cdot (k_1k_2k_3), & \mathcal{S}_3 &= (k_1^3 + k_2^3 + k_3^3)\\
\mathcal{K}_0 &= 1; \\
\mathcal{K}_1 &= \theta_1^2 \cdot S_1; \\
\mathcal{K}_2 &= \theta_2^2 \cdot \frac{1}{2}\left(\mathcal{K}_1 \cdot S_1 - S_2\right); \\
\mathcal{K}_3 &= \theta_3^2 \cdot \frac{1}{3}\left(\mathcal{K}_2 \cdot S_1 - \mathcal{K}_1 \cdot S_2 + S_3\right); \\
\mathcal{K}_{HYB} &= \mathcal{K}_1 + \mathcal{K}_2 + \mathcal{K}_3
\end{aligned}
$$

B. Additional Experimental Details

B.1. Real world benchmarks

1) Pressure vessel design optimization. The objective function (cost of cylindrical pressure vessel design) $\mathcal{F}(x)$ for this domain is given below:

$$
\begin{aligned}
\min_{\{x_1,x_2,x_3,x_4\}} \; & 0.6224x_1x_3x_4 + 1.7781x_2x_3^2 + \\
& 3.1661x_1^2x_4 + 19.84x_1^2x_3
\end{aligned} \tag{B.1}
$$

where x_1, x_2 are discrete variables (thickness of shell and head of pressure vessel) lying in $\{1, \cdots, 100\}$ and $x_3 \in [10, 200], x_4 \in [10, 240]$ are continuous variables (inner radius and length of cylindrical section).

2) Welded beam design optimization. The objective function (cost of fabricating welded beam) $\mathcal{F}(x)$ for this domain is:

$$
\min_{\{x_1,x_2,x_3,x_4,x_5,x_6\}} (1 + G_1)(x_1x_5 + x_4)x_3^2 + G_2x_5x_6(L + x_4) \tag{B.2}
$$

where $x_1 \in \{0, 1\}, x_2 \in \{0, 1, 2, 3\}$ are discrete variables, $x_3 \in [0.0625, 2], x_4 \in [0, 20], x_5 \in [2, 20], x_6 \in [0.0625, 2]$ are continuous variables, G_1 is the cost per volume of the welded material, and G_2 is the cost per volume of the bar stock. The constants (G_1, G_2, L), which are dependent on the second discrete variable x_2, are given in (Deb & Goyal, 1996; Reklaitis et al., 1983).

3) Speed reducer design optimization. The objective function (weight of speed reducer) $\mathcal{F}(x)$ for this domain is:

$$
\begin{aligned}
\min_{\{x_1,x_2,x_3,x_4,x_5,x_6,x_7\}} \; & 0.79x_2x_3^2(3.33x_1^3 + 14.93x_1 - 43.09) \\
& - 1.51x_2(x_6^2 + x_7^2) + 7.48(x_6^3 + x_7^3) + 0.79(x_4x_6^2 + x_5x_7^2)
\end{aligned} \tag{B.3}
$$

where $x_1 \in \{17, 18 \cdots, 28\}$ represents the discrete variable (number of teeth on pinion), $x_2 \in [2.6, 3.6], x_3 \in [0.7, 0.8], x_4 \in [7.3, 8.3], x_5 \in [0.7, 0.8], x_6 \in [2.9, 3.9], x_7 \in [5, 5.5]$ represents the continuous variables (face width, teeth module, lengths of shafts between bearings, and diameters of the shafts respectively).

The above three benchmarks are usually described with *known* constraints in a declarative manner. However, for simplicity, we consider their unconstrained version for evaluation in this paper. If required, since the constraints are known, we can easily avoid searching for invalid solutions by using an appropriate acquisition function optimizer within HyBO.

4) Optimizing control for robot pushing. This domain was taken from this URL [7]. We consider a hybrid version of this problem by discretizing the location parameters ($x_1, x_2, x_3, x_4 \in \{-5, -4, \cdots, 5\}$ and $x_5, x_6, x_7, x_8 \in \{-10, -9, \cdots, 10\}$). There are two other discrete variables corresponding to simulation steps $x_9, x_{10} \in \{2, 3, 4, \cdots, 30\}$ and two continuous variables x_{11}, x_{12} lying in $[0, 2\pi]$.

5) Calibration of environmental model. The details of the objective function for this domain are available in (Bliznyuk et al., 2008; Astudillo & Frazier, 2019). The single discrete variable has 284 candidate values lying in the set $\{30.01, 30.02, \cdots 30.285\}$. There are three continuous variables lying in the range: $x_2 \in [7, 13], x_3 \in [0.02, 0.12], x_4 \in [0.01, 3]$.

6) Hyper-parameter optimization. The type and range for different hyper-parameters considered in this domain are given in Table 5. We employed the scikit-learn (Pedregosa et al., 2011) neural network implementation for this benchmark.

[7] `https://github.com/zi-w/Ensemble-Bayesian-Optimization/tree/master/test_functions`

Hyperparameter	Type	Range
Hidden layer size	Discrete	$\{40, 60, \cdots, 300\}$
Type of activation	Discrete	{'identity', 'logistic', 'tanh', 'relu'}
Batch size	Discrete	$\{40, 60, \cdots, 200\}$
Type of learning rate	Discrete	{'constant', 'invscaling', 'adaptive'}
Early stopping	Discrete	True/False
Learning rate initialization	Continuous	$[0.001, 1]$
Momentum	Continuous	$[0.5, 1]$
Alpha parameter	Continuous	$[0.0001, 1]$

Table 5. Type and range of hyper-parameters considered for the HPO benchmark.

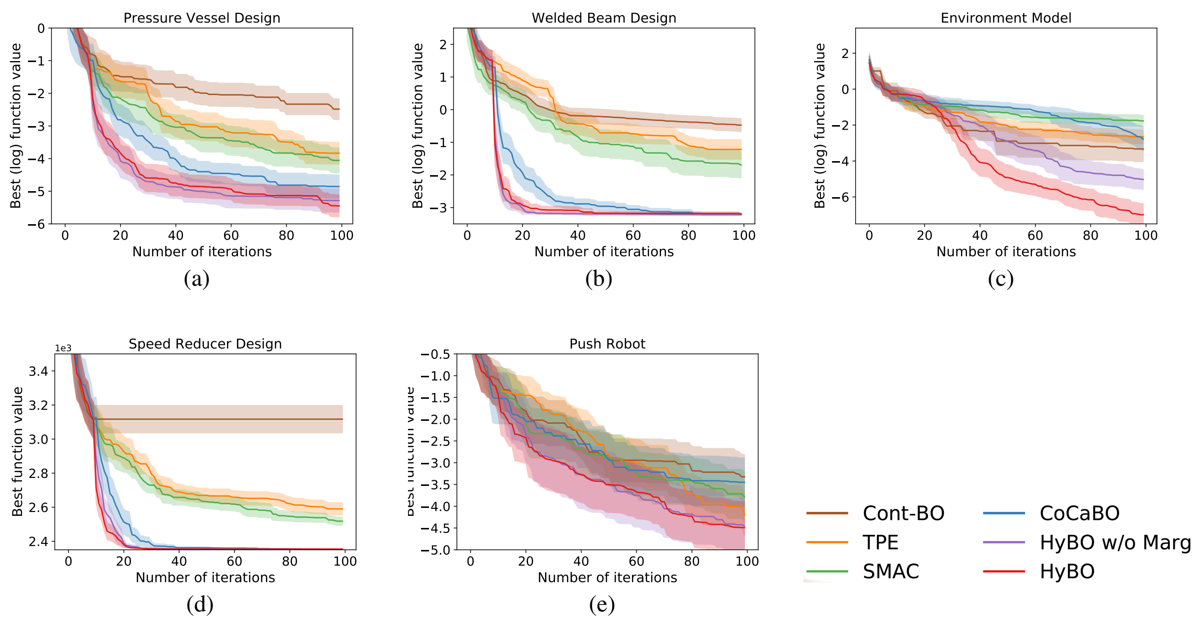

Figure 4. Results comparing the proposed HyBO approach with state-of-the-art baselines on multiple real world benchmarks. These figures also contain HyBO without marginalization and Cont-BO results.

C. Additional Results

Results for real-world benchmarks. Figure 4 extends the plots of Figure 3 by including the performance of Cont-BO and HyBO w/o Marg on the real-world benchmarks. The results show similar trend where Cont-BO performs worse than all other methods showing the need to take into account the hybrid input structure. Also, the performance of HyBO w/o Marg remains similar to HyBO (except on calibration of environment model) demonstrating the effective modeling strength of additive hybrid diffusion kernel.

Comparison with (Garrido-Merchán & Hernández-Lobato, 2020) As mentioned in our related work, this is an interesting approach for BO over discrete spaces but it is specific to discrete spaces alone. Since our problem setting considers hybrid input spaces, we performed experiments using this method for the discrete part and using the standard BO approach for the continuous part with HyBO's AFO procedure. Results of this approach (referred as G-M et al.,) on the 4 synthetic benchmarks are shown in Table 6. The best function value achieved after 200 iterations and averaged over 25 different runs (same configuration as described in the main paper) is shown. We also add another baseline named Vanilla BO (GP with RBF kernel to model hybrid space + HyBO's AFO procedure) in Table 6. It is evident from the results that HyBO performs significantly better.

Benchmark	**HyBO**	**G-M et al.**	**Vanilla BO**
Synthetic Function 1	79.7	99.4	86.2
Synthetic Function 2	394.6	420	407
Synthetic Function 3	81.1	143	135
Synthetic Function 4	395.2	458	456.8

Table 6. Results for additional baseline experiments

References

Astudillo, R. and Frazier, P. Bayesian optimization of composite functions. volume 97 of *Proceedings of Machine Learning Research*, pp. 354–363. PMLR, 09–15 Jun 2019.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. The nonstochastic multiarmed bandit problem. *SIAM Journal of Computing*, 32(1):48–77, 2002.

Baptista, R. and Poloczek, M. Bayesian optimization of combinatorial structures. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 462–471, 2018.

Belakaria, S., Deshwal, A., and Doppa, J. R. Max-value entropy search for multi-objective Bayesian optimization. In *NeurIPS*, 2019.

Belakaria, S., Deshwal, A., and Doppa, J. R. Max-value entropy search for multi-objective Bayesian optimization with constraints. *CoRR*, abs/2009.01721, 2020a. URL https://arxiv.org/abs/2009.01721.

Belakaria, S., Deshwal, A., and Doppa, J. R. Multi-fidelity multi-objective Bayesian optimization: An output space entropy search approach. In *AAAI conference on Artificial Intelligence (AAAI)*, 2020b.

Belakaria, S., Deshwal, A., and Doppa, J. R. Uncertainty aware search framework for multi-objective Bayesian optimization with constraints. *CoRR*, abs/2008.07029, 2020c. URL https://arxiv.org/abs/2008.07029.

Belakaria, S., Deshwal, A., and Doppa, J. R. Information-theoretic multi-objective Bayesian optimization with continuous approximations. *CoRR*, abs/2009.05700, 2020d. URL https://arxiv.org/abs/2009.05700.

Belakaria, S., Deshwal, A., Jayakodi, N. K., and Doppa, J. R. Uncertainty-aware search framework for multi-objective Bayesian optimization. In *AAAI*, 2020e.

Belakaria, S., Jackson, D., Cao, Y., Doppa, J. R., and Lu, X. Machine learning enabled fast multi-objective optimization for electrified aviation power system design. In *IEEE Energy Conversion Congress and Exposition (ECCE)*, 2020f.

Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. Algorithms for hyper-parameter optimization. In Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems 24*, pp. 2546–2554, 2011.

Bliznyuk, N., Ruppert, D., Shoemaker, C., Regis, R., Wild, S., and Mugunthan, P. Bayesian calibration and uncertainty analysis for computationally expensive models using optimization and radial basis function approximation. *Journal of Computational and Graphical Statistics*, 17 (2):270–294, 2008.

Cagnina, L., Esquivel, S., and Coello, C. Solving engineering optimization problems with the simple constrained particle swarm optimizer. *Informatica*, 32:319–326, 2008.

Chung, F. R. and Graham, F. C. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.

Daxberger, E. Personal communication about MiVaBO implementation and code.

Daxberger, E., Makarova, A., Turchetta, M., and Krause, A. Mixed-variable bayesian optimization. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pp. 2633–2639, 7 2020.

Deb, K. and Goyal, M. A combined genetic adaptive search (GeneAS) for engineering design. *Computer Science and Informatics*, 26:30–45, 1996.

Deshwal, A., Belakaria, S., and Doppa, J. R. Scalable combinatorial Bayesian optimization with tractable statistical models. *CoRR*, abs/2008.08177, 2020a. URL https://arxiv.org/abs/2008.08177.

Deshwal, A., Belakaria, S., Doppa, J. R., and Fern, A. Optimizing discrete spaces via expensive evaluations: A learning to search framework. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2020b.

Deshwal, A., Belakaria, S., and Doppa, J. R. Mercer features for efficient combinatorial Bayesian optimization. In *AAAI*, 2021.

Duvenaud, D. K., Nickisch, H., and Rasmussen, C. E. Additive gaussian processes. In Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems 24*, pp. 226–234, 2011.

Falkner, S., Klein, A., and Hutter, F. BOHB: robust and efficient hyper-parameter optimization at scale. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pp. 1436–1445, 2018.

Frazier, P. I. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.

Garrido-Merchán, E. C. and Hernández-Lobato, D. Dealing with categorical and integer-valued variables in bayesian optimization with gaussian processes. *Neurocomputing*, 380:20–35, 2020.

Gijsbers, P., LeDell, E., Poirier, S., Thomas, J., Bischl, B., and Vanschoren, J. An open source automl benchmark. *arXiv preprint arXiv:1907.00909.*, 2019. Accepted at AutoML Workshop at ICML 2019.

Greenhill, S., Rana, S., Gupta, S., Vellanki, P., and Venkatesh, S. Bayesian optimization for adaptive experimental design: A review. *IEEE Access*, 8:13937–13948, 2020.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. A kernel two-sample test. *Journal of Machine Learning Research*, 13(1):723–773, 2012.

Hansen, N. The cma evolution strategy: A tutorial. *arXiv preprint arXiv:1604.00772*, 2016.

Hutter, F., Hoos, H. H., and Leyton-Brown, K. Sequential model-based optimization for general algorithm configuration (extended version). Technical Report TR-2010-10, University of British Columbia, Department of Computer Science, 2010. Available online: http://www.cs.ubc.ca/~hutter/papers/10-TR-SMAC.pdf.

Hutter, F., Hoos, H. H., and Leyton-Brown, K. Sequential model-based optimization for general algorithm configuration. In *International conference on Learning and Intelligent Optimization*, pp. 507–523, 2011.

Kannan, B. K. and Kramer, S. N. An augmented lagrange multiplier based method for mixed integer discrete continuous optimization and its applications to mechanical design. *Journal of Mechanical Design*, 116(2):405–411, 06 1994.

Kondor, R. and Lafferty, J. Diffusion kernels on graphs and other discrete structures. In *Proceedings of the 19th International Conference on Machine Learning*, volume 2002, pp. 315–322, 2002.

Kondor, R. and Vert, J.-P. Diffusion kernels. *Kernel methods in computational biology*, pp. 171–192, 2004.

Lafferty, J. and Lebanon, G. Diffusion kernels on statistical manifolds. *Journal of Machine Learning Research*, 6(1): 129–163, 2005.

Li, L., Jamieson, K. G., DeSalvo, G., Rostamizadeh, A., and Talwalkar, A. Hyperband: a novel bandit-based approach to hyper-parameter optimization. *Journal of Machine Learning Research (JMLR)*, 18:185:1–185:52, 2017.

Mania, H., Ramdas, A., Wainwright, M. J., Jordan, M. I., and Recht, B. On kernel methods for covariates that are rankings. *Electronic Journal of Statistics*, 12(2):2537–2577, 2018.

Micchelli, C. A., Xu, Y., and Zhang, H. Universal kernels. *Journal of Machine Learning Research*, 7(Dec):2651–2667, 2006.

Mockus, J., Tiesis, V., and Zilinskas, A. The application of bayesian methods for seeking the extremum. *Towards Global Optimization*, 2(117-129), 1978.

Neal, R. M. Slice sampling. *Annals of statistics*, 31(3): 705–741, 6 2003.

Oh, C., Tomczak, J., Gavves, E., and Welling, M. Combinatorial bayesian optimization using the graph cartesian product. In *Advances in Neural Information Processing Systems*, pp. 2910–2920, 2019.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Rahimi, A. and Recht, B. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems 20*, pp. 1177–1184, 2007.

Rasmussen, C. E. and Williams, C. K. I. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, 2006.

Reklaitis, G. V., Ravindran, A., and Ragsdell, K. M. *Engineering optimization: methods and applications*. Wiley New York, 1983.

Roustant, O., Padonou, E., Deville, Y., Clément, A., Perrin, G., Giorla, J., and Wynn, H. Group kernels for gaussian process metamodels with categorical inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 8(2):775–806, 2020.

Ru, B., Alvi, A. S., Nguyen, V., Osborne, M. A., and Roberts, S. J. Bayesian optimisation over multiple continuous and categorical inputs. In *International Conference on Machine Learning (ICML)*, 2020.

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and de Freitas, N. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104 (1):148–175, 2016.

Sriperumbudur, B. K., Fukumizu, K., and Lanckriet, G. R. Universality, characteristic kernels and rkhs embedding of measures. *Journal of Machine Learning Research*, 12 (7), 2011.

Steinwart, I. On the influence of the kernel on the consistency of support vector machines. *Journal of Machine Learning Research*, 2(Nov):67–93, 2001.

Steinwart, I. and Christmann, A. *Support vector machines*. Springer Science & Business Media, 2008.

Steinwart, I., Thomann, P., and Schmid, N. Learning with hierarchical gaussian kernels. *arXiv preprint arXiv:1612.00824*, 2016.

Tanabe, R. and Ishibuchi, H. An easy-to-use real-world multi-objective optimization problem suite. *Applied Soft Computing*, 89:106078, 2020.

Tušar, T., Brockhoff, D., and Hansen, N. Mixed-integer benchmark problems for single-and bi-objective optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 718–726, 2019.

Verel, S., Derbel, B., Liefooghe, A., Aguirre, H., and Tanaka, K. A surrogate model based on walsh decomposition for pseudo-boolean functions. In *International Conference on Parallel Problem Solving from Nature*, pp. 181–193. Springer, 2018.

Wang, Z., Gehring, C., Kohli, P., and Jegelka, S. Batched large-scale bayesian optimization in high-dimensional spaces. In *International Conference on Artificial Intelligence and Statistics*, pp. 745–754, 2018.

Zhou, Z., Belakaria, S., Deshwal, A., Hong, W., Doppa, J. R., Pande, P. P., and Heo, D. Design of multi-output switched-capacitor voltage regulator via machine learning. In *DATE*, 2020.