**Supplementary Information**

# Bayesian optimization with experimental failure for high-throughput materials growth

Yuki K. Wakabayashi,[1,*,†] Takuma Otsuka,[2,*,‡] Yoshiharu Krockenberger,[1] Hiroshi Sawada,[2] Yoshitaka Taniyasu,[1] and Hideki Yamamoto[1]

[1]*NTT Basic Research Laboratories, NTT Corporation, Atsugi, Kanagawa 243-0198, Japan*
[2]*NTT Communication Science Laboratories, NTT Corporation, Soraku-gun, Kyoto 619-0237, Japan*

[*]These authors contributed equally to this work.

[†]Corresponding author: yuuki.wakabayashi.we@hco.ntt.co.jp

[‡]Corresponding author: takuma.otsuka.uf@hco.ntt.co.jp

**Algorithm** Bayesian optimization with experimental failure

**Input:** initial data $\mathcal{D} \coloneqq \{(\mathbf{x}_i, y_i)\}_{i=1,\dots,\#\text{ of init}}$, budget $N$

**Output:** best parameter $\mathbf{x}_i$ such that $i = \mathrm{argmax}_{1\le i\le N}\, \tilde{y}_i$

**while** $|\mathcal{D}| < N$ **do**

$n \leftarrow |\mathcal{D}| + 1$

preprocess data into $\widetilde{\mathcal{D}} = \{(\tilde{\mathbf{x}}_n, \tilde{y}_n)\}_{i=1}^{n-1}$ from $\mathcal{D}$ by Eqs. (1) and (2)

fit Gaussian process using $\widetilde{\mathcal{D}}$ to calculate $m$ and $s^2$ in Eqs. (3) and (4)

find next parameter $\tilde{\mathbf{x}}_n = \mathrm{argmax}_{\tilde{\mathbf{x}}'}\, a_{\mathrm{EI}}(\tilde{\mathbf{x}}')$

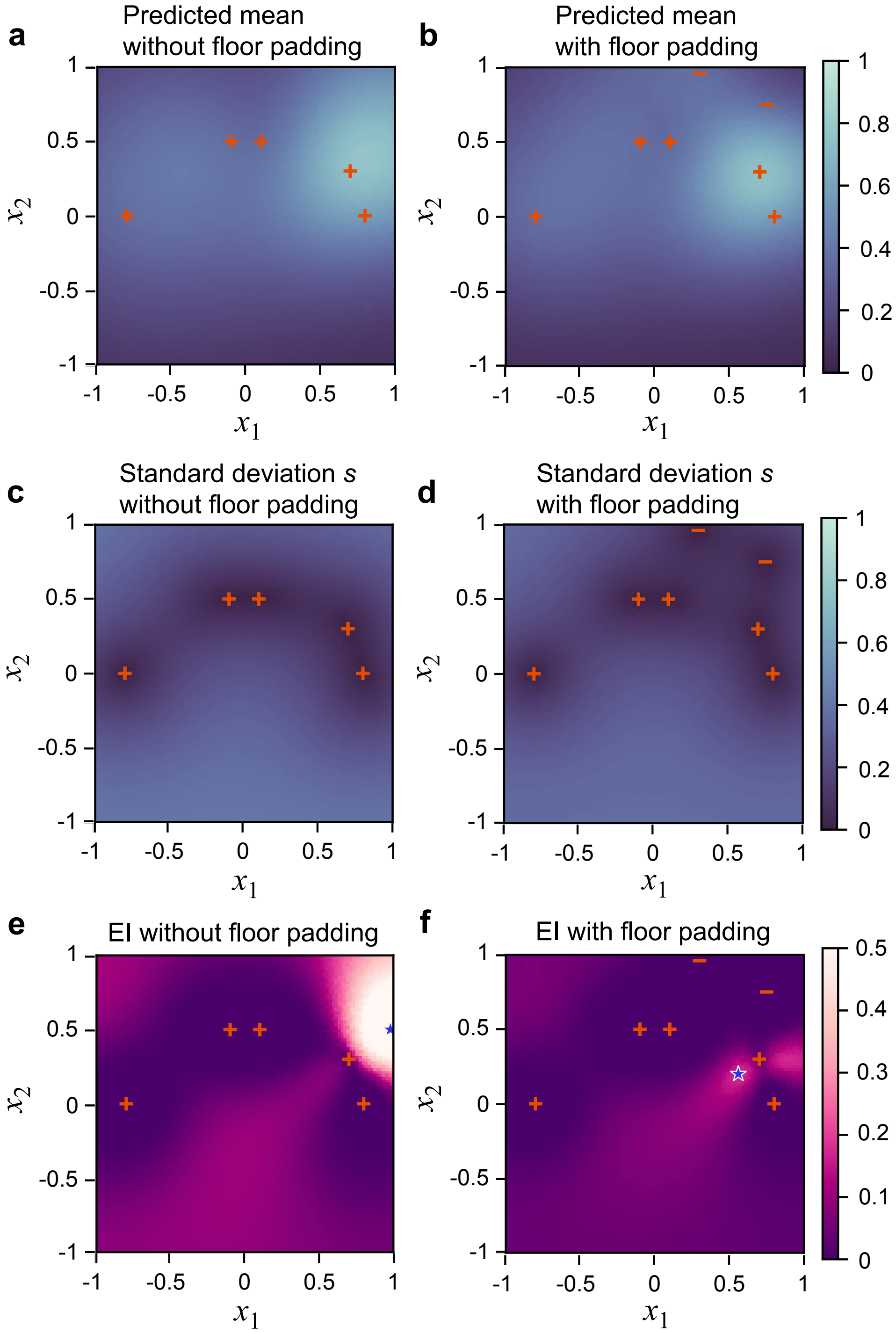
rescale to original parameter $\mathbf{x}_n$ from $\tilde{\mathbf{x}}_n$

obtain evaluation for parameter $\mathbf{x}_n$ as $y_n$

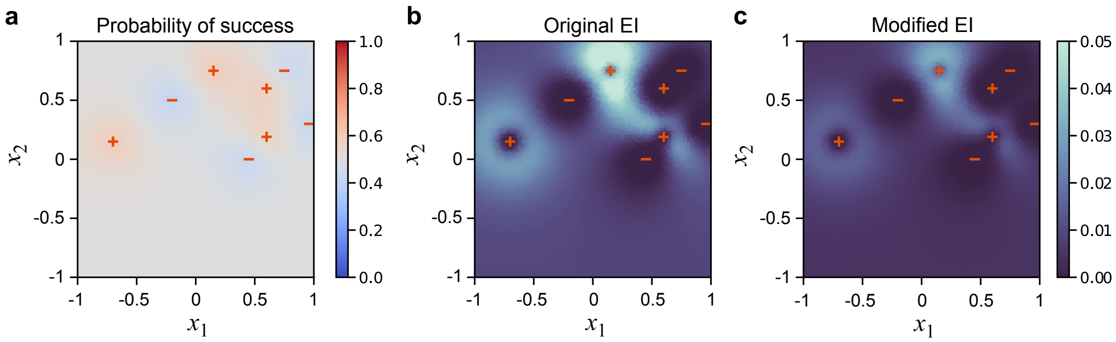
add to raw data $\mathcal{D} \leftarrow \mathcal{D} \cup \{(\mathbf{x}_n, y_n)\}$

**end while**

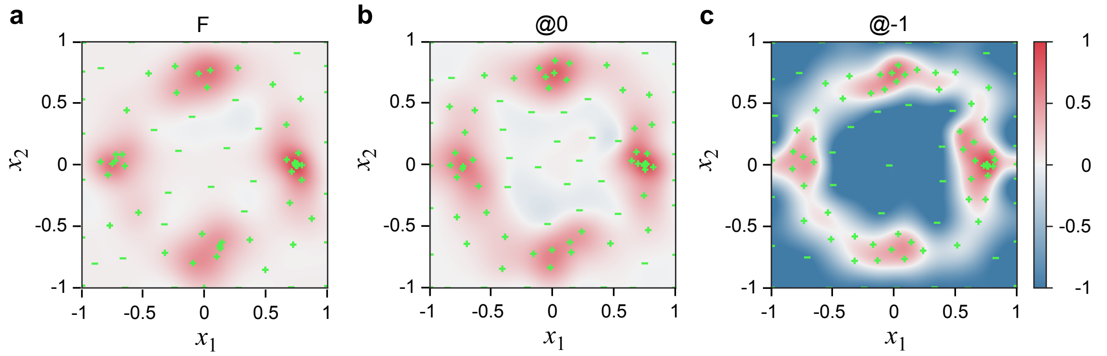
**Supplementary Figure 1 Pseudocode of Bayesian optimization.**




**Supplementary Figure 2 Effect of the floor padding on the prediction and EI values.** Predicted mean (**a,b**), standard deviation *s* (**c,d**), and EI (**e,f**) for the Circle function with five observations '+' (**a,c,e**) and with five observations '+' as well as two failed ones '–' (**b,d,f**). In (**e**) and (**f**), the blue stars indicate the next conditions at which the highest EI values are obtained.




**Supplementary Figure 3 Effect of binary classifier on expected improvement.** The prediction by method FB from eight observations of the Hole function with four successful '+' and four failed '−' observations. (**a**) Predicted success probability. (**b**) The original EI function used by method F. (**c**) Modified EI function of method FB, given as the product of (**a**) and (**b**).




**Supplementary Figure 4 Predicted mean after running BO with 100 experiments.** Successful observations were marked with ‘+’ while failed observations were marked by ‘−’. Failed observations were treated by floor padding (F) trick (**a**), replacement with 0 (@0) (**b**) and replacement with −1 (@−1) (**c**).

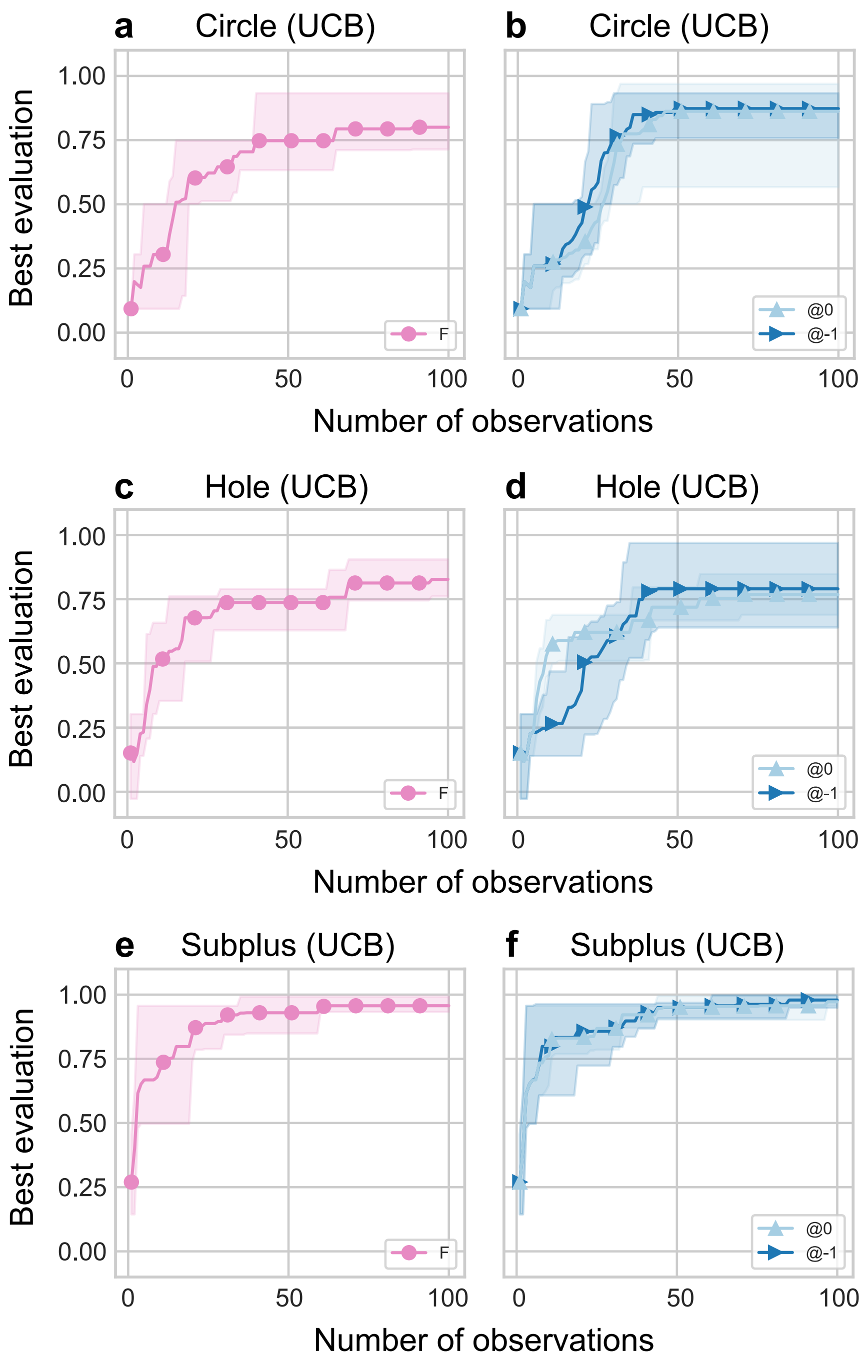
**Supplementary Note - Floor padding trick with other acquisition function**

The floor padding trick can also be combined with other acquisition functions in the literature other than the expected improvement defined as Eq. (5). This Note shows the result with upper confidence bound (UCB) criterion. Given the predicted mean $m(\tilde{\mathbf{x}}')$ and variance $s^2(\tilde{\mathbf{x}}')$ as in Eqs. (3) and (4), the UCB is calculated as follows:

$$a_{\mathrm{UCB}}(\tilde{\mathbf{x}}') = m(\tilde{\mathbf{x}}') + \kappa_n s(\tilde{\mathbf{x}}'),$$

where $\kappa_n > 0$ trades off the exploration and exploitation of the prediction from $n$ observations[S1]. We adopted $\kappa_n = \sqrt{n}$ to encourage the exploration as the number of observations increases.

Supplementary Figure 5 presents the optimization results by methods F, @0, and @−1 with the UCB acquisition function for the Circle, Hole and Softplus functions. Since the combination of UCB and binary classifier is non-trivial, we omitted the results of methods using the binary classifier. For the Circle and Hole functions, method F outperformed @0 and @−1 within the first 30 iterations. After that, method F tended to start extensive exploration in the parameter space due to the choice of increasing $\kappa_n$, which resulted in frequent failures. In contrast, methods @0 and @−1 made conservative decisions to avoid failures despite increasing $\kappa_n$ owing to the low padding constants. As a result, the method F failed to reach the maximum value of the Circle and Hole functions once out of five runs. While different design of $\kappa_n$ may improve the optimization, we leave it to future work. Finally, the relatively smooth shape of Softplus function helped all the methods find near-optimal parameter in the early stage of the optimization. Thus, the EI function is more suitable since the methods using the UCB function did not reach the maximum value of the Circle and Hole functions.




**Supplementary Figure 5 Optimization results by the methods F (a,c,e), @0, and @−1 (b,d,f) using the UCB functions.** Best evaluation values averaged over five runs as a function of the number of observations for the Circle (**a**,**b**), Hole (**c**,**d**), and Subplus (**e**,**f**) functions. Shaded areas indicate the range of the best and worst values of five runs.

**Supplementary Reference**

S1. Srinivas, N. Krause, A. Kakade, S. and Seeger, M. Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. *Proc. of the 27th Int'l. Conf. Mach. Learn.* (2010).