

410 6 Supplementary material

411 6.1 Animal ethics statement

412 All experiments on animals were conducted with approval of the Animal Care and Use Committee of
413 the University of California, Berkeley.

414 6.2 Compute

415 All computational procedures were performed either on a desktop workstation running Ubuntu 18.04
416 with an Intel Xeon E5-2620 v4 CPU, four GTX 1080 Ti GPUs, and 112GB RAM, or on the Axon
417 computer cluster based at [redacted for anonymity] using nodes comprised of two Xeon E5-2660 v4
418 CPUs, eight GTX 1080 Ti GPUs, and 125GB RAM.

419 6.3 Broader societal impact

420 Our work is significant for interventional approaches to studying the brain and its connection
421 to disease. By minimising off-target activation, Bayesian target optimisation could enable (e.g.)
422 more precise synaptic connectivity mapping, improving our understanding of neural circuitry. This
423 advancement has potential implications for understanding brain disorders like epilepsy, where
424 abnormal synaptic connections are central to seizure generation and propagation. Deepening our
425 understanding of these diseases can lead to enhanced targeted interventions and more effective
426 therapeutic strategies, benefiting individuals with neurological disorders.

427 6.4 Code availability

428 An open-source implementation of Bayesian target optimisation is available in Python at <https://anonymous.4open.science/r/bataro-4401/>.
429

430 6.5 Single-target holographic stimulus optimisation with posterior uncertainty

431 Here we provide further mathematical details for optimising holographic stimuli. First, we develop
432 the approach for single optogenetic targets, as this is most closely related to existing GP-based
433 receptive field inference techniques. The single-target case also allows us to have a full treatment
434 of posterior uncertainty (unlike for optimising ensemble stimuli) which may be desired in certain
435 applications.

436 **Optogenetic receptive field model.** We use a GP-Bernoulli approach to model the response y_{nt} of
437 neuron n on trial t to a single-target stimulus \mathbf{x}_t ,

$$y_{nt} \sim \text{Bernoulli}(\sigma(g_n(\mathbf{x}_t))), \quad (9)$$

438 where the stimulus $\mathbf{x}_t = (c_{1t}, c_{2t}, I_t) \in \mathbb{R}^3$ represents the two-dimensional coordinates and laser
439 power of the t -th hologram. Each ORF follows a three-dimensional GP prior $g_n \sim \mathcal{GP}(m_n(\cdot), k(\cdot, \cdot))$,
440 where m_n and k again are the mean and covariance functions of the GP.

441 **Posterior inference.** Unlike for ensemble stimulation, for single-target stimulation we do not require
442 that the ORF g_n is non-negative. Consequently, the posterior of g_n is a GP, which allows us to work
443 with a full description of posterior uncertainty. To compute the posterior, we use the conventional
444 Laplace approximation. Briefly, this consists of approximating the posterior using a multivariate
445 normal $q(g_n | \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n) = \text{Normal}(g_n | \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n) \approx p(g_n | \mathbf{y}_n, \mathbf{X}, \phi)$. The mean $\boldsymbol{\mu}_n$ is obtained by
446 maximising the log-posterior, given by the expression

$$\ln p(g_n | \mathbf{y}_n, \mathbf{X}, \phi) = \sum_{t=1}^T \ln p(y_{nt} | \mathbf{x}_t, g_n) + \ln p(g_n(\mathbf{x}_1), \dots, g_n(\mathbf{x}_T) | \phi) + \text{const}, \quad (10)$$

447 where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_T)$ and where const does not depend on g_n . Since the posterior is log-concave
448 in g_n , we use Newton’s method to identify the global optimum of [Equation 10](#), and adaptively set the
449 Newton step-size using a standard backtracking line-search method. Letting $\mathbf{H} = \nabla \nabla_{g_n} \ln p(g_n |$
450 $\mathbf{y}_n, \mathbf{X}, \phi)$ be the Hessian of the log-posterior, the posterior covariance matrix is obtained by setting
451 $\boldsymbol{\Sigma}_n = -\mathbf{H}^{-1} |_{g_n=\boldsymbol{\mu}_n}$.

452 **Target optimisation.** Let $G = (g_1, \dots, g_N)$, and define the predicted evoked activity for single
453 holographic targets as $\hat{y}(\mathbf{x}, G) = (\sigma(g_1(\mathbf{x})), \dots, \sigma(g_N(\mathbf{x})))$. To minimise the error between a target
454 binary activity pattern $\Omega \in \{0, 1\}^N$ and the predicted evoked activity, we solve an optimisation
455 problem that accounts for the uncertainty in the ORF estimates:

$$\mathbf{x}_{\text{optimal}} = \underset{\mathbf{x}}{\operatorname{argmin}} \mathbb{E}_{q(G|\mu, \Sigma)} \left[\|\Omega - \hat{y}(\mathbf{x}, G)\|^2 \right] \quad \text{such that} \quad 0 \leq I \leq I_{\max}, \quad (11)$$

456 where $q(G | \mu, \Sigma) = \prod_{n=1}^N q(g_n | \mu_n, \Sigma_n)$ gives the joint posterior across all ORFs. To solve
457 **Equation 11**, we first sample ORFs $g_n^{(s)}$ (for $s = 1, \dots, S$) from their posterior distributions to
458 approximate the expected error at the current estimate \mathbf{x}^* ,

$$\mathbb{E}_{q(G|\mu, \Sigma)} \left[\|\Omega - \hat{y}(\mathbf{x}^*, G)\|^2 \right] \approx \frac{1}{S} \sum_{s=1}^S \sum_{n=1}^N \left(\Omega_n - \sigma(g_n^{(s)}(\mathbf{x}^*)) \right)^2. \quad (12)$$

459 Then, we compute the partial derivative (in dimension d) of the expected error by differentiating
460 through the Monte Carlo approximation,

$$\frac{\partial}{\partial x_d^*} \mathbb{E}_{q(G|\mu, \Sigma)} \left[\|\Omega - \hat{y}(\mathbf{x}^*, G)\|^2 \right] \approx -\frac{2}{S} \sum_{s=1}^S \sum_{n=1}^N (\Omega_n - \sigma(g_n^{(s)}(\mathbf{x}^*))) \sigma'(g_n^{(s)}(\mathbf{x}^*)) \frac{\partial}{\partial x_d^*} g_n^{(s)}(\mathbf{x}^*). \quad (13)$$

461 Next we must evaluate the partial derivative on the right-hand side of **Equation 13**. We use the fact
462 that a GP and its derivative are jointly GP-distributed, and hence infer the derivative from observations
463 of the ORF. The covariance between a GP and its derivative is given by [40, Sec 9.4]

$$\operatorname{cov} \left(g_n(\mathbf{x}_t), \frac{\partial}{\partial x_d^*} g_n(\mathbf{x}^*) \right) = \frac{\partial k(\mathbf{x}_t, \mathbf{x}^*)}{\partial x_d^*} = \frac{\alpha^2}{\lambda_d^2} (x_{dt} - x_d^*) \exp \left(-\frac{\|\mathbf{x}_t - \mathbf{x}^*\|^2}{2\lambda_d^2} \right), \quad (14)$$

464 where the second equality is specific to the RBF covariance. Thus, we can use **Equation 14** to obtain
465 the posterior predictive mean for the derivative GPs in closed form as [46, Sec 2.7]

$$\mathbb{E}_{q(g_n|\mu_n, \Sigma_n)} \left[\frac{\partial g_n(\mathbf{x}^*)}{\partial x_d^*} \right] = \frac{\partial m_n(\mathbf{x}^*)}{\partial x_d^*} + \operatorname{cov} \left(g_n(\mathbf{X}), \frac{\partial g_n(\mathbf{x}^*)}{\partial x_d^*} \right)^\top \mathbf{K}^{-1}(\mu_n - m_n(\mathbf{X})). \quad (15)$$

466 Here $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_T)$ is the collection of unique points on the ORF probed during calibration.
467 If **Equation 15** is combined with an expression for the posterior predictive variance, one obtains a
468 full predictive distribution over derivative functions consistent with the observed neural responses.
469 However, rather than working with this full distribution, we instead use **Equation 15** to approximate
470 the derivatives of the Monte Carlo samples by replacing the posterior mean μ_n with a Monte Carlo
471 sample,

$$\frac{\partial g_n^{(s)}(\mathbf{x}^*)}{\partial x_d^*} \approx \frac{\partial m_n(\mathbf{x}^*)}{\partial x_d^*} + \operatorname{cov} \left(g_n(\mathbf{X}), \frac{\partial g_n(\mathbf{x}^*)}{\partial x_d^*} \right)^\top \mathbf{K}^{-1}(g_n^{(s)}(\mathbf{X}) - m_n(\mathbf{X})). \quad (16)$$

472 **Equation 16** then allows us to define a closed-form approximate gradient $\tilde{\nabla}_{\mathbf{x}^*} g_n^{(s)}$ at test point \mathbf{x}^* ,
473 defined as

$$\tilde{\nabla}_{\mathbf{x}^*} g_n^{(s)} = \left[\frac{\partial g_n^{(s)}(\mathbf{x}^*)}{\partial x_1^*}, \dots, \frac{\partial g_n^{(s)}(\mathbf{x}^*)}{\partial x_D^*} \right]^\top, \quad (17)$$

474 which we use in the single-target projected gradient descent algorithm (**Algorithm 2**). Note that one
475 could also consider a quadrature approach to solving **Equation 12**, which may be more efficient than
476 Monte Carlo sampling. However, the presentation of the Monte Carlo approach is instructive for
477 deriving the optimisation of ensemble stimuli below.

478 6.6 Additional details on ensemble stimulus optimisation approach

479 The approach for optimising holographic ensemble stimuli is based on the approach for single-target
480 optimisation, but modified to account for differences in the ORF model and inference. In particular,
481 we again seek to minimise the error between a target activity pattern Ω and the predicted evoked

Algorithm 2: Projected Monte Carlo gradient descent algorithm for optimising single-target holograms

- 1 Infer ORF posterior $q(G \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ from calibration data $\{\mathbf{y}_n\}_{n=1}^N$, \mathbf{X} using the Laplace approximation.
 - 2 Precompute the negative of the Hessian $\mathbf{W}_n = -\nabla\nabla \ln p(\mathbf{y}_n \mid \mathbf{X}, g_n) \big|_{g_n=\mu_n}$ for each n .
 - 3 Initialise \mathbf{x} to random location near soma of target neuron and with random laser power.
 - 4 **while** *target not converged* **do**
 - 5 **for** $n = 1, \dots, N$ **do**
 - 6 Compute mean and variance of posterior predictive distribution at current target estimate \mathbf{x} via $\mu_n(\mathbf{x}) = m_n(\mathbf{x}) + k(\mathbf{X}, \mathbf{x})^\top \mathbf{K}^{-1}(\boldsymbol{\mu}_n - m_n(\mathbf{X}))$, and $\sigma_n^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - k(\mathbf{X}, \mathbf{x})^\top (\mathbf{K} + \mathbf{W}_n^{-1})^{-1} k(\mathbf{X}, \mathbf{x})$.
 - 7 Sample ORFs at the current target estimate, $g_n^{(s)}(\mathbf{x}) \sim \text{Normal}(\mu_n(\mathbf{x}), \sigma_n^2(\mathbf{x}))$ for $s = 1, \dots, S$.
 - 8 Construct approximate gradients $\tilde{\nabla}_{\mathbf{x}} g_n^{(s)}$ for $s = 1, \dots, S$ using Equation 17.
 - 9 **end**
 - 10 Set $\boldsymbol{\delta}_{\mathbf{x}} = -\frac{2}{S} \sum_{s=1}^S \sum_{n=1}^N (\Omega_n - \sigma(g_n^{(s)}(\mathbf{x})) \sigma'(g_n^{(s)}(\mathbf{x})) \tilde{\nabla}_{\mathbf{x}} g_n^{(s)}(\mathbf{x}))$ as per Equation 13.
 - 11 Perform gradient descent update, $\mathbf{x} \leftarrow \mathbf{x} + \beta \boldsymbol{\delta}_{\mathbf{x}}$ with step-size β .
 - 12 Project laser power onto feasible domain, $I \leftarrow \min(I, I_{\max})$.
 - 13 **end**
-

482 activity, but now using the MAP estimates $\mathcal{G} = \{\hat{g}_n, \hat{\theta}_n\}_{n=1}^N$ in place of the full posterior distributions.
483 Let $\hat{y}(\mathbf{x}, \mathcal{G}) = (\sigma(\hat{\gamma}_1(\mathbf{x}) - \hat{\theta}_1), \dots, \sigma(\hat{\gamma}_N(\mathbf{x}) - \hat{\theta}_N))$ be the predicted population response to an
484 ensemble stimulus, where $\hat{\gamma}_n(\mathbf{x}) = \sum_{j=1}^J \hat{g}_n(\mathbf{x}^j)$. The optimal ensemble stimulus is now

$$\mathbf{x}_{\text{optimal}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\boldsymbol{\Omega} - \hat{y}(\mathbf{x}, \mathcal{G})\|^2 = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{n=1}^N \left(\Omega_n - \sigma(\hat{\gamma}_n(\mathbf{x}) - \hat{\theta}_n) \right)^2 \quad (18)$$

485 such that $0 \leq I \leq I_{\max}$. Evaluating the partial derivative of Equation 18 with respect to dimension d
486 of a test point \mathbf{x}^* yields,

$$\frac{\partial}{\partial x_d^*} \|\boldsymbol{\Omega} - \hat{y}(\mathbf{x}^*, \mathcal{G})\|^2 = -2 \sum_{n=1}^N (\Omega_n - \sigma(\hat{\gamma}_n(\mathbf{x}^*) - \hat{\theta}_n)) \sigma'(\hat{\gamma}_n(\mathbf{x}^*) - \hat{\theta}_n) \frac{\partial}{\partial x_d^*} \hat{\gamma}_n(\mathbf{x}^*). \quad (19)$$

487 The derivative on the right-hand side of Equation 19 is given by $\frac{\partial}{\partial x_d^*} \hat{\gamma}_n(\mathbf{x}) = \sum_{j=1}^J \frac{\partial}{\partial x_d^*} \hat{g}_n(\mathbf{x}^j)$,
488 which requires computing the derivative of $\hat{g}_n(\mathbf{x}^j)$. To evaluate this derivative, we use a similar trick
489 to Equation 16, but substituting the MAP estimate in place of the posterior mean or Monte Carlo
490 sample,

$$\frac{\partial}{\partial x_d^*} \hat{g}_n(\mathbf{x}^*) = \frac{\partial}{\partial x_d^*} m_n(\mathbf{x}^*) + \operatorname{cov} \left(g_n(\mathbf{X}), \frac{\partial}{\partial x_d^*} g_n(\mathbf{x}^*) \right)^\top \mathbf{K}^{-1} (\hat{g}_n(\mathbf{X}) - m_n(\mathbf{X})). \quad (20)$$

491 This expression can also be arrived at by first evaluating the posterior predictive mean of $g_n(\mathbf{x}^*)$, and
492 then differentiating with respect to x_d^* .

493 We use Equation 20 to define a closed-form gradient $\nabla_{\mathbf{x}^*} \hat{\gamma}_n$ at test point \mathbf{x}^* via

$$\nabla_{\mathbf{x}^*} \hat{\gamma}_n = \left[\frac{\partial \hat{\gamma}_n(\mathbf{x}^*)}{\partial x_1^*}, \dots, \frac{\partial \hat{\gamma}_n(\mathbf{x}^*)}{\partial x_D^*} \right]^\top. \quad (21)$$

494 Finally, Equation 21 is used in the projected gradient descent algorithm for optimising ensemble
495 stimuli (Algorithm 1).

496 6.7 Further details on simulations and "synthetic" optogenetics experiments

497 Simulations consisted of both ORF mapping and stimulus optimisation phases. ORF mapping
498 required probing responses to stimulation at a range of laser powers and stimulus locations. We

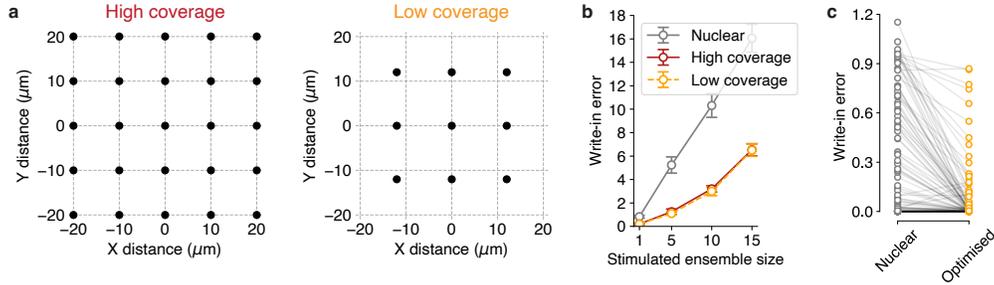


Figure S1: Effect of reducing the number of points at which each ORF is probed. (a) In the high coverage case (left), each ORF is probed by stimulating at a 5×5 grid of points near the soma (grid points separated by $10 \mu\text{m}$), at three different laser powers. In the low coverage case (right), this reduces to stimulating at just a 3×3 grid (grid points separated by $12 \mu\text{m}$) at three powers. However, as the density of opsin-expressing neurons increases, ORFs are probed at high density even in the low coverage case as the grids from different neurons increasingly overlap. (b) Minimal performance difference between the high and low coverage cases in simulations with 50 neurons. (c) Reduction in optical write-in error using cell-attached recordings as in [Figure 4](#), but with low coverage. Reduction in average write-in error, 74% (c.f. 85% with high coverage, [Figure 4c](#)).

499 defined a grid of stimulation points surrounding each neuron. In the spatial dimensions, the grid
500 ranged from $-20 \mu\text{m}$ to $20 \mu\text{m}$ relative to the centroid of the neuron in steps of $10 \mu\text{m}$, and powers
501 ranged from 30 mW to 70 mW in steps of 20 mW. The complete grid was thus given by the Cartesian
502 product $\{-20, -10, 0, 10, 20\} \times \{-20, -10, 0, 10, 20\} \times \{30, 50, 70\}$. For opsin-expressing neurons
503 that were spaced far apart, this coarse-resolution grid was sufficient because risk of OTS was low,
504 and therefore ORF mapping was not needed at high detail. On the other hand, as the density of
505 opsin-expressing neurons increased, the grids surrounding each neuron increasingly overlapped with
506 each other, resulting in much denser sampling of the ORFs.

507 For the synthetic optogenetics experiments (based on the cell-attached recordings), we used the same
508 spatial grid spacing but used laser powers of 10, 25, and 40 mW to match the range of powers used
509 in the underlying slice experiment, though note that the slice experiment had a denser spacing than
510 our chosen 15 mW (see example loose-patch recordings below), which we chose to reduce the ORF
511 mapping time. For the optogenetics experiments involving three spatial dimensions, we extended
512 the grid sampling to include depths of $-60 \mu\text{m}$ to $60 \mu\text{m}$ in steps of $30 \mu\text{m}$. We also explored the
513 effect of reducing the number of probed grid points to further reduce the time spent mapping ORFs,
514 and found that Bayesian target optimisation maintained high performance when probing with a 3×3
515 spatial grid of $\{-12, 0, 12\} \times \{-12, 0, 12\}$ ([Figure S1](#)).

516 We selected the parameters of the GP covariance kernel using 5-fold cross-validation on a separate set
517 of recordings that were made on the same set of four cells, ensuring the hyperparameter selection was
518 using out-of-sample data. Cross-validation was performed using a grid search over a set of possible
519 hyperparameters: the possible radial lengthscales were 2, 4, 8, 16, the power lengthscales were 2,
520 4, 8, 16, and the amplitudes were 1, 2, 4, 8, 16. For each hyperparameter combination θ and for
521 each cell, we used Newton's method to fit the GP-Bernoulli model to 80% of the trials in the loose-
522 patch data, yielding an ORF estimate \hat{g}_θ . On the remaining 20% of the trials (denoted as $\mathcal{T}_{\text{held-out}}$),
523 we evaluated the log-likelihood, $\sum_{t \in \mathcal{T}_{\text{held-out}}} \{y_t \ln(\sigma(\hat{g}_\theta(\mathbf{x}_t))) + (1 - y_t) \ln(1 - \sigma(\hat{g}_\theta(\mathbf{x}_t)))\}$. We
524 averaged the log-likelihood across all five folds and across all four cells, and chose the hyperparameter
525 combination θ that yielded the largest average log-likelihood, resulting in a radial lengthscale of 8, a
526 power lengthscale of 16, and a kernel amplitude of 8.

527 The GP parameters for generating the simulations in [Figure 3](#), inferring the resulting ORFs, and
528 generating synthetic optogenetics experiments with two and three spatial dimensions are given in
529 [Table S1](#). For reference, a typical ORF mean function is given in [Figure S2](#).

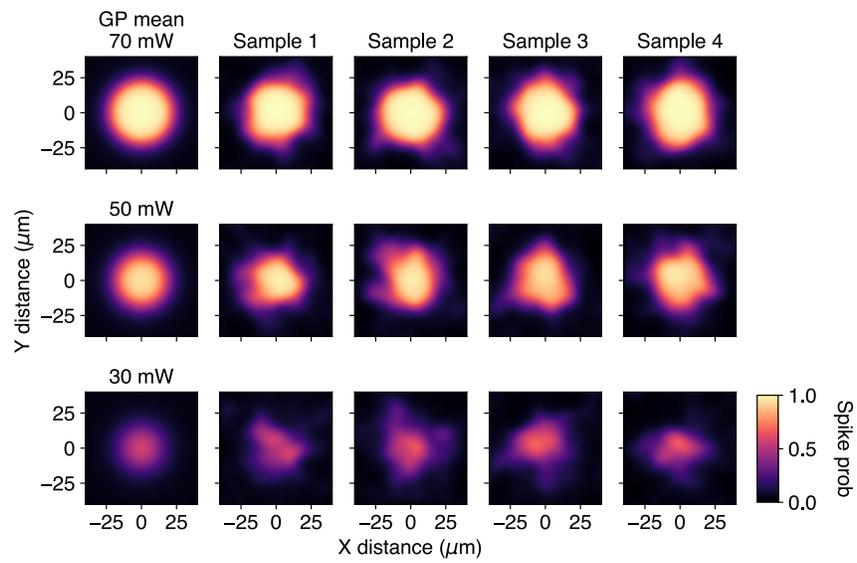


Figure S2: Example mean function (shown at three powers) used for simulations (left column). Also shown are four samples from the ORF prior corresponding to this mean function (right four columns). Parameters given in [Table S1](#).

Parameter	Symbol	Value
Simulations (data generation)		
Mean function excitability	ρ	0.125
Mean function width	σ_m^2	$3 \times 10^2 \mu\text{m}$
Spike threshold	θ	3.5
Kernel radial lengthscale	λ_s	$8 \mu\text{m}$
Kernel power lengthscale	λ_I	20 mW
Kernel amplitude	α^2	0.2
Kernel marginal variance	σ_d^2	10^{-5}
Simultaneously stimulated neurons during ORF mapping	J	10
Simulations (ORF inference)		
Mean function excitability	ρ	0.125
Mean function width	σ_m^2	$3 \times 10^2 \mu\text{m}$
Kernel radial lengthscale	λ_s	$5 \mu\text{m}$
Kernel power lengthscale	λ_I	16 mW
Kernel amplitude	α^2	1
Kernel marginal variance	σ_d^2	10^{-5}
Learning rate for spike thresholds ($\{\theta_n\}_{n=1}^N$)	—	5
Number of random initialisations	—	5
Synthetic optogenetics experiments (two spatial dimensions)		
Mean function excitability	ρ	0.175
Mean function width	σ_m^2	$3 \times 10^2 \mu\text{m}$
Kernel radial lengthscale	λ_s	$8 \mu\text{m}$
Kernel power lengthscale	λ_I	16 mW
Kernel amplitude	α^2	8
Kernel marginal variance	σ_d^2	10^{-5}
Learning rate for spike thresholds ($\{\theta_n\}_{n=1}^N$)	—	5
Number of random initialisations	—	5
Synthetic optogenetics experiments (three spatial dimensions)		
Mean function excitability	ρ	0.175
Mean function width (x/y dimensions)	σ_m^2	$3 \times 10^2 \mu\text{m}$
Mean function width (z dimension)	—	$3 \times 10^3 \mu\text{m}$
Kernel radial lengthscale (x/y dimensions)	λ_s	$8 \mu\text{m}$
Kernel axial lengthscale (z dimension)	λ_z	$32 \mu\text{m}$
Kernel power lengthscale	λ_I	16 mW
Kernel amplitude	α^2	8
Kernel marginal variance	σ_d^2	10^{-5}
Learning rate for spike thresholds ($\{\theta_n\}_{n=1}^N$)	—	5
Number of random initialisations	—	5

Table S1: Parameters used for simulations and generating synthetic optogenetics experiments.

530 **6.8 Additional examples of optogenetic receptive fields from cell-attached recordings**

531 Figures S3 to S6 show examples of four ORFs that have been comprehensively mapped using two-
 532 photon optogenetic stimulation and cell-attached recordings of evoked spikes. Note the unpredictable
 533 differences in ORF shape across laser powers and depths, motivating a nonparametric approach.

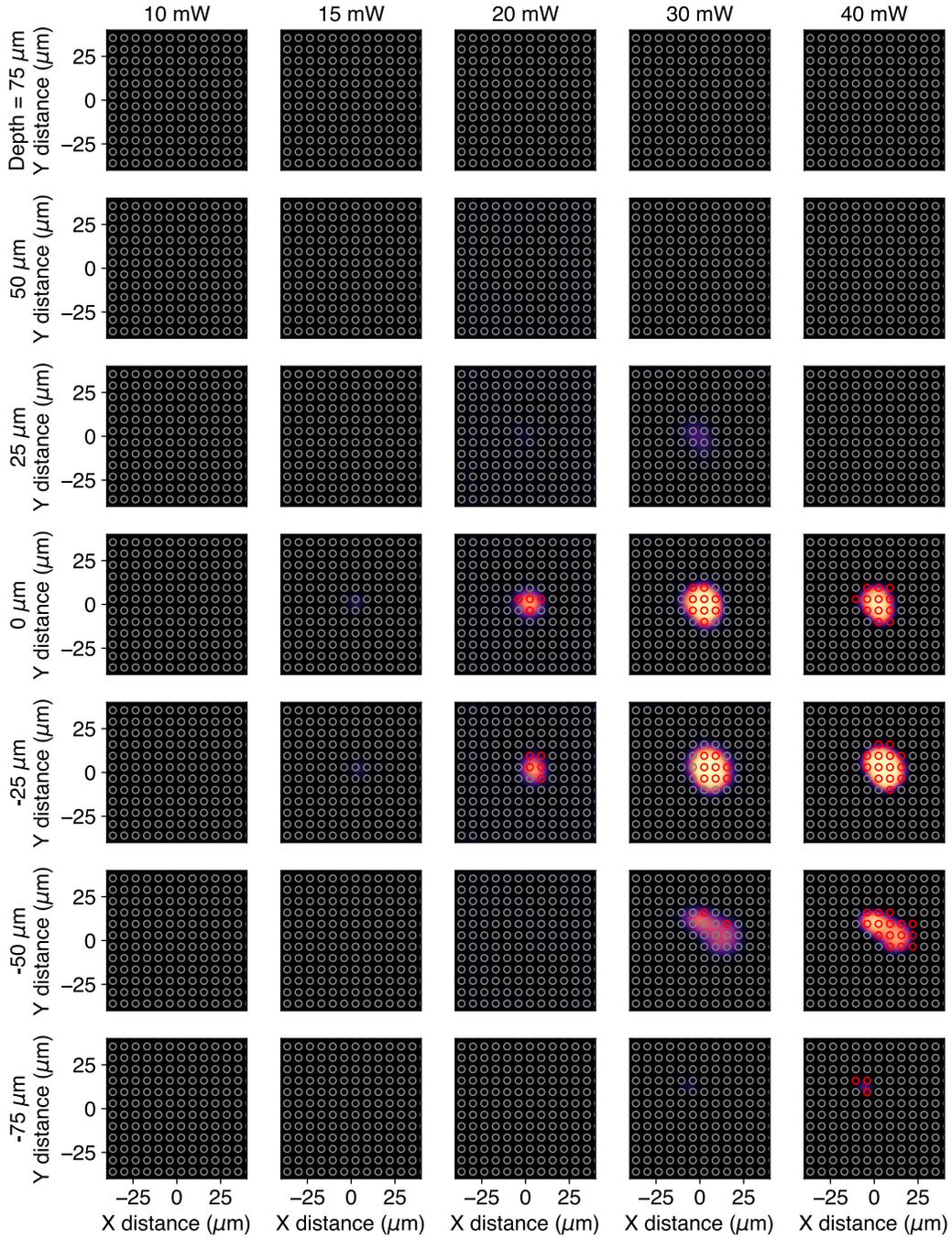


Figure S3: Loose-patch recording and inferred ORF (experiment 1/4).

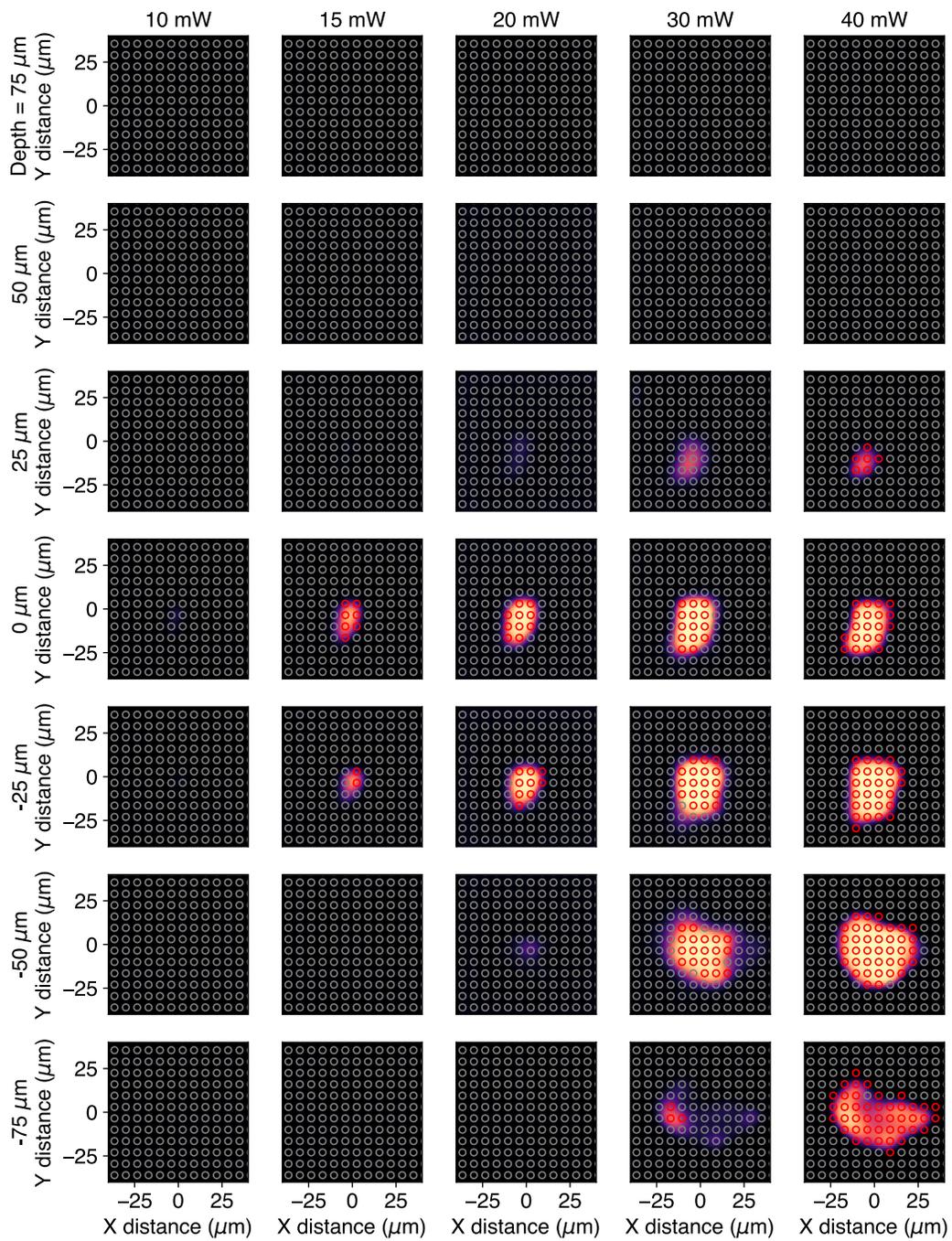


Figure S4: Loose-patch recording and inferred ORF (experiment 2/4)

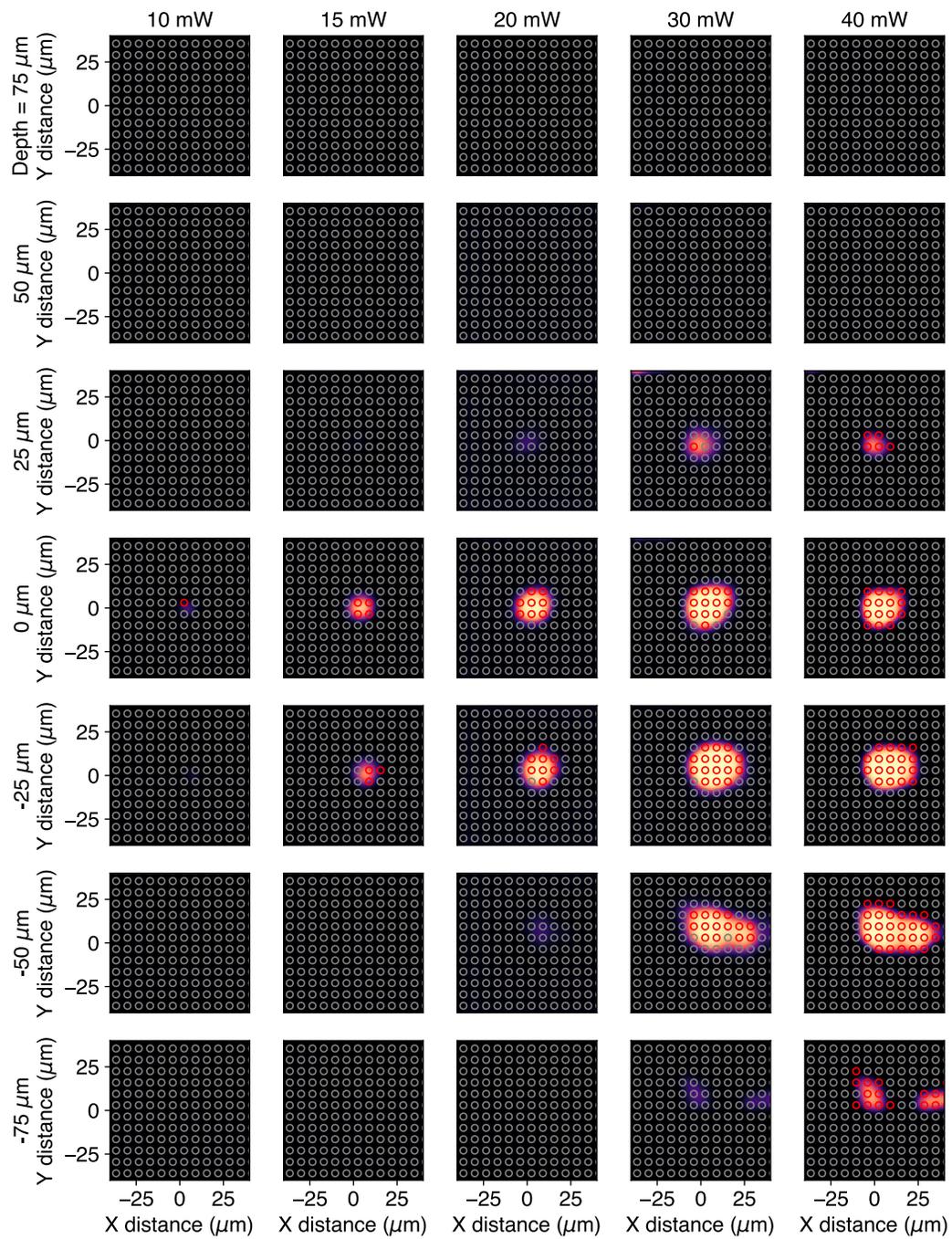


Figure S5: Loose-patch recording and inferred ORF (experiment 3/4)

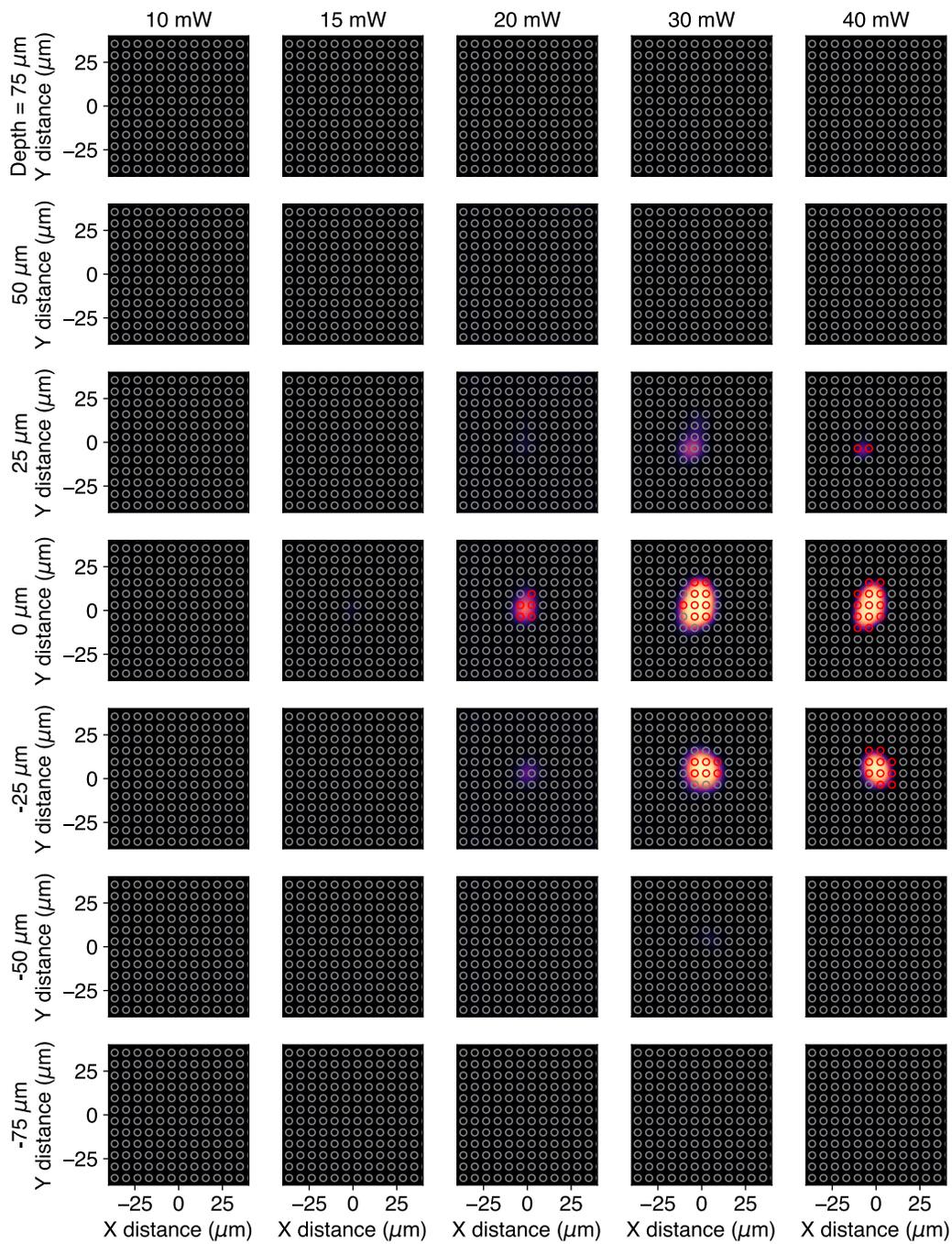


Figure S6: Loose-patch recording and inferred ORF (experiment 4/4)