
Automated discovery of large-scale, noise-robust experimental designs in super-resolution microscopy

Carla Rodríguez

Max Planck Institute for the Science of Light, Erlangen, Germany
carla.rodriguez@mpl.mpg.de

Sören Arlt

Max Planck Institute for the Science of Light, Erlangen, Germany
soeren.arlt@mpl.mpg.de

Leonhard Möckl

Max Planck Institute for the Science of Light, Erlangen, Germany
Friedrich-Alexander-University Erlangen-Nuremberg, Faculty of Sciences,
Department of Physics, Erlangen, Germany
Friedrich-Alexander-University Erlangen-Nuremberg, Faculty of Medicine
1/CITABLE, Erlangen, Germany
Deutsches Zentrum Immuntherapie (DZI), Erlangen, Germany
leonhard.moeckl@mpl.mpg.de

Mario Krenn

Max Planck Institute for the Science of Light, Erlangen, Germany
mario.krenn@mpl.mpg.de

Abstract

The discovery of super-resolution techniques, which circumvent the classical diffraction limit of light, represent a leap in optical microscopy. Yet, the vast space of all possible experimental configurations suggests that powerful techniques remain undiscovered. We demonstrate the automated discovery of SR microscopy techniques using XLUMINA, an open-source JAX-based computational framework which demonstrates a speed-up of 4 orders of magnitude compared to well-established numerical optimization methods. We implement a highly-efficient optimization scheme that incorporates random noise sampling at each iteration to ensure robustness, leading to the discovery of a novel, noise-resilient experimental blueprint featuring sub-diffraction imaging capabilities. This work advances AI-driven discovery in optics and microscopy, emphasizing both high-performance and robustness.

1 Introduction

The space of all possible experimental optical configurations is enormous. It contains all possible discrete arrangements of optical elements (e.g., lenses, beam splitters, phase shifters, lasers...) and their tunable parameters (such as lenses' focal lengths, beam splitter ratios, phases or laser power), which leads to additional high-dimensional continuous parameter space for each of the mentioned discrete possibilities. This vast search space contains all experimental designs possible, including

those with exceptional properties. AI-based exploration techniques could provide enormous benefit by exploring the space in a fast, unbiased way [1, 2].

Optical microscopy has impact on diverse fields, ranging from material sciences all the way to medicine [3–6]. It is currently most widely used in biological sciences, where precise labeling of imaging targets enables fluorescence microscopy with exquisite sensitivity and specificity [7, 8]. In the past two decades, several breakthroughs have broadened the scope of optical microscopy in this area even further. Among them, through the ingenuity and creativity of human researchers, the discovery of super-resolution (SR) methods [9–15], which circumvent the classical diffraction limit of light, stand out in particular. These SR techniques, highlighted by the 2014 Chemistry Nobel Prize [16], have considerable impact in biology [17–19], chemistry [20] and material sciences [21]. In this work, we demonstrate the automated discovery of SR microscopy techniques using XLUMINA. Crucially, the motivation of our work goes far beyond small-scale optimization of already known optical techniques. Rather, this work sets out to discover novel, experimentally viable concepts for advanced optical microscopy that are at-present entirely untapped.

2 Software performance

XLUMINA is a highly-efficient open-source framework developed using JAX [22], which allows for the simulation of classical optics hardware configurations and enables the optimization and automated discovery of new setup designs. XLUMINA is equipped with an optics simulator which contains a diverse set of optical manipulation, interaction and measurement technologies. To include the automated discovery feature, XLUMINA’s optical simulator and optimizer are tied together by the loss function. The automated discovery tool is designed to explore the vast parameter space encompassing all possible optical designs. We adopt a gradient-based strategy, where the experimental setup’s parameters are adjusted iteratively in the steepest descent direction. We evaluate the time it takes for numerical and analytical (auto-differentiation) methods to compute one gradient evaluation times over different resolutions and devices. We compare XLUMINA’s performance with the optics framework *Diffractio* [23], which is a high-quality, open-source NumPy-based Python module for optics simulation with an active developer community, and is employed in numerous studies in optics and physics in general. The acquired results are depicted in Fig. 1a. Clearly, autodiff consistently outperforms numerical methods on the gradient evaluation time by up to 5 orders of magnitude on GPU and 4 orders on CPU. Given that certain optical elements may operate at resolutions as high as 2048×2048 pixels, the resulting search space can easily expand to around 8.4 million parameters. This makes the use of autodiff within GPU-accelerated frameworks more appropriate for efficient experimentation. Overall, the computational performance of XLUMINA highlights its suitability for running complex simulations and optimizations with a high level of efficiency. More details on the software and its performance are discussed in the Appendix A.

3 Large-scale discovery framework

Configuring both the optical topology and the optical parameters is essentially a hybrid discrete-continuous optimization problem, which is extremely difficult computationally. We define a quasi-universal computational *ansatz*, illustrated in Fig. 1b, which translates this hybrid scheme into a purely continuous optimization framework that can be solved with efficient gradient-based methods. Essentially, setting different continuous parameters leads to different optical setup topologies: e.g., adjusting beam splitter ratios the optimizer can “turn off” the optical paths. Remarkably, for a very discrete approach of available parameters, the number of possible discrete arrangements within this framework scales up to $\sim 10^{20}$. Details on the *ansatz* can be found in the Appendix section B.

Loss function: The loss function, \mathcal{L} , is calculated as the inverse of the density of the total detected intensity over a certain threshold, I_ε . Thus, minimizing \mathcal{L} aims to maximize the generation of small, high intensity beams. In particular,

$$\mathcal{L} = \frac{1}{\text{Density}} = \frac{\text{Area}}{I_\varepsilon} \quad (1)$$

where I_ε is the sum of pixel intensity values greater than the threshold value $\varepsilon \cdot i_{\max}$, where $0 \leq \varepsilon \leq 1$ and i_{\max} corresponds to the maximum detected intensity. The Area corresponds to the total number

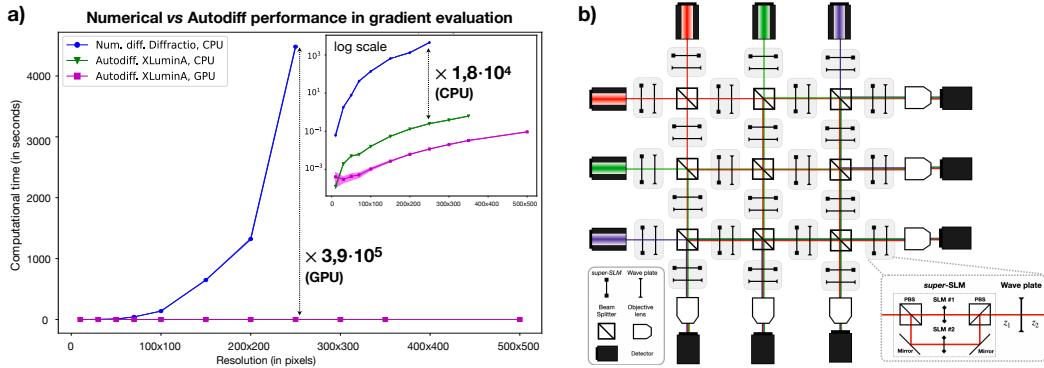


Figure 1: Performance of XLUMINA and large-scale discovery framework. (a) Average time (in seconds) over 5 runs for single gradient evaluation using numerical differentiation with *Diffractio*'s optical simulator (blue dots) and autodiff methods (green triangles for CPU and magenta squares for GPU) with XLUMINA's optical simulator for different resolutions. The use of XLUMINA with autodiff methods improves the gradient evaluation and convergence time by a factor of $\times 3.9 \cdot 10^5$ and $\times 2.1 \cdot 10^4$ in the GPU and CPU, respectively, for a resolution of 250×250 pixels. Shaded regions correspond to standard deviation values. The numerical and autodiff methods are computed using BFGS and Adam optimizers, respectively. (b) Quasi-universal computational ansatz. The setup's complexity and size can be arbitrarily extended by incorporating additional connections.

of camera pixels fulfilling the same condition. The loss function \mathcal{L} is common to all the optical setups henceforth described. Importantly, light gets detected across various devices. Thus, we compute the loss function at each detector and the parameter update is driven by the device demonstrating the minimum loss value. This selection is performed in a fully differentiable manner. Details on the derivation of the loss function and camera selection are provided in the Appendix section B.1.

4 Results

In this section we showcase the optical designs generated by XLUMINA. The goal is to discover both the setup topology and the patterns to imprint onto the light beams using the available optimizable optical parameters (i.e., SLMs, distances, beam splitter ratios and wave plate's angles). To ensure the discovered optical designs are robust, we implement a novel noise-robust optimization procedure that emulates real-world experimental imperfections. Our approach involves optimizing, simultaneously, three multiple instances of the optical setup, each subjected to randomly sampled noise in its physical parameters. At each step, we compute the mean loss across these noisy variants and optimize based on the average performance. This iterative process, where the noise is resampled at each step, allows us to explore optical solutions that are inherently resilient to noise and maintain high performance even under non-ideal conditions. The detailed description of the noise implementation are provided in the Appendix section C. The showcased solutions are the result from running multiple optimizations.

Rediscovery through exploration: As a benchmark we aim to rediscover the generation of an ultra-sharp focus, a feature that breaks the diffraction limit in the longitudinal direction as demonstrated by Dorn, Quabis and Leuchs in Ref. [24]. We initialize XLUMINA with the virtual optical table in Supplementary Fig. 3. The loss function corresponds to equation (1) considering the measured intensity as the field's longitudinal component, $|E_z|^2$. The discovered topology and SLMs patterns are depicted in Figs. 2a and 2b, respectively. Surprisingly, XLUMINA found an alternative way to imprint a phase singularity onto the beam and produce pronounced longitudinal components on the focal plane. The longitudinal intensity profiles of the discovered solution and the reference experiment are depicted in Fig. 2c. Remarkably, the identified solution demonstrates a spot size close to the reference and does not feature side lobes, which can limit practical imaging techniques. Details on the optimization process are provided in the Appendix section D.

Discovery of a new experimental blueprint: Finally, we demonstrate the capability of XLUMINA for genuine discovery. For this purpose, we simulate the effect of stimulated emission-depletion, one of the fundamental concepts of STED microscopy [9, 25]. The key idea of this technique is the use of two diffraction-limited laser beams, one probe to activate (excite) the light emitters of the sample and one, doughnut-shaped beam to deactivate its excitation in a controlled way (depletion). Thus, the ultimately detected light is that of the emitters laying in the central region of the doughnut-shaped beam. This effectively reduces the area of normal fluorescence, which leads to super-resolution imaging. Without having to rely on time-dependent processes, such as the energy level relaxation times of the excited emitters, we perform a nonlinear modulation of the focused light based on the Beer-Lambert law [26], commonly used to describe the optical attenuation in light-matter interaction. The details of our fluorescence model are provided in the Appendix section E.

We initialize the system in the virtual setup depicted in Supplementary Fig. 3. The details of the optimization are provided in the Appendix section E.1. The loss function corresponds to equation (1) considering the total intensity of the effective beam resulting from the STED process, $|E_x|^2 + |E_y|^2 + |E_z|^2$. The discovered topology and identified SLMs patterns are depicted in Figs. 2d and 2e, respectively. The detected intensity topologies reveal the system generates a doughnut-shaped and a Gaussian-like beams. We compute the vertical cross-section of the focused intensity patterns for both beams and the resulting effective beam (green, orange and dotted blue lines in Fig. 2f, respectively). The horizontal cross-section exhibits analogous features. We further compare the effective beam intensity with the simulated STED reference [9] and the the discovered Gaussian-like beam with the simulated sharp focus reference [24]. Strikingly, the discovered solution exploits the underlying physical concepts of two aforementioned optical systems. In one hand, it generates a doughnut-shaped “depletion” beam as demonstrated in Ref. [9]. On the other hand, it generates a Gaussian-like “excitation” signal with a sharper focus, achieving smaller effective intensity spots resulting from the STED process. The discovered solution showcases an effective beam profile which is sharper than the simulated STED reference. This occurs due to the enhanced sharpening of the longitudinal component of the excitation beam, which demonstrates similar profile as the simulated sharp focus reference [24]. To the best of our knowledge, this technique has never been discussed in the scientific literature before. Regardless of its physical realizability, this solution demonstrates the ability of XLUMINA to uncover interesting solutions within highly complex systems.

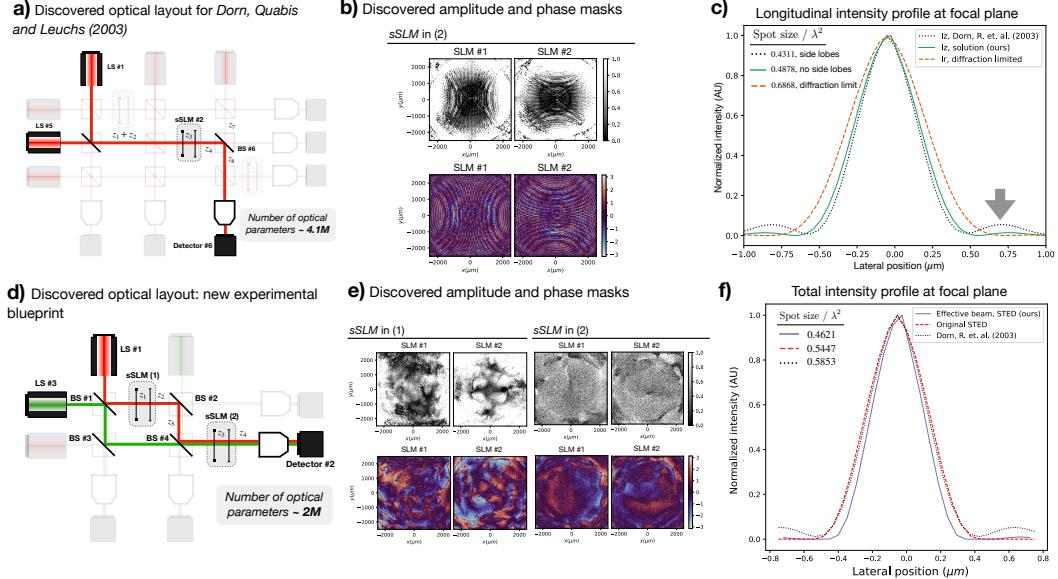


Figure 2: Rediscovery of robust optical solutions (a-c) and discovery of a new experimental blueprint (d-f) within a highly parameterized, noise-robust optical setup. The parameter space and the values of the identified optical parameters are specified in the Appendix section E.1. (a) Discovered optical topology for Dorn, Quabis and Leuchs (2003). (b) Discovered amplitude and phase patterns for *super-SLMs* (*sSLM*) in (2). (c) Normalized longitudinal intensity profile for Dorn, Quabis, and Leuchs (2003) and the identified solution (black dotted, and green lines, respectively) and radial intensity profile of the diffraction-limited linearly polarized beam (orange dotted line). Lateral position indicates lateral distance from the optical axis. The spot size is computed as $\phi = (\pi/4)\text{FWHM}_x\text{FWHM}_y$, where FWHM denotes for Full Width Half Maximum. The discovered approach breaks the diffraction limit with a spot size close to the reference. Remarkably, it does not feature side lobes (indicated with a gray arrow), which can limit practical imaging techniques. (d) Discovered optical topology for the new experimental blueprint. The minimum value of the loss is demonstrated in detector #2. (e) Discovered amplitude and phase masks corresponding to the *sSLM* in (1) and (2). (f) Total intensity ($|E_x|^2 + |E_y|^2 + |E_z|^2$) horizontal cross-section of the detected light beams of 650 nm (orange), 530 nm (green), and effective beam emulating stimulated emission (dashed blue). (d) Horizontal cross-section of the normalized total intensity of the effective beam from the discovered solution (blue), the simulated STED reference (dashed red), and the simulated reference (dotted black) using 560 nm wavelength. The discovered solution outperforms both simulated references for STED microscopy and the sharp focus from Dorn, Quabis and Leuchs (2003), demonstrating a spot size of 15% smaller than both references.

References

- [1] H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, S. Liu, P. Van Katwyk, A. Deac, *et al.*, “Scientific discovery in the age of artificial intelligence,” *Nature*, vol. 620, no. 7972, pp. 47–60, 2023.
- [2] M. Krenn, R. Pollice, S. Y. Guo, M. Aldeghi, A. Cervera-Lierta, P. Friederich, G. dos Passos Gomes, F. Häse, A. Jinich, A. Nigam, *et al.*, “On scientific understanding with artificial intelligence,” *Nature Reviews Physics*, vol. 4, no. 12, pp. 761–769, 2022.
- [3] A. M. Reigoto, S. A. Andrade, M. C. R. R. Seixas, M. L. Costa, and C. Mermelstein, “A comparative study on the use of microscopy in pharmacology and cell biology research,” *PLOS ONE*, vol. 16, no. 1, pp. 1–13, 2021.
- [4] S. Weisenburger and V. Sandoghdar, “Light microscopy: an ongoing contemporary revolution,” *Contemporary Physics*, vol. 56, no. 2, pp. 123–143, 2015.
- [5] A. Bullen, “Microscopic imaging techniques for drug discovery,” *Nature Reviews Drug Discovery*, vol. 7, no. 1, pp. 54–67, 2008.

- [6] P. Antony, C. Trefois, A. Stojanovic, A. Baumuratov, and K. Kozak, “Light microscopy applications in systems biology: opportunities and challenges,” *Cell Communication and Signaling*, vol. 11, no. 1, 2013.
- [7] J. B. Grimm and L. D. Lavis, “Caveat fluorophore: an insiders’ guide to small-molecule fluorescent labels,” *Nature Methods*, vol. 19, no. 2, 2022.
- [8] D. K. M. and A. E. Palmer, “Advances in fluorescence labeling strategies for dynamic cellular imaging,” *Nature Chemical Biology*, vol. 10, no. 7, 2014.
- [9] S. W. Hell and J. Wichmann, “Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy,” *Optics Letters*, vol. 19, no. 11, pp. 780–782, 1994.
- [10] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, “Imaging intracellular fluorescent proteins at nanometer resolution,” *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.
- [11] S. T. Hess, T. P. Girirajan, and M. D. Mason, “Ultra-high resolution imaging by fluorescence photoactivation localization microscopy,” *Biophysical Journal*, vol. 91, no. 11, pp. 4258–4272, 2006.
- [12] M. Rust, M. Bates, and X. Zhuang, “Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm),” *Nature Methods*, vol. 3, p. 793–796, 2006.
- [13] S. van de Linde, A. Löschberger, T. Klein, M. Heidbreder, S. Wolter, M. Heilemann, and M. Sauer, “Direct stochastic optical reconstruction microscopy with standard fluorescent probes,” *Nature Protocols*, vol. 6, p. 991–1009, 2011.
- [14] M. G. L. Gustafsson, “Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 37, pp. 13081–13086, 2005.
- [15] F. Balzarotti, Y. Eilers, K. C. Gwosch, A. H. Gynnå, V. Westphal, F. D. Stefani, J. Elf, and S. W. Hell, “Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes,” *Science*, vol. 355, no. 6325, pp. 606–612, 2017.
- [16] L. Möckl, D. C. Lamb, and C. Bräuchle, “Super-resolved fluorescence microscopy: Nobel prize in chemistry 2014 for eric betzig, stefan hell, and william e. moerner,” *Angewandte Chemie International Edition*, vol. 53, no. 51, pp. 13972–13977, 2014.
- [17] L. Möckl, K. Pedram, A. R. Roy, V. Krishnan, A.-K. Gustavsson, O. Dorigo, C. R. Bertozzi, and W. Moerner, “Quantitative super-resolution microscopy of the mammalian glycocalyx,” *Developmental Cell*, vol. 50, no. 1, pp. 57–72.e6, 2019.
- [18] K. Xu, G. Zhong, and X. Zhuang, “Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons,” *Science*, vol. 339, no. 6118, pp. 452–456, 2013.
- [19] A. Yildiz, J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin, “Myosin v walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization,” *Science*, vol. 300, no. 5628, pp. 2061–2065, 2003.
- [20] Y. Zhang, J. M. Lucas, P. Song, B. Beberwyck, Q. Fu, W. Xu, and A. P. Alivisatos, “Superresolution fluorescence mapping of single-nanoparticle catalysts reveals spatiotemporal variations in surface reactivity,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 29, pp. 8959–8964, 2015.
- [21] P. Müller, R. Müller, L. Hammer, C. Barner-Kowollik, M. Wegener, and E. Blasco, “Sted-inspired laser lithography based on photoswitchable spirothiopyran moieties,” *Chemistry of Materials*, vol. 31, no. 6, pp. 1966–1972, 2019.
- [22] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, “JAX: composable transformations of Python+NumPy programs,” 2018.
- [23] L. M. S. Brea, “Diffractio, python module for diffraction and interference optics,” 2019.
- [24] R. Dorn, S. Quabis, and G. Leuchs, “Sharper focus for a radially polarized light beam,” *Physical Review Letters*, vol. 91, p. 233901, 2003.
- [25] M. Hofmann, C. Eggeling, S. Jakobs, and S. W. Hell, “Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 49, pp. 17565–17569, 2005.

- [26] T. G. Mayerhöfer, S. Pahlow, and J. Popp, “The bouguer-beer-lambert law: Shining light on the obscure,” *Chemphyschem*, vol. 21, pp. 2029 – 2046, 2020.
- [27] F. Shen and A. Wang, “Fast-fourier-transform based numerical integration method for the rayleigh-sommerfeld diffraction formula,” *Applied Optics*, vol. 45, no. 6, pp. 1102–1110, 2006.
- [28] H. Ye, C.-W. Qiu, K. Huang, J. Teng, B. Luk’Yanchuk, and S. Yeo, “Creation of a longitudinally polarized subwavelength hotspot with an ultra-thin planar lens: Vectorial rayleigh-sommerfeld method,” *Laser Physics Letters*, vol. 10, no. 6, 2013.
- [29] Y. Hu, Z. Wang, X. Wang, *et al.*, “Efficient full-path optical calculation of scalar and vector diffraction using the bluestein method,” *Light: Science & Applications*, vol. 9, p. 119, 2020.
- [30] J. Li, Z. Fan, and Y. Fu, “FFT calculation for Fresnel diffraction and energy conservation criterion of sampling quality,” in *Lasers in Material Processing and Manufacturing*, vol. 4915, pp. 180 – 186, SPIE, 2002.

A Software workflow and performance

In this section we provide the detailed description of XLUMINA’s simulation features and performance. The simulator enables, among many other features, to define light sources (of any wavelength and power), phase masks (i.e., spatial light modulators, SLMs), polarizers, variable retarders (e.g., wave plates, WPs), diffraction gratings, and high numerical aperture (NA) lenses to replicate strong focusing conditions. Light propagation and diffraction is simulated by two methods, each available for both scalar and vectorial regimes: the fast-Fourier-transform (FFT) based numerical integration of the Rayleigh-Sommerfeld (RS) diffraction integral [27, 28] and the Chirped z-transform (CZT) [29]. The CZT is an accelerated version of the RS algorithm, which allows for arbitrary selection and sampling of the region of interest. These algorithms are based on the FFT and require a reasonable sampling for the calculation to be accurate [30]. In our simulations we consider light sources emitting Gaussian beams of 1.2 mm beam waist. To avoid possible boundary-generated artifacts during the simulation, we define these beams in larger computational spaces of 4 mm or 5 mm. Thus, the pixel resolutions often span 1024×1024 , or 2048×2048 .

Some functionalities of XLUMINA’s optics simulator (e.g., optical propagation algorithms, planar lens or amplitude masks) are inspired in an open-source NumPy-based Python module for diffraction and interferometry simulation, *Diffractio* [23], although we have rewritten and modified these approaches to combine them with JAX just-in-time (*jit*) functionality. In essence, *jit* compilation optimizes sequences of operations together and runs them at once. For this purpose, the first run of a jitted function builds an abstract representation of the sequence of operations specified by the function. This representation encodes the shape and the *dtype* of the arrays - but is agnostic to the values of such arrays. If the input shapes and *dtypes* are not modified, the abstract structure of the function can be then re-used for subsequent runs, without re-compilation, which allows to execute the subsequent calls faster. However, if the input shape or *dtype* is modified, the function automatically gets re-compiled. This will cause an extra overhead time due to the extraction of a new abstract structure of the function for the new shapes/*dtypes*. On top of that, we developed completely new functions (e.g., beam splitters, WPs or propagation through high NA objective lens with CZT methods, to name a few) which significantly expand the software capabilities. The most important hardware addition on the optical simulator are the SLMs, each pixel of which possesses an independent (and variable) phase value. They serve as a universal approximation for phase masks, including lenses, and offer a computational advantage: given a specific pixel resolution, they allow for unrestricted phase design selection. Such flexibility is crucial during the parameter space exploration, as it allows the software to autonomously probe all potential solutions. In addition, we defined under the name of *super-SLM* (*sSLM*) a hardware-box-type which consists of two SLMs, each one independently imprinting an amplitude and phase mask on the horizontal and vertical polarization components of the field.

To evaluate the performance of numerical and auto-differentiation methods we chose to use BFGS (from SciPy’s Python library) and Adam (included in the JAX library) as optimizers. Further comparison including SGD (Stochastic-Gradient-Descent), AdaGrad (Adaptive Gradient) and AdamW (Adam with weight decay) is presented in Supplementary Fig. 1.

As the optical system, we set-up a Gaussian beam propagating over a distance z and interacting with a phase mask. The objective function is the mean squared error between the detected light and the ground truth, characterized by a Gaussian beam with a spiral phase imprinted on its wavefront. We initialize the system with an arbitrary phase mask configuration. We first evaluate the computational time for a single gradient evaluation for numerical and autodiff methods across different computational window sizes (from 10×10 up to 500×500 pixels) and devices (CPU and GPU). We keep the default settings for BFGS. For Adam, the step size is set to 0.1. The optimization process is terminated if there is no improvement in the loss value (meaning it has not decreased below the best value recorded), over 50 consecutive iteration steps. For each resolution window, we collect the convergence time of both optimizers and divide it by the total number of gradient evaluations for BFGS and the total number of steps for Adam. The acquired gradient evaluation times correspond to the mean value over 5 runs. Obtained results are depicted in Figs. 1a. It is clear how autodiff outperforms numerical methods by up to 4 orders of magnitude on CPU and 5 orders of magnitude when running in the GPU. The advantage over larger sizes is clear given that we run simulations with resolutions of 1024×1024 and 2048×2048 pixels.

We then conduct the evaluation of the convergence time for both methods. We keep the aforementioned settings for the optimizers. We initialize the systems 5 times and compute their mean value. The acquired results are depicted in Fig. 1b. On the CPU, numerical methods exhibit exponential scaling in convergence time, reaching about $4.5 \cdot 10^4$ seconds (roughly 12 hours) for 250×250 pixel resolution. In contrast, autodiff demonstrates superior efficiency, reducing it to roughly 53 seconds. GPU optimization performance is even more pronounced, reaching convergence in 0.24 seconds for 250×250 pixels, and 16 seconds for a resolution of 500×500 .

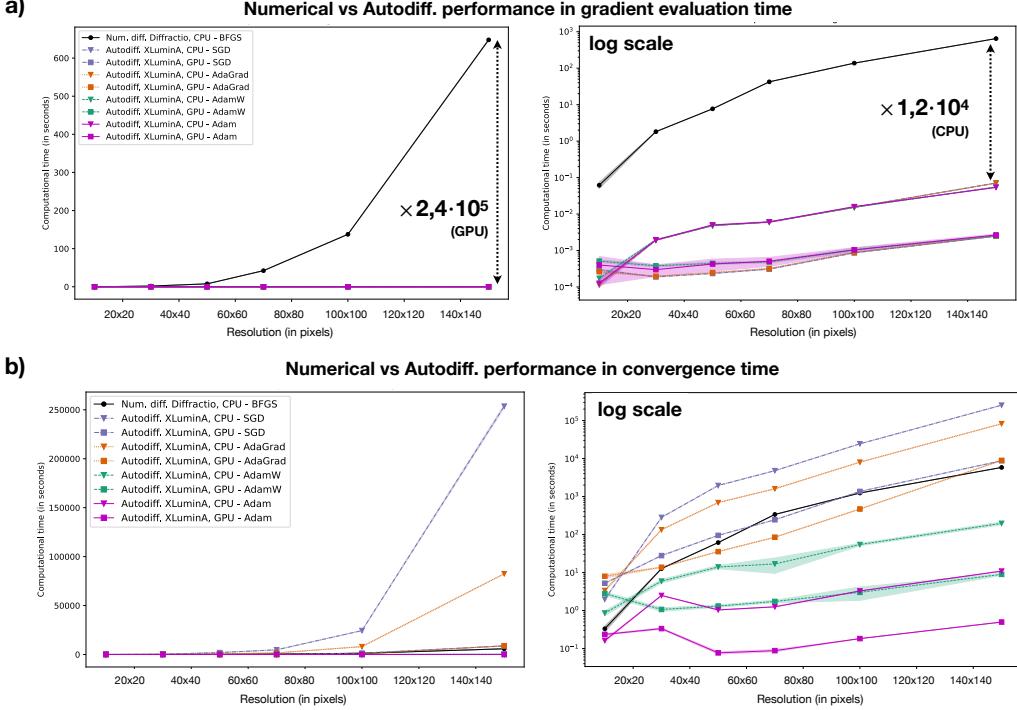


Figure 1: Performance of XLUMINA (auto-differentiation) compared to *Diffractio* (numerical methods) across different resolutions and optimizers in (a) single gradient evaluation and (b) convergence time. Data corresponds to the average time over 5 runs. Numerical differentiation is computed using *Diffractio*'s optical simulator and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimizer (black dots) and auto-differentiation (triangles for CPU and squares for GPU) on XLUMINA. The Stochastic-Gradient-Descent (SGD), Adaptive Gradient (AdaGrad), Adaptive moment estimation with weight decay (AdamW) and Adaptive moment estimation (Adam) correspond to blue (dash-dot line), orange (dotted line), green (dash line) and magenta (continuous line), respectively. Shaded regions correspond to standard deviation values. The learning rate is set to 0.1 and is common to all the optimizers. For AdamW, the weight decay is set to 10^{-4} . The stopping condition is common to all the frameworks: the optimization is terminated if there is no improvement in the loss value (i.e., it has not decreased below the best value recorded), over 500 consecutive iteration steps. This condition is checked every 100 steps. The use of XLUMINA with autodiff methods improves the gradient evaluation time by a factor of $\times 2.4 \cdot 10^5$ in the GPU and a factor of $\times 1.2 \cdot 10^4$ on the CPU for resolutions of 150×150 pixels. This behavior is common to all the tested optimizers (Adam, AdamW, SGD and AdaGrad). When evaluating the convergence time, the use of Autodiff methods on XLUMINA using the Adam and AdamW optimizers improve the performance with respect to numerical methods by a factor of $\times 1.1 \cdot 10^4$ and $\times 6.5 \cdot 10^2$ in the GPU, respectively, for a resolution of 150×150 pixels. The performance of Adam and AdamW in the CPU demonstrates factors of $\times 5.8 \cdot 10^2$ and $\times 2.9 \cdot 10^1$, respectively, for the same resolution. Remarkably, the use of *Diffractio* with numerical methods (BFGS) outperforms both AdaGrad and SGD in convergence time. In particular, numerical methods outperform AdaGrad by a factor of $\times 1.53$ in the GPU and $\times 14.14$ in the CPU, for a resolution of 150×150 pixels. Similar behavior is demonstrated for SGD: numerical methods outperform it by a factor of $\times 1.50$ in the GPU and $\times 43.50$ in the CPU, for a resolution of 150×150 pixels. Overall, the use of Autodiff methods (in particular, using Adam or AdamW) within GPU-accelerated frameworks is a more appropriate choice to conduct efficient optimization.

B Large-scale discovery framework

In this section we detail the methodology for the optimizations conducted using our quasi-universal computational *ansatz*, a purely continuous framework. We first discuss the enormous search space corresponding to large-scale optical setups. Afterwards, we provide the derivation of the loss function in equation (1).

The large-scale optical setup depicted in Fig. 1b consists of 6 light sources that emit linearly polarized Gaussian beams with different wavelengths (e.g., 625 nm, 530 nm and 470 nm). Through 82 vectorial propagation (vectorial Rayleigh-Sommerfeld, VRS), these beams interact with a total of 9 beam splitters, 24 *sSLMs* (i.e., 48 SLMs), 24 wave plates, and get ultimately detected by 6 high NA objective lenses focusing on light detectors.

We analyze the number of possible discrete arrangements within this general optical setup. For a very discrete approach of beam splitter ratios (either transmit, reflect or have light in both arms) and only allowing the SLMs and wave plates (WP) to be switched ON/OFF (i.e., displaying a constant zero phase and amplitude or adding zero retardance to the incoming light), the number of possible discrete layouts is of

$$N_{\text{Discrete layouts}} = 3^{9_{\text{BS}}} \cdot 2^{48_{\text{SLM}}} \cdot 2^{24_{\text{WP}}} = 2 \cdot 10^{20}. \quad (2)$$

All these are considering that the available beam splitter ratios are restricted to 3 values and the SLMs and wave plates to turn ON/OFF, respectively. In practice, beam splitter ratios and phase values are continuous variables and can take any value (from 0 to 1 and $-\pi$ to π , respectively) which increases even more the dimension of our search space.

In the following Table 1 we present a summary detailing the main properties of the optimizations conducted within our large-scale ansatz: the number of tunable elements, the dimension of the parameter space and the available number of topologies (for a very discrete approach).

Table 1: Outline of the main properties of the five digital experiments conducted within our large-scale ansatz. Displays the total number of tunable elements, the dimension of the parameter space and the available topologies (for a very discrete approach).

Experiment (Fig. #)	# tunable elements	Parameter space	# available topologies
Fig. 2a	26	~ 4 million	10^7
Fig. 2d	16	~ 2 million	10^3

B.1 Loss function

The loss function, \mathcal{L} , is inversely proportional to the total detected intensity density that is above a specified intensity threshold, I_ε . Thus, minimizing \mathcal{L} aims to maximize the generation of small, high intensity beams. In particular, it reads

$$\mathcal{L} = \frac{1}{\text{Density}} = \frac{\text{Area}}{I_\varepsilon}. \quad (3)$$

The total intensity I_ε above the threshold is computed as

$$I_\varepsilon = \sum_{k,l}^N i_\varepsilon(k, l), \quad (4)$$

where N is the total number of pixels in the camera's sensor and $i_\varepsilon(k, l)$ represents the intensity value at each pixel once the threshold condition is applied. This condition is defined as follows:

$$i_\varepsilon(k, l) = \begin{cases} i_{\text{det}}(k, l) & \text{if } i_{\text{det}}(k, l) > \varepsilon i_{\text{max}}, \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where $i_{\text{det}}(k, l)$ is the intensity value at the i -th row and j -th column in the detected 2D intensity pattern, $\varepsilon i_{\text{max}}$ (with $0 \leq \varepsilon \leq 1$) is the threshold value, with i_{max} being the maximum intensity value in the entire 2D detector array.

The Area is determined using a variation of the Heaviside function Θ applied to i_ε , quantifying the area where the intensity is above the threshold:

$$\text{Area} = \sum_{k,l}^N \Theta(i_\varepsilon(k, l)), \quad (6)$$

where N is the total number of pixels in the camera's sensor and $\Theta(i_\varepsilon(k, l))$ is defined as:

$$\Theta(i_\varepsilon(k, l)) = \begin{cases} 1 & \text{if } i_\varepsilon(k, l) > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

Therefore, the loss function can be read as follows:

$$\mathcal{L} = \frac{1}{\text{Density}} = \frac{\text{Area}}{I_\varepsilon} = \frac{\sum_{k,l}^N \Theta(i_\varepsilon(k, l))}{\sum_{k,l}^N i_\varepsilon(k, l)}. \quad (8)$$

Importantly, the camera pixel selection in equation (7) is a discrete operation. However, JAX offers some interesting capabilities due to its integrated autodiff framework. In particular, control flow operations in JAX are supported and differentiable. Therefore, we compute the loss function in a fully differentiable manner using `jax.numpy.where()`.

Crucially, light is detected across six different devices. Therefore, we compute the loss function at each detector and the parameter update is driven by the detector that shows the minimum value of the loss. We conduct this selection by using a differentiable, smooth approximation using `jax.nn.logsumexp()` as:

```
def softmin(l_det, beta):
    return - logsumexp(-beta * l_det) / beta,
```

where `l_det` is the array of the loss values corresponding to each detector and `beta` is the strength of the modulation.

C Noise-aware optimization

To ensure robustness of the optical setup solutions, we implement a noise-aware optimization procedure. The method is depicted in Supplementary Fig. 2. This approach allows us to account for potential real-world variations and uncertainties in the physical parameters of the optical system. The optimization procedure is as follows: for each optimization step, we execute N parallel optical tables (in this work, $N = 3$) using JAX's `vmap` functionality. Then, we sample random noise and apply it to all available physical variables across each of the N optical tables. The random noise is uniformly distributed in the phase values for spatial light modulators (SLMs) and wave plates (WP) in the range of \pm (0.01 to 0.1) radians, covering all qualities available in current experimental devices. Also, we incorporate uniformly distributed random misalignment ranging from \pm (0.01 to 0.1) millimeters, covering both expert-level precision (\pm 0.01 mm) and beginner-level accuracy (\pm 0.1 mm). Finally, we include a 1% imperfection on the transmissivity/reflectivity of beam splitters (BS), which is a realistic approach given the high quality of the currently available hardware.

We then simulate the optical setup for each of the N tables simultaneously, incorporating the sampled noise. The loss function is computed independently for each of the setups. Afterwards, we calculate the mean loss value across all optical tables, which provides an average performance metric that accounts for the introduced experimental variability (noise). The gradients are computed based on this mean loss value and so the update of the system parameters^{*}.

Importantly, before applying the updated parameters and proceeding to the next iteration, we resample new random noise for each optical table. This ensures that each optimization step encounters different noise values, further enhancing the robustness of the solution. This procedure is repeated iteratively until convergence.

By incorporating this noise-aware optimization approach, we aim to discover optical designs that perform well in the presence of real-world experimental imperfections.

Single optimization step:

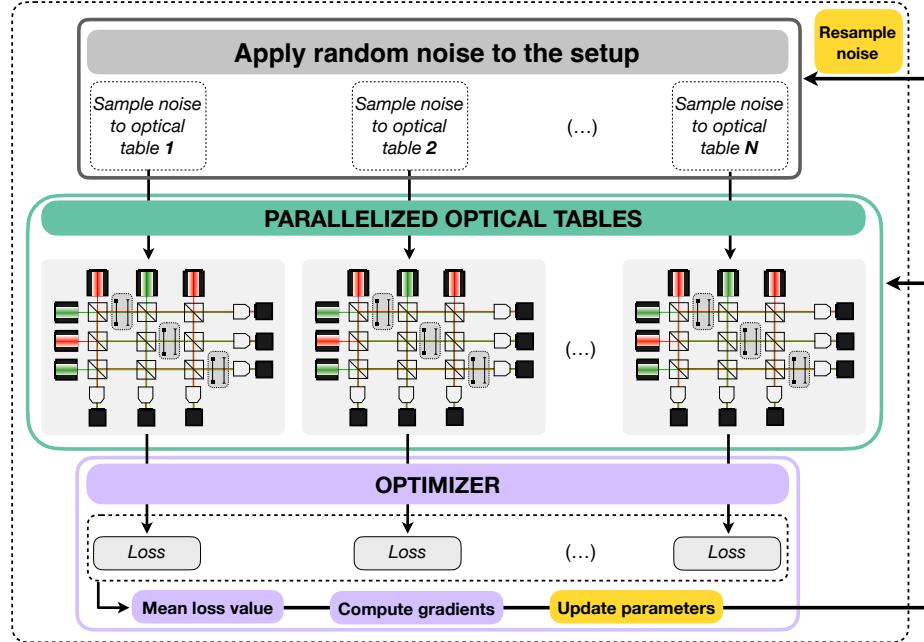


Figure 2: Noise-robust optimization procedure. It consists of optimizing, simultaneously, N multiple instances of the optical setup, each subjected to randomly sampled noise in its physical parameters. At each step, we compute the mean loss across N -different noisy variants of the optical setup and optimize based on the average performance. In this iterative process, the noise is resampled at each step. This method allows us to explore optical solutions that are inherently resilient to noise and maintain high performance even under non-ideal conditions.

D Rediscovery through exploration

We task XLUMINA to rediscover the super-resolution techniques of Dorn, Quabis and Leuchs [24]. For this purpose we build the 3×3 optical setup depicted in Supplementary Fig. 3. It consists of six light sources emitting linearly polarized Gaussian beams of wavelength 650 nm. Three building blocks, which contain one *super-SLM* (i.e., two SLMs imprinting independent phase and amplitude masks to orthogonal polarization states) and a wave plate separated a distance z , are placed within the diagonal of the grid (grey boxes in Supplementary Fig. 3). Light gets ultimately detected across six detectors. As discussed in the previous Appendix section B.1, the loss function is computed at each detector, the parameter update is driven by the device demonstrating the minimum value. This selection is conducted in a fully differentiable manner using `jax.nn.logsumexp()`.

The loss function corresponds to equation (1), in this instance considering the intensity from the longitudinal component of the electric field, $|E_z|^2$, and $\varepsilon = 0.7$. We set-up the AdamW optimizer with a learning rate of 0.05 and a weight decay of 10^{-5} . The system is initialized with three optical tables in parallel, each one showcasing random optical parameters with values between 0 and 1. The parallelized optimization procedure is detailed in the previous Appendix section C. The optimization is terminated if there is no improvement in the loss value over 500 consecutive iteration steps. This condition is checked every 100 steps. The system converged into the topology highlighted in Fig. 2a, demonstrating the smallest loss value in the detector #6. The identified optical parameters correspond to: the wave plate's $\eta = 0.01$, $\theta = -2.1$; propagation distances (in cm): $z_1 = 116.76$, $z_2 = 48.07$, $z_3 = 19.77$, $z_4 = 40.80$, $z_7 = 115.40$ and $z_8 = 96.04$. The beam splitter ratios, in [Transmittance, Reflectance] pairs: BS#4: [0.4837, 0.5162], BS#5: [0.999, 0.000], BS#6: [0.000, 0.999], and BS#9: [0.999, 0.000].

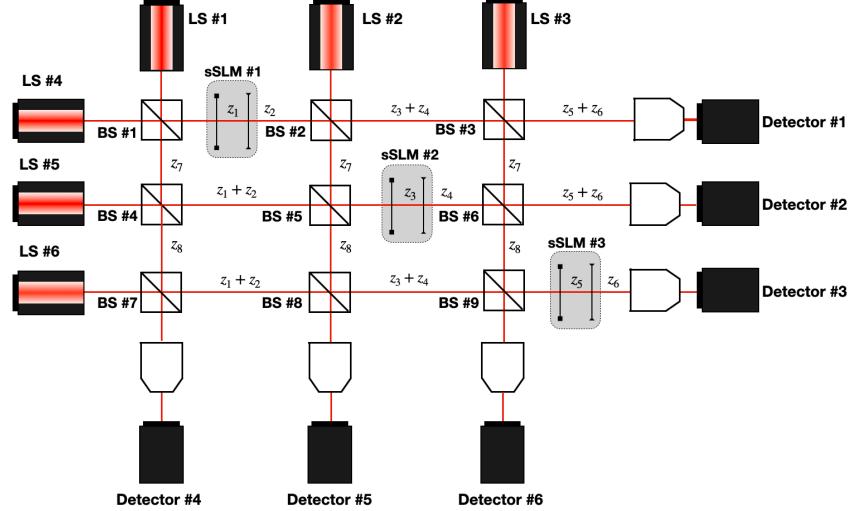


Figure 3: Initial virtual optical setup utilized for rediscover the SR-technique of Dorn, Quabis and Leuchs (2003). It features light sources that emit Gaussian beams with wavelength of 650 nm that are linearly polarized at 45° . Gray boxes, numbered from (1) to (3), represent the building units, each comprising one *super-SLM* (sSLM) and one wave plate (WP). Distances are denoted as z_i where $i = 1, \dots, 8$. After interacting with a high NA objective lens (of NA = 0.9), light gets detected across six detectors (#1 - #6) with $0.05\mu\text{m}$ pixel size screen. The parameter space (~ 4 million optical parameters) contains three sSLM, four wave plates (WPs) with variable phase retardance η and orientation angle θ , eight distances and nine beam splitter ratios. The resolution is set to 1024×1024 pixels with a computational pixel size of $4.8\mu\text{m}$.

E Stimulated emission-depletion model

STED microscopy [9, 25] is based on excitation and spatially targeted depletion of fluorophores. In order to achieve this, a Gaussian-shaped excitation beam and a doughnut-shaped depletion beam (generated by imprinting a spiral phase into its wavefront) are concentrically overlapped. The depletion beam has zero intensity in the center, where the excitation beam has its maximum. Fluorophores that are not in the center of the beams are forced to emit at the wavelength of the depletion beam. Their emission is spectrally filtered out. Only fluorophores in the center of the beams are allowed to fluoresce normally, and only their emission is ultimately detected. This effectively reduces the area of normal fluorescence, which leads to super-resolution imaging.

We simulate one of the fundamental concepts of STED microscopy without having to rely on time-dependent processes related to absorption and fluorescence. To do so, we perform a nonlinear modulation of the intensity of the excitation and depletion beams based on the Beer-Lambert law [26]. We define the effective fluorescence that would ultimately be detected as:

$$I_{\text{eff}} = I_{\text{ex}} \left[1 - \beta \left(1 - e^{-(I_{\text{dep}}/I_{\text{ex}})} \right) \right], \quad (9)$$

where I_{ex} and I_{dep} correspond to the excitation and depletion intensities, respectively, and $0 \leq \beta \leq 1$ captures the quenching efficiency of the depletion beam. This expression bounds the effect of the depletion beam such that scenarios with negative effective intensity or unrealistically high values are avoided. In particular, assuming a perfect efficiency of the depletion beam in suppressing the excitation (i.e., $\beta = 1$), we obtain an expression resembling the Beer-Lambert law:

$$I_{\text{eff}} = I_{\text{ex}} \cdot e^{-(I_{\text{dep}}/I_{\text{ex}})}. \quad (10)$$

Thus, the effective detected light falls off exponentially with the intensity ratio $I_{\text{dep}}/I_{\text{ex}}$. In the limit case where there is no excitation intensity, $I_{\text{ex}} = 0$, the detected light is zero as well, $I_{\text{eff}} = 0$. If there is no depletion intensity, $I_{\text{dep}} = 0$, the detected light corresponds to the excitation beam $I_{\text{eff}} = I_{\text{ex}}$. The trivial case of null efficiency in the quenching, $\beta = 0$, leads to the same result.

To evaluate the nonlinear effect we consider $\beta = 1$ and $I_{\text{dep}} = \frac{1}{2}I_{\text{ex}}$. From equation (9) we obtain

$$I_{\text{eff}} = I_{\text{ex}} e^{-1/2} \approx 0.6I_{\text{ex}}. \quad (11)$$

Now, by slightly increasing the depletion energy, e.g., $I_{\text{dep}} = \frac{3}{2}I_{\text{ex}}$, it reads

$$I_{\text{eff}} = I_{\text{ex}} e^{-3/2} \approx 0.2I_{\text{ex}}. \quad (12)$$

Therefore, a small change in the depletion energy causes a large effect in the effective intensity. As a further example, if we set an intermediate efficiency of $\beta = 0.5$ and $I_{\text{dep}} = \frac{1}{2}I_{\text{ex}}$ we obtain

$$I_{\text{eff}} = I_{\text{ex}} \left[\frac{1 + e^{-1/2}}{2} \right] \approx 0.8I_{\text{ex}} . \quad (13)$$

which clearly demonstrates the effect of diminishing the efficiency of the suppression. Overall, we successfully imprinted the nonlinear behavior of the quenching for different range of effectiveness, achieving a realistic, bounded physical model for STED.

E.1 Discovery of a new experimental blueprint

Finally, we demonstrate the capabilities of XLUMINA for genuine discovery. We use the initial optical setup in Supplementary Fig. 4. The parameter space (~ 2 million parameters) corresponds to two *super-SLMs* (i.e., 4 SLMs) with a resolution of 512×512 pixels with a computational pixel size of $9.6 \mu\text{m}$, two wave plates, six distances and four beam splitter ratios. The loss function corresponds to equation (1), in this instance considering the total intensity of the effective light emerging from the STED process, $|E_x|^2 + |E_y|^2 + |E_z|^2$, and $\varepsilon = 0.5$. We simulate the stimulated emission depletion effect using equation (9) with the efficiency set to $\beta = 1$. We set-up the AdamW optimizer with a step size of 10^{-3} and a weight decay of 10^{-3} .

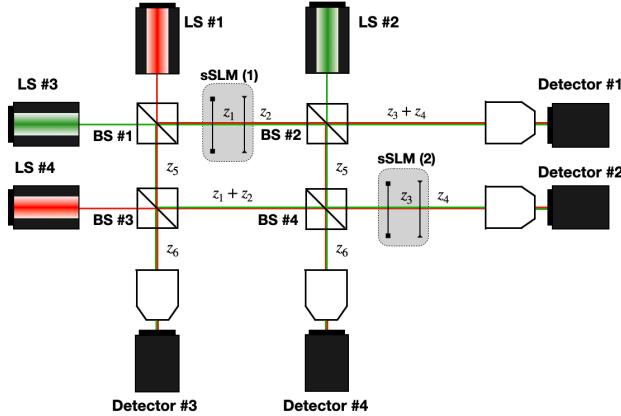


Figure 4: Initial virtual optical setup utilized for discover a new experimental blueprint. It features light sources that emit Gaussian beams with wavelengths of 650 nm and 530 nm linearly polarized at 45° . Gray boxes, numbered (1) and (2), represent the building units, each comprising one *super-SLM* (sSLM) and one wave plate (WP). Distances are denoted as z_i where $i = 1, \dots, 8$. After interacting with a high NA objective lens (of NA= 0.9), light gets detected across four detectors (#1 - #4) with $0.05 \mu\text{m}$ pixel size screen. The parameter space (~ 2 million optical parameters) contains two sSLM, two wave plates (WPs) with variable phase retardance η and orientation angle θ , six distances and four beam splitter ratios. The resolution is set to 512×512 pixels with a computational pixel size of $9.6 \mu\text{m}$.

The system is initialized with three optical tables in parallel, each one showcasing random optical parameters with values between 0 and 1. The parallelized optimization procedure is detailed in the previous Appendix section C. The optimization is terminated if there is no improvement in the loss value over 500 consecutive iteration steps. This condition is checked every 100 steps. The system converged into the topology highlighted in Fig. 2d, demonstrating the smallest loss value in the detector #2. The identified optical parameters correspond to: the beam splitter ratios, in [Transmittance, Reflectance] pairs: BS#1: [0.000, 0.999], BS#2: [0.000, 0.999], BS#3: [0.000, 0.999], and BS#4: [0.306, 0.6937]. The wave plates, in radians (1): $\eta = 0.00$, $\theta = -2.39$, and (2): $\eta = 0.13$, $\theta = 2.50$. The propagation distances (in cm) are $z_1 = 140.109$, $z_2 = 165.67$, $z_3 = 563.69$, $z_4 = 645.44$ and $z_5 = 116.27$.