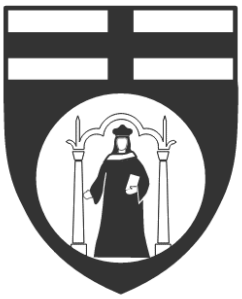

UNIVERSITÀ
DEGLI STUDI
DI GENOVA

Reconstruction and regularization approaches for image scanning microscopy

Ph.D. Thesis submitted by:
Giacomo Garrè

Supervisors:
Dr. Alessandro Zunino
Dr. Giuseppe Vicidomini

Ph.D. Course in *Bioengineering and robotics - bionanotechnology* –
XXXVIII cycle

Contents

Acknowledgements

List of Acronyms

APR	Adaptive Pixel Reassignment
AU	Airy Units
BSA	Bovine Serum Albumin
CLSM	Confocal Laser Scanning Microscopy
CO2	Carbon Dioxide
DFT	Discrete Fourier Transform
DOF	Depth of Field
EGFP	Enhanced Green Fluorescent Protein
FISTA	Fast Iterative Shrinkage–Thresholding Algorithm
FLIM	Fluorescence Lifetime Imaging Microscopy
FLU	Fluorescence Lifetime Unmixing
fPALM	Fluorescence Photoactivation Localization Microscopy
GC	Gradient Consensus
IRF	Impulse Response Function
ISM	Image Scanning Microscopy
KL	Kullback-Leibler
LICS	Live Cell Imaging Solution
LSM	Laser Scanning Microscopy

MID	Multi-Image Deconvolution
MLE	Maximum Likelihood Estimation
MRA	Multi-Resolution Algorithm
NA	Numerical Aperture
PALM	Photoactivated Localization Microscopy
PBS	Phosphate Buffered Saline
PIE	Pulsed Interleaved Excitation
PSF	Point Spread Function
RL	Richardson–Lucy
s2FLISM	Super-Resolved Sectioned Fluorescence Lifetime Image Scanning Microscopy
s2ISM	Super-Resolved Sectioned Image Scanning Microscopy
SIM	Structured Illumination Microscopy
SMLM	Single-Molecule Localization Microscopy
SNR	Signal-to-Noise Ratio
SPAD	Single-Photon Avalanche Diode
SPIM	Selective Plane Illumination Microscopy
SSIM	Structural Similarity Index Measure
STED	Stimulated Emission Depletion
STORM	Stochastic Optical Reconstruction Microscopy
TCSPC	Time-Correlated Single-Photon Counting

Abstract

Confocal laser-scanning microscopy (CLSM), one of the most important imaging tools in biomedical and biological research, is undergoing a major transition towards image-scanning microscopy (ISM). In ISM, a fast detector array replaces the conventional single-element detector and pinhole of the confocal microscope, providing a true image of the region probed during laser scanning while collecting the majority of photons emitted by the specimen. By reconstructing this highly informative dataset, ISM produces images that overcome the traditional trade-off between transverse spatial resolution and signal-to-noise ratio (SNR) that limits confocal microscopy. This transition is further driven by the introduction of single-photon avalanche diode (SPAD) array detectors, whose single-photon timing capabilities and asynchronous readout enable seamless integration with fluorescence lifetime imaging microscopy (FLIM) – one of the most powerful and widely used applications of CLSM – and fluorescence signal sampling at sub-microsecond temporal resolution. However, current ISM reconstruction algorithms face several limitations. First, they lack optical sectioning and fail with thick samples unless the detector size is reduced to act as a pinhole, thereby introducing a new trade-off between sectioning capability and SNR. Second, they are not optimised for FLIM. Finally, the most advanced ISM reconstruction methods are prone to artefacts caused by noise amplification, particularly when iterative optimisation is not properly regularised. In this work, we propose a method that overcomes these limitations. Based on the observation that detector arrays inherently encode axial information from the specimen, we developed a maximum-likelihood estimation (MLE) reconstruction algorithm, termed s²ISM (super-resolved sectioned ISM), which inverts the physical model of ISM image formation. From a single-plane ISM acquisition, s²ISM reconstructs images with super-resolution, high SNR, and enhanced optical sectioning. Within the MLE framework, s²ISM also exploits the redundant information contained in the ISM dataset to achieve digital super-resolution. We further extend s²ISM to FLIM by incorporating the instrument response function (IRF) of the system into the image-formation model to compensate for temporal distortions. The resulting s²FLISM algorithm provides super-resolved FLIM images with enhanced optical sectioning and robust lifetime estimation. Moreover, we introduce a modified version of s²FLISM that enables lifetime-based multi-target imaging, allowing distinct fluorophores to be separated based on their fluorescence lifetimes rather than their spectral properties. Finally, we address the noise amplification inherent to maximum-likelihood estimation by leveraging the SPAD array's temporal sampling capabilities to implement an integrated denoising framework. This strategy effectively suppresses noise amplification, enhances numerical stability, and improves the robustness of the reconstruction process. We validate our method through a comprehensive theoretical analysis and experimental results obtained from calibration standards and biological specimens imaged with a custom-built linear and non-linear laser-scanning microscope equipped with a SPAD array detector. We believe that s²ISM and its lifetime-resolved extension s²FLISM represent a significant advancement for ISM, establishing a unified and versatile framework for super-resolved, lifetime-resolved, and optically sectioned fluorescence microscopy. This approach paves the way for new imaging modalities capable of addressing increasingly complex questions in biological and biomedical research.

1 Introduction

Optical microscopy is an indispensable tool in both scientific research and industrial applications, with a central role in disciplines ranging from materials science to the life sciences. Among the various optical imaging modalities, fluorescence microscopy stands out as a powerful technique, enabling highly specific and minimally invasive visualization of biomolecular spatiotemporal distributions within living cells, tissues, and model organisms [1], thus studying the most fundamental biological processes directly in their native environment.

Despite its versatility, conventional fluorescence microscopy is fundamentally limited by diffraction. Over the past two decades, this limitation has been overcome through the development of super-resolution microscopy techniques, which have extended optical imaging beyond the diffraction limit [2] – which, as a rule of thumb, is around half of the wavelength used to probe the specimen (i.e., around 250 nm for blue light). The super-resolution revolution began with the demonstration of stimulated emission depletion (STED) microscopy in 2000 [3]. In STED microscopy, the effective fluorescent emission region of a laser-scanning microscope – commonly referred to as the probed region – is spatially confined by co-aligning a conventional Gaussian excitation beam with a doughnut-shaped depletion beam. The depletion beam drives excited fluorophores back to the ground state via stimulated emission, thereby reducing the effective size of the fluorescent region. Scanning this reduced probed region across the sample enables lateral resolutions on the order of tens of nanometers [4]. Moreover, the confinement of the probed region can be extended to the axial direction, providing true three-dimensional resolution enhancement. Shortly thereafter, single-molecule localization microscopy (SMLM) techniques were introduced, including stochastic optical reconstruction microscopy (STORM) [5], photoactivated localization microscopy (PALM), and fluorescence photoactivation localization microscopy (fPALM) [6]. These methods achieve nanometric resolution by localizing sparse subsets of stochastically blinking fluorophores with a precision far exceeding the diffraction limit [7]. The resulting localization coordinates are then used to reconstruct a spatial map of fluorophore positions within the cell,

yielding images with sub-diffraction resolution. In SMLM, localization can also be extended along the axial direction, enabling three-dimensional super-resolution imaging. While STED and SMLM provide exceptional spatial resolution – in principle unlimited – their practical application is often constrained by phototoxicity, acquisition speed, and limited compatibility with live-cell imaging. In parallel with the development of super-resolution techniques, increasing attention has been devoted to fast and gentle volumetric imaging methods capable of capturing dynamic processes in three-dimensional samples. Selective plane illumination microscopy (SPIM) is the prominent example [8–10], it achieves optical sectioning – the ability to selectively image a single focal plane within a three-dimensional specimen and thus the key ingredient for volumetric imaging – by restricting excitation to a thin plane within the sample, allowing rapid and gentle volumetric imaging. Advanced implementations, such as lattice light-sheet microscopy [11] and single-objective light-sheet architectures [12], further improve spatial resolution and imaging flexibility. However, achieving high resolution in SPIM typically requires sophisticated optical designs, which can limit its general applicability.

In contrast to the rapid and transformative advances that have characterized super-resolution and fast volumetric imaging techniques, confocal laser-scanning microscopy (CLSM) – a method originally developed in the late 1980s and widely adopted in the 1990s [13–15] – has undergone a more gradual but continuous evolution. Thanks to this incremental development, CLSM has steadily incorporated major technological innovations and remains one of the most versatile and widely used imaging platforms. In terms of performance, CLSM provides a spatial resolution lower than that of super-resolution microscopy techniques but comparable to that of conventional light-sheet–based approaches, while offering volume-rate imaging higher than super-resolution methods yet typically lower than those achievable with SPIM. Optical sectioning is achieved in CLSM through the use of a confocal pinhole, which rejects out-of-focus light at the expense of reduced photon collection efficiency and signal-to-noise ratio (SNR). Nevertheless, CLSM has been continuously improved through technological innovations. Two-photon excitation microscopy [16, 17] enhances optical sectioning and enables deep imaging in scattering tissues by confining fluorescence excitation to the focal volume. Fast scanning technologies, such as resonant scanners, allow video-rate imaging and make CLSM suitable for dynamic live-cell studies. Moreover, thanks to its laser-scanning architecture, namely the sequential point-by-point excitation and detection of the sample, CLSM is naturally compatible with advanced spectroscopic techniques such as fluorescence lifetime imaging microscopy (FLIM) and spectral detection, enabling functional and multiplexed imaging. In particular, FLIM provides contrast based on fluorescence decay dynamics rather than intensity alone, allowing the quantitative probing of molecular interactions, environmental conditions, and functional states.

The need to combine optical sectioning, high spatial and temporal resolution, low phototoxicity, and spectroscopic versatility, while maintaining low the artchitecure complexity, has motivated the development of image scanning microscopy (ISM). ISM was originally proposed as an extension of CLSM [18, 19], exploiting the dual role of the confocal pinhole in both background rejection and resolution enhancement [20]. In conventional confocal microscopy, closing the pinhole improves lateral resolution but severely degrades SNR. ISM addresses this limitation by replacing the single-element detector with a detector array. Each detector element acts as a virtual pinhole, while the array as a whole preserves photon collection efficiency. As a result, ISM

can be interpreted as multiple confocal microscopes observing the same sample in parallel from slightly different perspectives. Through appropriate computational processing, the acquired data can be fused to produce an image with an effective resolution corresponding to that of an ideal confocal microscope with a small pinhole, while preserving high SNR due to the full collection of fluorescence photons. The original implementation of ISM employed a camera-based detection scheme [21], but its temporal resolution was limited by camera frame rates. Subsequent optomechanical multi-spot implementations [22, 23] improved acquisition speed at the cost of increased experimental complexity and reduced compatibility with advanced spectroscopy. More recently, compact detector arrays such as the AiryScan [24] and asynchronous-readout single-photon avalanche diode (SPAD) arrays [25, 26] have enabled practical ISM implementations within conventional laser-scanning architectures. In particular, SPAD arrays offer single-photon sensitivity and timing, excellent temporal resolution, and negligible readout noise [27], while remaining fully compatible with modern CLSM advancements, including resonant scanning and two photon-excitation. Furthermore, the availability of single-photon timing capability enables straightforward integration with time-correlated single-photon counting (TCSPC), allowing each detected photon to be time-stamped relative to the excitation pulse. This time-stamping capability is fundamental for fluorescence lifetime measurements. More generally, single-photon timing also permits the encoding of additional temporal dimensions in the ISM dataset by referencing photon arrival times to external signals, such as the start of the acquisition or changes in the laser scanning position.

The raw data acquired in ISM are inherently four-dimensional in space, comprising a set of confocal-like images corresponding to each detector element. To exploit this information, dedicated reconstruction algorithms are required. The two main classes of ISM reconstruction methods are pixel reassignment [28–30] and multi-image deconvolution [31, 32]. Both approaches improve lateral resolution and SNR beyond the diffraction limit. However, they do not intrinsically provide optical sectioning, and out-of-focus light rejection is typically achieved by limiting the detector size or by applying virtual pinholes [33]. These strategies reduce collection efficiency and cannot fully suppress background contributions. Recent theoretical work has shown that axial information is inherently encoded in the ISM dataset [34, 35].

Building on this insight, a first proof-of-concept method, focus-ISM [36], demonstrated that optical sectioning can be achieved computationally. While effective, this approach relies on simplifying approximations that limit its robustness and integration with other reconstruction strategies. Beyond these spatial limitations, an even more fundamental shortcoming of existing ISM reconstruction algorithms is their restriction to the purely spatial dimensions of the dataset. In particular, they neglect the temporal information carried by photon arrival times, which is essential for fluorescence lifetime imaging microscopy (FLIM). As a result, there currently exists no reconstruction algorithm specifically designed for fluorescence lifetime ISM that jointly exploits the full spatio-temporal information content of the data. In practice, standard spatial ISM reconstruction methods are first applied, and fluorescence lifetime analysis is subsequently performed as a separate post-processing step, leading to a decoupled workflow that does not fully leverage the capabilities of time-resolved ISM data.

In this thesis, we present a novel computational reconstruction framework for ISM that fully exploits the rich spatio-temporal information provided by SPAD-array–based detection. By op-

erating directly on the complete dataset, our approach enables the simultaneous realization of optical super-resolution, optical sectioning, and high signal-to-noise ratio without imposing constraints on detector size or introducing additional optical complexity. Importantly, this framework applies seamlessly to both intensity-based imaging and fluorescence lifetime imaging, thereby extending the capabilities of ISM beyond conventional reconstruction paradigms.
Within the structured illumination framework [37], ISM is modeled through a physically grounded forward formulation, which is then inverted using a maximum-likelihood estimation approach akin to the Richardson–Lucy algorithm [38, 39],(App. Note G). The reconstruction explicitly separates in-focus and out-of-focus contributions, allowing background rejection to be performed entirely in post-processing. Unlike methods based on engineered illumination or detection, such as RESCH [40, 41], and engineered ISM [42], our approach does not require wavefront shaping and preserves the simplicity and lateral super-resolution capability of conventional ISM. Since optical sectioning and super-resolution are achieved simultaneously, we refer to this technique as Super-resolution Sectioning Image Scanning Microscopy (s^2ISM). We naturally extend the s^2ISM framework to incorporate photon arrival times and the system impulse response function, enabling direct reconstruction from TCSPC data. This extension allows super-resolved fluorescence lifetime imaging with high precision, yielding more accurate estimation of fluorescence lifetime values. In addition, we introduce a variant of the algorithm that integrates multi-target imaging based on distinct fluorescence lifetime signatures. This approach enables the unmixing of signals originating from multiple fluorophores with different lifetimes while simultaneously reconstructing a super-resolved and sectioned image of the underlying cellular structures. By jointly performing image reconstruction and lifetime-based unmixing within a single computational framework, our method avoids the loss of information inherent to conventional two-step or phasor-based approaches [43].

Finally, the inversion of the ISM forward model constitutes an ill-posed problem and is susceptible to noise amplification. Rather than relying on *a posteriori* stopping rules, this thesis explores regularization strategies that promote stable convergence. In particular, we adopt a Gradient Consensus approach [44], which exploits the single-photon time-tagging capability of SPAD array detectors to generate statistically independent measurements intrinsic to SPAD-based ISM. This strategy provides effective noise suppression without compromising resolution or physical interpretability.

Chapter 1 is devoted to a general introduction of the image restoration problem that we are going to tackle. Starting with a deep description of the IS microscope to a modelization of the physical phenomena describing the formation of the final ISM dataset. The advantages and drawbacks of its application are discussed, also considering the state-of-the-art methods available for this microscopy framework.

Chapter 2 presents the inversion of the forward model leading to the explicit form of the reconstruction algorithm. Also, different side data-driven methods capable of extracting needed parameters from the raw data are presented and eviscerated.

Chapter 3 focuses on the validation of the algorithm on synthetic data. In this part, we quantitatively characterize a lot of physical quantities that must be properly tuned in order to

optimize the algorithm capabilities depending on the boundary conditions of the problem.

Chapter 4 shows application of the optimized reconstruction method to various experimental samples, varying in structure dimensions, signal-to-noise ratios, and for different incoherent LSM techniques. Also, we demonstrate the capability of our method for performing digital super-resolution under proper sample acquisition conditions. We prove the latter through qualitative and quantitative analysis of a few experimental data.

Chapter 5 extends the physical-mathematical forward model of raw data formation to incorporate the temporal dimension. We highlight the unique features of the SPAD array detector that enable new capabilities within the FLIM estimation framework. These capabilities are first validated using phantom datasets and subsequently demonstrated on an experimental spatio-temporal dataset.

Chapter 6 studies the ill-posed nature of the inversion problem at hand, highlighting that our reconstruction algorithm is also susceptible to noise overfitting during the iterative process. To address this, we explore a range of regularization strategies—including explicit loss-based regularizers, Plug-and-Play methods, deep learning-based regularizers, and inherent regularization techniques. We analyze the strengths and limitations of each approach, ultimately motivating our choice of the Gradient Consensus method as the regularization strategy for our framework.

Chapter 7 presents our conclusions and outlines related ongoing work. We also highlight future research directions enabled by our findings.

References

[1] Jeff W Lichtman and José-Angel Conchello. Fluorescence microscopy. *Nature Methods*, 2:910–919, 12 2005.

[2] Kirti Prakash, Benedict Diederich, Rainer Heintzmann, and Lothar Schermelleh. Super-resolution microscopy: a brief history and new avenues. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 380, 4 2022.

[3] Thomas A. Klar, Stefan Jakobs, Marcus Dyba, Alexander Egner, and Stefan W. Hell. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proceedings of the National Academy of Sciences*, 97(15):8206–8210, 2000.

[4] Giuseppe Vicidomini, Paolo Bianchini, and Alberto Diaspro. Sted super-resolved microscopy. *Nature Methods*, 15:173–182, 2018.

[5] Michael Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nature methods*, 3:793–5, 11 2006.

[6] Samuel T. Hess, Thanu P.K. Girirajan, and Michael D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical Journal*, 91(11):4258–4272, 2006.

[7] Mickaël Lelek, Melina Gyparaki, Gerti Beliu, Florian Schueder, Juliette Griffie, Suliana Manley, Ralf Jungmann, Melike Lakadamyali, and Christophe Zimmer. Single-molecule localization microscopy. *Nature Reviews Methods Primers*, 1, 12 2021.

[8] Jan Huisken, Jim Swoger, Filippo Del Bene, Joachim Wittbrodt, and Ernst H K Stelzer. Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science*, 305:1007–1009, 8 2004.

[9] Alessandro Zunino, Francesco Garzella, Alberta Trianni, Peter Saggau, Paolo Bianchini, Alberto Diaspro, and Martí Duocastella. Multiplane encoded light-sheet microscopy for enhanced 3d imaging. *ACS Photonics*, 8:3385–3393, 11 2021.

[10] Ernst H. K. Stelzer, Frederic Strobl, Bo-Jui Chang, Friedrich Preusser, Stephan Preibisch, Katie McDole, and Reto Fiolka. Light sheet fluorescence microscopy. *Nature Reviews Methods Primers*, 1:73, 11 2021.

[11] Bi-Chang Chen, Wesley R. Legant, Kai Wang, Lin Shao, Daniel E. Milkie, Michael W. Davidson, Chris Janetopoulos, Xufeng S. Wu, John A. Hammer, Zhe Liu, Brian P. English, Yuko Mimori-Kiyosue, Daniel P. Romero, Alex T. Ritter, Jennifer Lippincott-Schwartz, Lillian Fritz-Laylin, R. Dyche Mullins, Diana M. Mitchell, Joshua N. Bembenek, Anne-Cecile Reymann, Ralph Böhme, Stephan W. Grill, Jennifer T. Wang, Geraldine Seydoux, U. Serdar Tulu, Daniel P. Kiehart, and Eric Betzig. Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science*, 346, 10 2014.

[12] Bin Yang, Merlin Lange, Alfred Millett-Sikking, Xiang Zhao, Jordão Bragantini, Shruthi VijayKumar, Mason Kamb, Rafael Gómez-Sjöberg, Ahmet Can Solak, Wanpeng Wang, Hirofumi Kobayashi, Matthew N. McCarroll, Lachlan W. Whitehead, Reto P. Fiolka, Thomas B. Kornberg, Andrew G. York, and Loic A. Royer. Daxi—high-resolution, large imaging volume and multi-view single-objective light-sheet microscopy. *Nature Methods*, 19:461–469, 4 2022.

[13] C.J.R. Sheppard and A. Choudhury. Image formation in the scanning microscope. *Optica Acta: International Journal of Optics*, 24:1051–1073, 10 1977.

[14] C. J. R. Sheppard and C. J. Cogswell. Three-dimensional image formation in confocal microscopy. *Journal of Microscopy*, 159:179–194, 8 1990.

[15] Peter O. Bayguinov, Dennis M. Oakley, Chien-Cheng Shih, Daniel J. Geanon, Matthew S. Joens, and James A. J. Fitzpatrick. Modern laser scanning confocal microscopy. *Current Protocols in Cytometry*, 85:e39, 7 2018.

[16] Winfried Denk, James H. Strickler, and Watt W. Webb. Two-photon laser scanning fluorescence microscopy. *Science*, 248:73–76, 4 1990.

[17] Alberto Diaspro, Paolo Bianchini, Giuseppe Vicidomini, Mario Faretta, Paola Ramoino, and Cesare Usai. Multi-photon excitation microscopy. *BioMedical Engineering OnLine*, 5:36, 12 2006.

[18] Colin Sheppard. Super-resolution in confocal imaging. *Optik*, 80:53–54, 1988.

[19] M Bertero, P Boccacci, M Defrise, C De Mol, and E R Pike. Super-resolution in confocal scanning microscopy: Ii. the incoherent case. *Inverse Problems*, 5:441–461, 8 1989.

[20] Colin Sheppard. Super-resolution by confocal fluorescent microscopy. *Optik*, 60:391–396, 1982.

[21] Claus B. Müller and Jörg Enderlein. Image scanning microscopy. *Physical Review Letters*, 104:198101, 5 2010.

[22] Stephan Roth, Colin J.R. Sheppard, Kai Wicker, and Rainer Heintzmann. Optical photon reassignment microscopy (opra). *Optical Nanoscopy*, 2:1–6, 2013.

[23] Andrew G York, Panagiotis Chandris, Damian Dalle Nogare, Jeffrey Head, Peter Wawrzusin, Robert S Fischer, Ajay Chitnis, and Hari Shroff. Instant super-resolution imaging in live cells and embryos via analog image processing. *Nature Methods*, 10:1122–1126, 11 2013.

[24] Joseph Huff, Annette Bergter, Jan Birkenbeil, Ingo Kleppe, Ralf Engelmann, and Uros Krzic. The new 2d superresolution mode for zeiss airyscan. *Nature Methods*, 14:1223–1223, 12 2017.

[25] Ivan Michel Antolovic, Claudio Bruschini, and Edoardo Charbon. Dynamic range extension for photon counting arrays. *Optics Express*, 26:22234, 8 2018.

[26] Mauro Buttafava, Federica Villa, Marco Castello, Giorgio Tortarolo, Enrico Conca, Mirko Sanzaro, Simonluca Piazza, Paolo Bianchini, Alberto Diaspro, Franco Zappa, Giuseppe Vicidomini, and Alberto Tosi. Spad-based asynchronous-readout array detectors for image-scanning microscopy. *Optica*, 7:755, 7 2020.

[27] Eli Slenders, Eleonora Perego, Mauro Buttafava, Giorgio Tortarolo, Enrico Conca, Sabrina Zappone, Agnieszka Pierzynska-Mach, Federica Villa, Enrica Maria Petrini, Andrea Barberis, Alberto Tosi, and Giuseppe Vicidomini. Cooled spad array detector for low light-dose fluorescence laser scanning microscopy. *Biophysical Reports*, 1:100025, 12 2021.

[28] Colin J. R. Sheppard, Shalin B. Mehta, and Rainer Heintzmann. Superresolution by image scanning microscopy using pixel reassignment. *Optics Letters*, 38:2889, 8 2013.

[29] Marco Castello, Giorgio Tortarolo, Mauro Buttafava, Takahiro Deguchi, Federica Villa, Sami Koho, Luca Pesce, Michele Oneto, Simone Pelicci, Luca Lanzanó, Paolo Bianchini, Colin J. R. Sheppard, Alberto Diaspro, Alberto Tosi, and Giuseppe Vicidomini. A robust and versatile platform for image scanning microscopy enabling super-resolution flim. *Nature Methods*, 16:175–178, 2 2019.

[30] Daniele Ancora, Alessandro Zunino, Giuseppe Vicidomini, and Alvaro H. Crevenna. Image scanning microscopy reconstruction by autocorrelation inversion. *Journal of Physics: Photonics*, 2024.

[31] Maria Ingaramo, Andrew G. York, Eelco Hoogendoorn, Marten Postma, Hari Shroff, and George H. Patterson. Richardson-lucy deconvolution as a general tool for combining images with complementary strengths. *ChemPhysChem*, 15:794–800, 3 2014.

[32] Alessandro Zunino, Marco Castello, and Giuseppe Vicidomini. Reconstructing the image scanning microscopy dataset: An inverse problem. *Inverse Problems*, 39, 6 2023.

[33] Colin J. R. Sheppard, Marco Castello, Giorgio Tortarolo, Takahiro Deguchi, Sami V. Koho, Giuseppe Vicidomini, and Alberto Diaspro. Pixel reassignment in image scanning microscopy: a re-evaluation. *Journal of the Optical Society of America A*, 37:154, 1 2020.

[34] Colin J. R. Sheppard, Marco Castello, Giorgio Tortarolo, Alessandro Zunino, Eli Slenders, Paolo Bianchini, Giuseppe Vicidomini, and Alberto Diaspro. Signal strength and integrated intensity in confocal and image scanning microscopy. *Journal of the Optical Society of America A*, 40:138, 1 2023.

[35] Colin J. R. Sheppard, Marco Castello, Giorgio Tortarolo, Alessandro Zunino, Eli Slenders, Paolo Bianchini, Giuseppe Vicidomini, and Alberto Diaspro. Background rejection in two-photon fluorescence image scanning microscopy. *Photonics*, 10:601, 5 2023.

[36] Giorgio Tortarolo, Alessandro Zunino, Francesco Fersini, Marco Castello, Simonluca Piazza, Colin J.R. Sheppard, Paolo Bianchini, Alberto Diaspro, Sami Koho, and Giuseppe Vicidomini. Focus image scanning microscopy for sharp and gentle super-resolved microscopy. *Nature Communications*, 13, 12 2022.

[37] Colin J. R. Sheppard. Structured illumination microscopy and image scanning microscopy: a review and comparison of imaging properties. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379:20200154, 6 2021.

[38] William Hadley Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62:55, 1 1972.

[39] L. B. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79:745, 6 1974.

[40] Alexander Jesacher, Monika Ritsch-Marte, and Rafael Piestun. Three-dimensional information from two-dimensional scans: a scanning microscope with postacquisition refocusing capability. *Optica*, 2:210, 3 2015.

[41] Clemens Roider, Rainer Heintzmann, Rafael Piestun, and Alexander Jesacher. Deconvolution approach for 3d scanning microscopy with helical phase engineering. *Optics Express*, 24:15456, 7 2016.

[42] Clemens Roider, Rafael Piestun, and Alexander Jesacher. 3d image scanning microscopy with engineered excitation and detection. *Optica*, 4:1373, 11 2017.

[43] Giorgio Tortarolo, Alessandro Zunino, Simonluca Piazza, Mattia Donato, Sabrina Zappone, Agnieszka Pierzyńska-Mach, Marco Castello, and Giuseppe Vicidomini. Compact and effective photon-resolved image scanning microscope. *Advanced Photonics*, 6, 1 2024.

[44] Jan Becker Craig Russell Andrew York, James Manton. Automatic data-driven parameter-free optimal deconvolution.

2

Beyond Confocal Microscopy: The Image Scanning Microscopy Framework

2.1 Confocal Laser Scanning Microscope (CLSM)

LSM approaches employ a point-scanning architecture that moves a focused excitation volume across the three-dimensional field of view of the microscope. At each scanning position, fluorescence emitted by all fluorophores within the probed volume is collected by a single-element detector, yielding pixel intensities proportional to the local fluorophore concentration.

In 1904, Ernst Abbe showed that the size of the focal volume – and hence the spatial resolution of a point-scanning microscope – is limited by diffraction:

$$d_{x_s} = d_{y_s} = \frac{\lambda}{2n \sin \alpha}, \tag{2.1}$$

where x_s and y_s denote lateral coordinates in the focal plane, λ is the excitation wavelength, n is the refractive index of the medium in which light is focused, and α is the half-angle of the focused illumination cone. Using the numerical aperture $\mathrm{NA} = n \sin \alpha$, it follows that resolution is principally determined by the wavelength λ and the numerical aperture NA. One of the first techniques, at least theoretically, able to exceed the diffraction limit was confocal LSM (CLSM). This type of microscope extends the resolution of conventional LSM by introducing a pinhole in front of the detector, which provides: (i) optical sectioning, out-of-focus fluorescence is physically rejected, improving the signal-to-background ratio (SBR) and enabling clear axial sectioning; (ii) resolution enhancement, with an ideally small pinhole, the lateral size of the probed region, and thus the resolution, reduces by a factor of $\sqrt{2}$ in terms of FWHM, which corresponds to a doubling of the spatial cut-off frequency of the system.

The excellent optical sectioning capability of confocal microscopy has been a major factor in its widespread adoption, making it one of the most prevalent imaging techniques in life sciences research. In contrast, the resolution improvement beyond the diffraction limit is seldom realised, as it requires an extremely small pinhole aperture. Under such conditions, only a limited number of fluorescence photons reach the detector, causing a substantial decrease in SNR (Fig.2.1b). Consequently, users typically open the pinhole to obtain acceptable SNR, a practice that progressively reduces – and ultimately eliminates – the intrinsic super-resolution benefit.

2.1.1 Image formation in confocal LSM

To the aim of this Phd work, it is useful to recall that a CLSM can be modelled as a linear and spatially invariant system, and can therefore be fully characterised by its impulse response function. In fluorescence microscopy, this function represents the intensity distribution recorded when imaging a point-like emitter, such as a single fluorescent molecule whose size is below the diffraction limit; hence, the term point-spread function (PSF). Owing to diffraction, the emitted light cannot be refocused into an ideal point; instead, it spreads over a finite region of the image plane, forming the characteristic diffraction spot. To analytically derive the PSF of a confocal system, it is convenient to treat the illumination and detection paths separately. Within the framework of LSM, this resulting intensity distribution is commonly referred to as the *illumination point spread function*, denoted as h_{exc}. The intensity distribution generated by a Dorac delta emitter is referred to as the *detection point spread function*, h_{det}, as it represents the image of a point source formed in the detection plane of the system.

Importantly, the overall PSF of an imaging system can be expressed directly in terms of the illumination and detection PSFs. To this end, we assume that both illumination and detection of the CLSM are constrained by finite apertures, denoted a_{exc} and a_{det}, respectively. In a confocal microscope, the illumination aperture is defined by the illumination pinhole, while the detection aperture is set by the size of the detection pinhole.

A point in the specimen is illuminated through the aperture a_{exc} of the light source, which can be modelled as a collection of point emitters distributed across the aperture with constant intensity C. Each point emitter is projected onto the sample by the objective lens. The resulting illumination intensity distribution can therefore be expressed as the convolution between the illumination aperture and the diffraction-limited PSF:

$$i_{exc}(\boldsymbol{x}_s) = C a_{exc}(\boldsymbol{x}_s) * h_{exc}(\boldsymbol{x}_s) \tag{2.2}$$

where $*$ is the convolution operator and $\boldsymbol{x}_s = (x_s, y_s)$ the set of scanning coordinate. Assuming that fading and saturation fluorescence effects can be neglected, the emitted light i_{em} is proportional to the illumination intensity i_{exc} , to the fluorescence molecules concentration in the sample, represented by the function $o(\boldsymbol{x}_s)$, and to the brightness B of the fluorescence molecules

$$i_{em}(\boldsymbol{x}_s) = i_{exc}(\boldsymbol{x}_s) o(\boldsymbol{x}_s) B \tag{2.3}$$

The emitted light is imaged by the same objective onto a photo-sensitive sensor. Behind the objective, the detection aperture limits the emission light with respect to the sensor. The

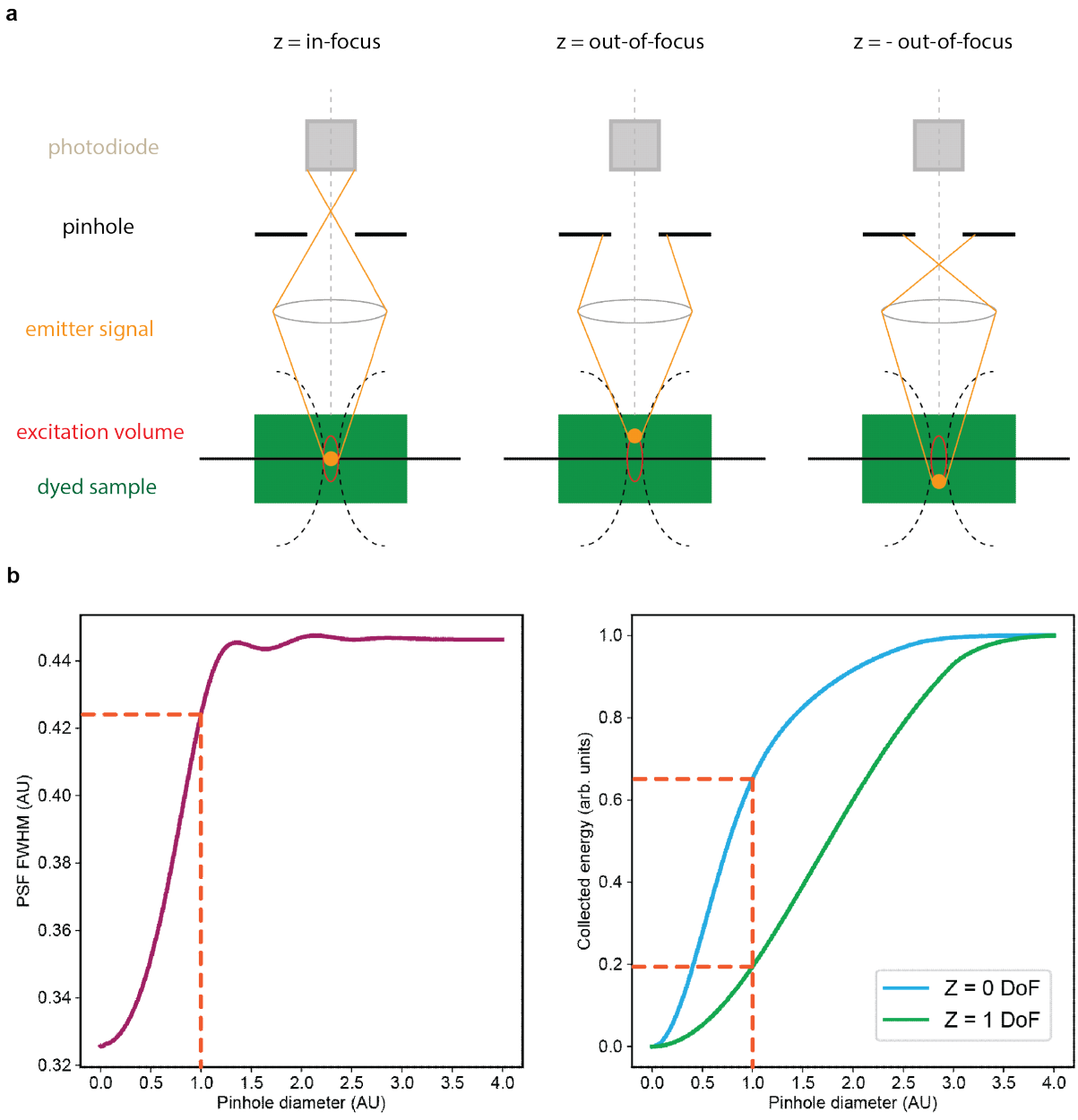

Figure 2.1: **a** Sketch of background rejection from the physical pinhole in a CLSM. **b** Trade-off representation between lateral resolution and photon budget collected (from focus and non) depending on the size of the physical pinhole. The orange line depicts a medium-sized pinhole for a CLSM architecture. Usually, small pinhole configurations stand around 0.28 AU, open pinhole configurations around 1.4 AU.

intensity behind this aperture can be modeled as

$$i_{det}(\boldsymbol{x}_s) = a_{det}(\boldsymbol{x}_s)[h_{det}(\boldsymbol{x}_s) * i_{em}(\boldsymbol{x}_s)] \tag{2.4}$$

The above formulations are valid for the general fluorescence microscope, and they can be tuned to different microscopy configurations by choosing the appropriate values for the illumination and detection apertures. In the case of confocal microscopy, a formulation for the signal i_{det} can be obtained assuming that both the excitation and the detection pinholes are infinitely small - condition which is formally modeled by using two Dirac functions $\delta(\boldsymbol{x}_s)$. Combining the previous equations, we can therefore write the detection intensity idet for a given point of the scan:

$$i_{\mathrm{det}}(\boldsymbol{x}_s) = \delta(\boldsymbol{x}_s)\left[h_{\mathrm{det}}(\boldsymbol{x}_s) * \big(\, h_{\mathrm{exc}}(\boldsymbol{x}_s) * \delta(\boldsymbol{x}_s)\,\big) o(\boldsymbol{x}_s)\right]. \tag{2.5}$$

The detection intensity is recorded by a single-element detector, resulting in a single intensity value for each point of the scan. Such a value can be obtained by integrating over the surface of the sensor:

$$\begin{aligned} i_{\mathrm{CLSM}}(0) &= \int_{-\infty}^{+\infty} i_{\mathrm{det}}(x')\,dx' = \int_{-\infty}^{+\infty} \delta(x')\Big[h_{\mathrm{det}}(x') * \big(h_{\mathrm{exc}}(x') * \delta(x')\big) o(x')\Big]\,dx' \\ &= \int_{-\infty}^{+\infty} \delta(x')\Big[h_{\mathrm{det}}(x') * \big(h_{\mathrm{exc}}(x') o(x')\big)\Big]\,dx' \\ &= \big(h_{\mathrm{det}} * (h_{\mathrm{exc}} o(x))\big)(0) = \int_{-\infty}^{+\infty} h_{\mathrm{det}}(-x)\, h_{\mathrm{exc}}(x)\, o(x)\,dx. \end{aligned} \tag{2.6}$$

Since the bi-dimensional image in a CLSM is obtained by scanning the specimen, the intensity recorded by the photodiode for each point can be derived by shifting both the illumination and detection over the object

$$\begin{aligned} i_{\mathrm{CLSM}}(x) &= \int_{-\infty}^{+\infty} \delta(x'-x)\Big[h_{\mathrm{det}}(x') * \big(h_{\mathrm{exc}}(x') * \delta(x'-x)\big)\, o(x')\Big]\,dx' \\ &= \int_{-\infty}^{+\infty} \delta(x'-x)\Big[h_{\mathrm{det}}(x') * \big(h_{\mathrm{exc}}(x'-x)\, o(x')\big)\Big]\,dx' \\ &= \int_{-\infty}^{+\infty} \delta(x'-x)\left[\int_{-\infty}^{+\infty} h_{\mathrm{det}}(x'-\xi)\, h_{\mathrm{exc}}(\xi-x)\, o(\xi)\,d\xi\right] dx' \\ &= \int_{-\infty}^{+\infty} h_{\mathrm{det}}(x-\xi)\, h_{\mathrm{exc}}(\xi-x)\, o(\xi)\,d\xi \\ &= \big(h_{\mathrm{det}}(x)\, h_{\mathrm{exc}}(-x)\big) * o(x). \end{aligned} \tag{2.7}$$

Substituting the object function with the Dirac delta function, we find the PSF of an infinitely closed CLSM:

$$h_\delta(\boldsymbol{x}_s) = h_{det}(\boldsymbol{x}_s)h_{exc}(\boldsymbol{x}_s) \tag{2.8}$$

The previous formulation is valid only under the condition of an infinitely closed pinhole. To extend this result to real scenarios, we replace the Dirac delta function with the finite-size aperture, obtaining the PSF of a CLSM with the finite pinhole aperture a_{det}:

$$h_{CLSM}(\boldsymbol{x}_s) = [a_{det}(\boldsymbol{x}_s) * h_{det}(\boldsymbol{x}_s)]h_{exc}(\boldsymbol{x}_s) \tag{2.9}$$

This equation highlights the fundamental trade-off between resolution and SNR discussed earlier. Consider a perfectly aligned system in which $\lambda_{\text{exc}} = \lambda_{\text{det}}$ (i.e., in the absence of a Stokes shift), and assume that both the illumination PSF h_{exc} and the detection PSF h_{det} can be approximated as identical Gaussian distributions. Under these assumptions, it holds true

$$h_{\text{CLSM}}(\boldsymbol{x}_s) = G(0,\sigma^2) * \big(G(0,\sigma^2) * a_{\text{det}}(\boldsymbol{x}_s)\big), \tag{2.10}$$

where $G(\mu,\sigma^2)$ denotes a Gaussian distribution centered at μ with variance σ^2.

If the confocal pinhole is considered infinitely open ($a_{\text{det}}(\boldsymbol{x}_s) = 1, \forall \boldsymbol{x}_s$), the resulting PSF reduces to the diffraction-limited illumination PSF, $G(0,\sigma^2)$. This configuration provides standard resolution but maximizes the SNR due to the large photon throughput.

Conversely, for an infinitely closed pinhole ($a_{\text{det}}(\boldsymbol{x}_s) = \delta(\boldsymbol{x}_s)$), the PSF becomes

$$h_{\text{CLSM}}(\boldsymbol{x}_s) = G^2(0,\sigma^2) = G\Big(0,\frac{\sigma^2}{\sqrt{2}}\Big), \tag{2.11}$$

indicating a lateral resolution improvement by a factor of $\sqrt{2}$ relative to the diffraction limit.

2.2 Image Scanning Microscopy (ISM)

Image scanning microscopy (ISM) overcomes the resolution–signal trade-off imposed by the physical pinhole in CLSM by removing the pinhole entirely and replacing the single photodiode with a detector array. In this configuration, ISM preserves high lateral resolution – since each element of the array effectively functions as a small pinhole – while simultaneously enabling efficient signal collection due to the detector's large photon-sensitive area. This makes ISM relatively easy to implement on existing CLSM platforms. The data collected by this new type of architecture must be processed to reconstruct a single image with super-resolution and high SNR. Currently, two methods are the most popular, namely pixel-reassignment and multi-image deconvolution.

However, both methods present a limitation: a large detector array not only captures strong in-focus signals but also collects substantial out-of-focus light from adjacent planes. As we will show, the raw ISM data embed a physical quantity that encodes the axial position of each emitter. By leveraging this information, we can build a method that simultaneously improves SNR, lateral resolution, and optical sectioning capabilities.

2.2.1 Principle of ISM

For this work, we used a custom ISM microscope (A.1). The instrument is a LSM equipped with a 5×5 asynchronous-readout single photon avalanche diode (SPAD) array detector [1]. A laser beam focused at the scan point $\boldsymbol{x}_s = (x_s, y_s)$ excites the fluorescent molecules from a diffraction-limited volume within the specimen. The array detector collects the emitted light from each pixel location $\boldsymbol{x}_d = (x_d, y_d)$, building a four-dimensional dataset. The dimensionality of the produced dataset is higher for time-resolved acquisitions. In this case, time acts as an additional scanning coordinate, which spam across the photon-arrival time dimension. For the sake of simplicity, we leave the treatment of the spatiotemporal dataset for another chapter, while here we focus only on the spatial dimensions.

The sensitive elements of the detector can be seen as independent pinholes, making the ISM microscope equivalent to multiple confocal microscopes observing the same sample in parallel, each from a different point of view. Thus, the ISM dataset can be interpreted as a set of confocal-like images, each blurred by a unique PSF. In detail, the PSF recorded by the sensitive element at position $\boldsymbol{x}_d$ is given by

$$h(\boldsymbol{x}_s, z_s|\boldsymbol{x}_d) = h_{\text{exc}}(-\boldsymbol{x}_s, z_s) \cdot [h_{em}(\boldsymbol{x}_s, z_s) * p(\boldsymbol{x}_s - \boldsymbol{x}_d)] = h_{\text{exc}}(-\boldsymbol{x}_s, z_s) \cdot h_{\text{det}}(\boldsymbol{x}_s - \boldsymbol{x}_d, z_s) \tag{2.12}$$

where $*$ is the convolution operator with respect to the coordinate $\boldsymbol{x}_s$, z_s is the axial coordinate of the emitter, p is the function describing the geometry of the detector element's sensitive area, h_{exc} and h_{em} are the excitation and emission PSFs of the microscope. We saw before that the detection PSFs h_{det} is the convolution of the emission PSFs with the pinhole. The above equation explains the physical reason behind the super-resolution in ISM. Indeed, for sufficiently small sensors, h_{det} is very similar to h_{exc}. Thus, the product of the two PSFs is sharper than the individual ones. Conversely, if the sensor size is too large, h_{det} approaches a constant value and the super-resolution effect is washed out – as for non-confocal laser-scanning microscopy.

The position $\boldsymbol{x}_d$ of the detector element plays a key role in determining the shape and the relative intensity of each PSF (Fig 2.2a). The modulation of the light intensity on the detector coordinates are then rescaled according to a normalization factor known as the *fingerprint*

$$f(\boldsymbol{x}_d, z_s) = \int_{\mathbb{R}^2} h(\boldsymbol{x}_s, z_s|\boldsymbol{x}_d)\, \mathrm{d}\boldsymbol{x}_s = h_{\text{exc}}(\boldsymbol{x}_d, z_s) \star h_{\text{det}}(\boldsymbol{x}_d, z_s) \tag{2.13}$$

where $\star$ is the cross-correlation operator. The fingerprint, in perfectly aligned optical conditions, is a bell-shaped distribution peaked at the centre of the detector, and the more the emitter is defocused, the broader the distribution is Fig 2.2b.

2.2.2 Forward operator - modeling data formation in ISM

Given the linearity of incoherent image formation, the ISM dataset $i(\boldsymbol{x}_s|\boldsymbol{x}_d)$ is given by the superposition of the fluorescence light emitted from any plane

$$i(\boldsymbol{x}_s|\boldsymbol{x}_d) = \int_{\mathbb{R}} o(\boldsymbol{x}_s, z) * h(\boldsymbol{x}_s, z|\boldsymbol{x}_d)\, \mathrm{d}z \tag{2.14}$$

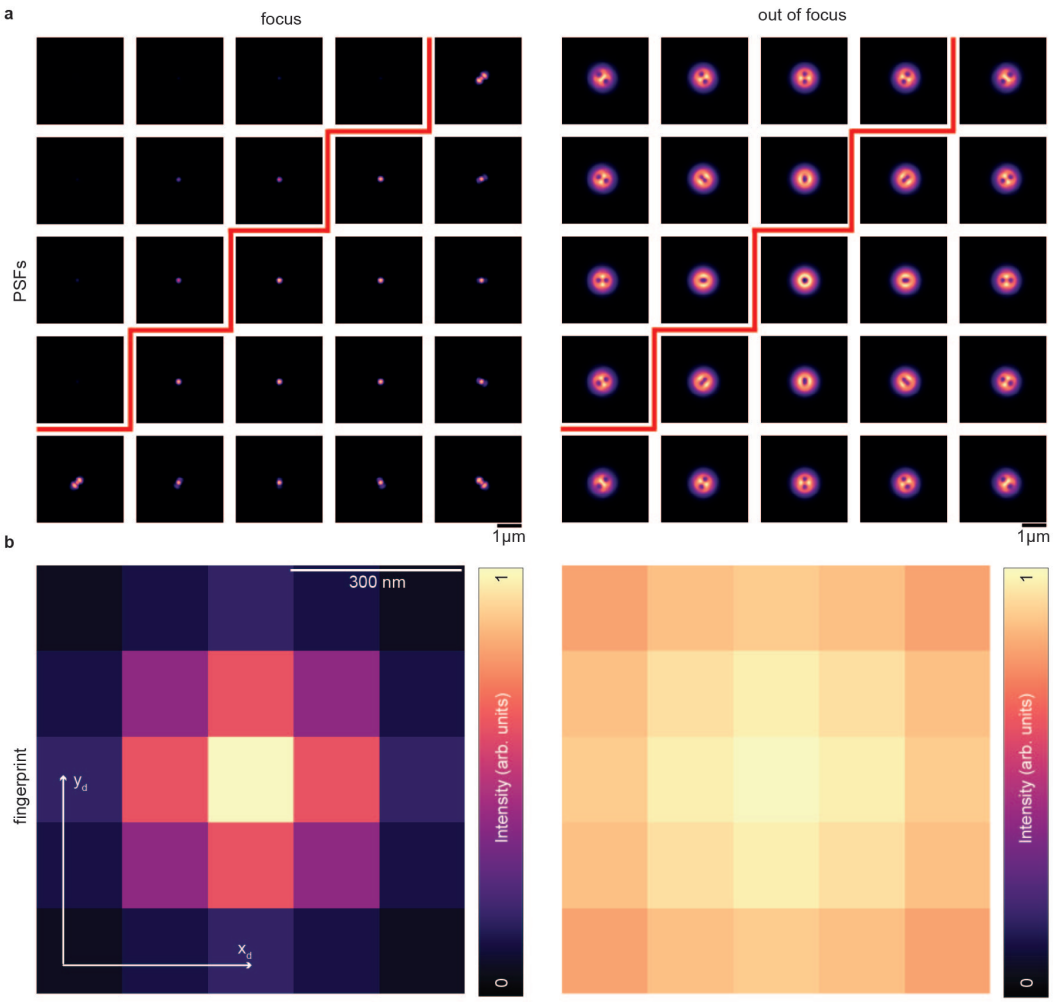

Figure 2.2: **In focus and out-of-focus PSFs and fingerprints.** **a** the in focus PSFs and the out-of-focus PSFs. Top left corner: each PSF is normalized with respect to the central element of the SPAD array. Bottom right corner: each PSF is normalized to itself. In **b**, the respective fingerprints.

where o is the object, namely the specimen's 3D distribution of emitters, and the integration is performed along the z-axis. Even though the shape of the PSFs depends on the axial coordinate, it cannot be directly extracted to identify the depth of the emitters. Indeed, the lateral structure of the specimen is convolved with the PSFs, and their relative contributions cannot be disentangled unless some prior knowledge is available. However, Eq. 2.13 suggests that the light stemming from different axial positions distributes differently on the detector plane, regardless of the lateral sample structure. Thus, the fingerprint inherently encodes the axial information into the detector plane dimension. Despite not allowing for a full volumetric reconstruction of the specimen, the fingerprint map is informative enough to enable the discrimination of the defocused light in a single-plane acquisition.

Assuming that the relevant out-of-focus contributions stem from a finite and discrete number

of planes, we re-write the forward model of ISM image formation as follows

$$i(\boldsymbol{x}_s|\boldsymbol{x}_d) = \sum_{k=1}^{N} f_k(\boldsymbol{x}_d)[o_k(\boldsymbol{x}_s) * \hat{h}_k(\boldsymbol{x}_s|\boldsymbol{x}_d)] \tag{2.15}$$

where k is the discretized axial position and $\hat{h}$ are the normalized PSFs. Thus, in-focus and out-of-focus emitters are weighted by a different fingerprint function. Namely, the modulation of contrast on the detector plane varies according to the axial position of the sample. The images generated by the central and peripheral elements of the detector array collect light mainly from in-focus and out-of-focus planes, respectively (Fig. 2.3). Our goal is to leverage the fingerprint map – uniquely provided by the ISM architecture – to design a reconstruction procedure that builds a single super-resolution image starting from a single-plane ISM dataset, excluding defocused contributions.

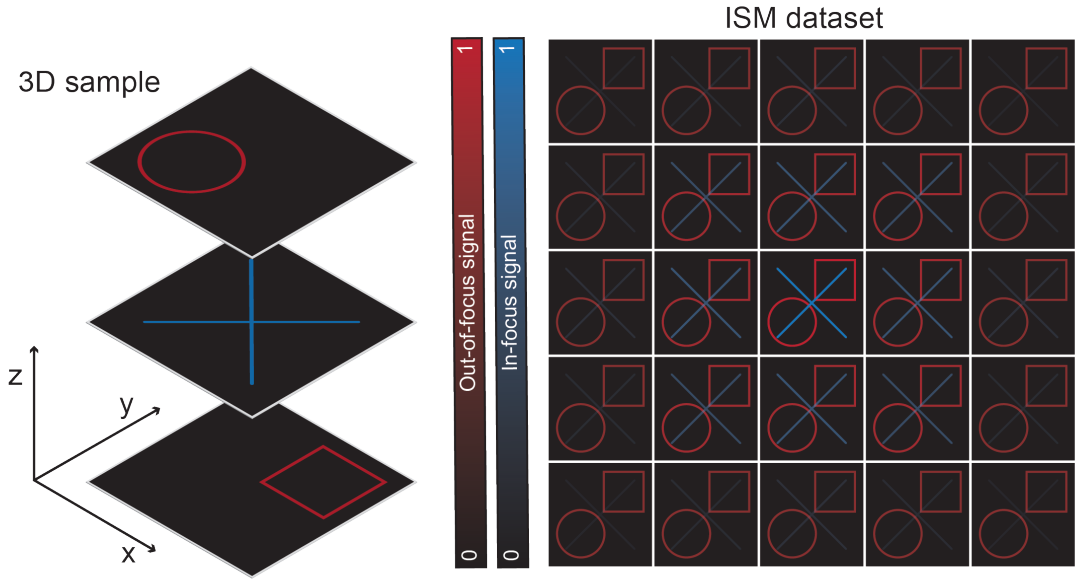

Figure 2.3: **ISM imaging a 3D sample.** Schematic representation of data formation in an ISM when imaging a thick sample. The axial information is encoded in the final data through the variation in pixel distribution width.

2.2.3 Lateral resolution and optical sectioning in ISM

ISM can be regarded as a special case of sequential structured illumination microscopy (SIM). Indeed, for each scan point $\boldsymbol{x}_s$, we are exciting the sample only on the diffraction-limited region described by the excitation PSF h_{exc}. The resulting image recorded by the detector array is

$$i(\boldsymbol{x}_d|\boldsymbol{x}_s)\Big|_{z_d=0} = [o(\boldsymbol{x}_d - \boldsymbol{x}_s) \cdot h_{\text{exc}}(\boldsymbol{x}_d)] * h_{\text{em}}(\boldsymbol{x}_d)\Big|_{z_d=0} \tag{2.16}$$

where $\boldsymbol{x}_d = (x_d, y_d, z_d)$, $\boldsymbol{x}_s = (x_s, y_s, z_s)$, and $*$ represents the 3D convolution with respect the the $\boldsymbol{x}_d$ coordinate. The corresponding Fourier transform is

$$I(\boldsymbol{k}_d|\boldsymbol{x}_s) = \left\{ \left[e^{i\boldsymbol{k}_d \cdot \boldsymbol{x}_s} O(\boldsymbol{k}_d) \right] * H_{\text{exc}}(\boldsymbol{k}_d) \right\} \cdot H_{\text{em}}(\boldsymbol{k}_d) \tag{2.17}$$

where $\boldsymbol{k}_d = (k_x, k_y, k_z)$ are the spectral coordinates, and the capital letters represent the Fourier transform of the corresponding quantity in real space. The excitation 3D optical transfer function (OTF) contains multiple spatial frequencies, up to following lateral and axial cut-off frequencies

$$k_{xy}^{\text{max}} = \frac{2n \sin \alpha}{\lambda} \qquad k_z^{\text{max}} = \frac{2n \sin^2 \alpha/2}{\lambda} \tag{2.18}$$

where α is the semi-angular aperture of the objective lens. Thus, the excitation OTF can be rewritten using the following identity

$$H_{\text{exc}}(\boldsymbol{k}_d) = \int_{\Omega} H_{\text{exc}}(\boldsymbol{k}) \delta(\boldsymbol{k}_d - \boldsymbol{k}) \, \mathrm{d}\boldsymbol{k} \tag{2.19}$$

where Ω is the support of the 3D OTF, namely $\Omega = \{\boldsymbol{k}_d \in \mathbb{R}^3 : |H_{\text{exc}}(\boldsymbol{k}_d)| > 0\}$. Plugging Eq. 2.19 into Eq. 2.17, we obtain

$$I(\boldsymbol{k}_d|\boldsymbol{x}_s) = H_{\text{em}}(\boldsymbol{k}_d) \cdot \int_{\Omega} H_{\text{exc}}(\boldsymbol{k}) \left[e^{i\boldsymbol{k}_d \cdot \boldsymbol{x}_s} O(\boldsymbol{k}_d) * \delta(\boldsymbol{k}_d - \boldsymbol{k}) \right] \mathrm{d}\boldsymbol{k} \tag{2.20}$$

$$= H_{\text{em}}(\boldsymbol{k}_d) \cdot \int_{\Omega} H_{\text{exc}}(\boldsymbol{k}) \left[e^{i(\boldsymbol{k}_d - \boldsymbol{k}) \cdot \boldsymbol{x}_s} O(\boldsymbol{k}_d - \boldsymbol{k}) \right] \mathrm{d}\boldsymbol{k} \tag{2.21}$$

Namely, the emitted fluorescence from the sample using a focused illumination spot has the effect of scanning the object's spatial spectrum in the frequency space. More in detail, the frequency content of the illumination enables access to the object's lateral frequencies up to twice the wide-field cut-off, allowing for super-resolution.

$$k_{xy}^{\text{max}} = 2\text{NA} \frac{\lambda_{\text{exc}} + \lambda_{\text{em}}}{\lambda_{\text{exc}} \lambda_{\text{em}}} \approx \frac{4NA}{\lambda} \tag{2.22}$$

where we assumed $\lambda_{\text{exc}} \approx \lambda_{\text{em}} = \lambda$. At the same time, the illumination frequencies enable filling the *missing cone* of the microscope's OTF, granting higher optical sectioning compared to wide-field microscopy. The object's accessible frequencies are those allowed by the ideal confocal 3D OTF, calculated as the auto-convolution of the wide-field 3D OTF. Unlike confocal microscopy, ISM does not discard light to extend the OTF, enabling high resolution and optical sectioning without compromising the SNR. The shifted spectra are available as a weighted sum in the Fourier transform of the raw images. Thus, reconstructing the sample's image with enhanced resolution and optical sectioning requires a computational approach.

2.2.4 ISM-specific state of the art algorithms

Adaptive Pixel Reassignment (APR)

We calculated the shift-vectors of an ISM dataset as

$$\boldsymbol{\mu}(\boldsymbol{x}_d) = \arg\max_{\boldsymbol{x}_s} \{\mathcal{R}(\boldsymbol{x}_s|\boldsymbol{x}_d)\} \tag{2.23}$$

where $\mathcal{R}$ is the phase correlation of the raw images to the central one,

$$\mathcal{R}(\boldsymbol{x}_s|\boldsymbol{x}_d) = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{i(\boldsymbol{x}_s|\boldsymbol{x}_d)\}\overline{\mathcal{F}\{i(\boldsymbol{x}_s|\boldsymbol{0})\}}}{|\mathcal{F}\{i(\boldsymbol{x}_s|\boldsymbol{x}_d)\}\overline{\mathcal{F}\{i(\boldsymbol{x}_s|\boldsymbol{0})\}}}\right\} \tag{2.24}$$

and $\mathcal{F}$ is the Fourier transform operator. The APR reconstruction [2] is calculated as the sum of the aligned images

$$i_{APR}(\boldsymbol{x}_s) = \sum_{\boldsymbol{x}_d} i(\boldsymbol{x}_s + \boldsymbol{\mu}(\boldsymbol{x}_d)|\boldsymbol{x}_d) \tag{2.25}$$

The total computation time is 20 seconds for a $2000 \times 2000 \times 25$ dataset on a computer equipped with an 8-core CPU ($3.6\,\mathrm{GHz}$) and 32 GB of RAM.

Multi-Image deconvolution (MID)

MID's approach to SNR and lateral resolution enhancement is slightly different from the APR one. We need to model physically the forward process of the formation of the image in ISM. Then, through a Maximum Likelihood Estimation (MLE) approach, and by estimating/acquiring the PSFs of the system, we invert the model, finding an iterative rule to recover the unknown investigated object. The forward model in ISM for a thin object is:

$$i(\boldsymbol{x}_s|\boldsymbol{x}_d) = o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s|\boldsymbol{x}_d) \tag{2.26}$$

Then, thanks to the great properties of the SPAD array detector, we can impose a Poisson distribution on the raw dataset:

$$i(\boldsymbol{x}_s|\boldsymbol{x}_d) \sim Poisson[o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s|\boldsymbol{x}_d)] \tag{2.27}$$

leading to a likelihood function and its maximization, in this scenario, through gradient ascent, to obtain an estimator of the noumenon $o(\boldsymbol{x}_s)$.

Let n be the number of events in a fixed interval. In our scenario, an event is the arrival of a photon on the impinging area of the detector; thus, with mean rate $\lambda = o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s|\boldsymbol{x}_d)$. Thanks to SPADs' robustness against dark noise, events occur independently, then n follows a Poisson distribution:

$$P(n = k) = P(i(\boldsymbol{x}_s|\boldsymbol{x}_d) = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad k \in \mathbb{N} \tag{2.28}$$

with

$$\mathbb{E}[i(\boldsymbol{x}_s|\boldsymbol{x}_d)] = \mathrm{Var}[i(\boldsymbol{x}_s|\boldsymbol{x}_d)] = \lambda \tag{2.29}$$

This property leads us to the inversion of the forward model through a Bayesian approach, which will be deeply explored in Section 3.1, dealing with the s^{2}ISM algorithm derivation.

This method further enhances the benefits of the APR algorithm, although it still does not provide optical sectioning. Careful control of the iterative process is essential, as stopping too

early or performing excessive iterations can result in noise overfitting and a reduction of the achievable reconstruction gains.

Both APR and MID are highly sensitive to the SNR of the raw data. In APR, low SNR can lead to inaccurate correlation estimates and misassignments, ultimately producing invalid reconstructions. In MID, low SNR may cause premature overfitting during the iterative optimization process.

A key difference between the two approaches is predictability: the SNR and lateral resolution gains achievable with APR can be estimated when the boundary conditions of the raw data are known. In contrast, although MID is deterministic due to its gradient descent implementation, the achieved gains depend strongly on the chosen number of iterations. This makes a theoretical characterization of MID performance—based solely on the boundary conditions of the raw data—considerably more challenging.

A key advantage of MID is its ability to explicitly incorporate the system's PSFs, in contrast to APR, which is entirely physics-agnostic. This enables MID to reliably reconstruct data under a wide range of non-ideal conditions, which is not achievable with APR.

The computation time is $\sim$8 seconds per iteration for a $2000 \times 2000 \times 25$ dataset on a computer equipped with an 8-core CPU ($3.6\,\mathrm{GHz}$) and 32 GB of RAM. Table 2.1 provides a multi-view comparison of the two algorithms.

Aspect	**APR**	**MID**
Core principle	Shift-vector estimation via phase correlation and summation of aligned images	Inversion of a physical forward model via Maximum Likelihood Estimation
Mathematical framework	Fourier-domain correlation, deterministic alignment	Poisson statistical model, gradient-based optimization
Forward model	Not required (physics-agnostic)	Explicit convolution model $i(x_s\|x_d) = o(x_s) * h(x_s\|x_d)$
Noise modeling	Implicit (no explicit statistical model)	Explicit Poisson statistics
Reconstruction strategy	Single-pass alignment + summation	Iterative deconvolution
Resolution enhancement	Predicted to be $\sqrt{2}$ in space domain	Potentially higher, iteration-dependent
SNR sensitivity	High: small decreases in SNR can cause inaccurate correlation estimates and misassignment of pixels, producing poor reconstructions.	High: low SNR may cause premature overfitting during iterative optimization; reconstruction quality strongly depends on iteration control. See Chapter 5)
SNR robustness	Low: it cannot compensate for low SNR or non-ideal conditions; performance drops quickly when data is noisy.	Moderate to High: MID explicitly incorporates system PSFs and uses a Bayesian/MLE approach, allowing it to better handle low SNR or non-ideal conditions if iterations are properly controlled.
Predictability of gain	Gains estimable from acquisition boundary conditions	Hard to theoretically characterize (depends on stopping rule)
Robustness to non-ideal PSFs	Limited	Strong (explicit PSF incorporation)
Optical sectioning	Not provided	Not provided
Computation time ($2000\times2000\times25$)	$\sim$20 s (single reconstruction)	$\sim$8 s per iteration
Implementation complexity	Relatively simple	Requires PSF estimation and stopping criterion tuning (See Chapter 5).

Table 2.1

Focus-ISM

The focus-ISM algorithm [3] exploits the APR approach to register the images of the ISM dataset. After APR and under reasonable assumptions, the micro-images of an ISM dataset are proportional to the fingerprint of the image. The micro-image of a 3D object can be written as:

$$i(\boldsymbol{x}_d \mid \boldsymbol{x}_s, z) = \int [o(\boldsymbol{x}_s - \boldsymbol{x}_d, z') \cdot h_{\text{exc}}(-\boldsymbol{x}_d, z - z')] * h_{\text{det}}(-\boldsymbol{x}_d, z - z')\, dz' \tag{2.30}$$

where the convolution is calculated along the $(\boldsymbol{x}_d, y_d)$ dimensions. Assuming that the sample is localized on a single plane, we have $o(\boldsymbol{x}_s - \boldsymbol{x}_d, z) = o(\boldsymbol{x}_s - \boldsymbol{x}_d)\, \delta(z - z_0)$. Using this assumption, the micro-image of an object placed at distance z_0 from the image plane z becomes:

$$i(\boldsymbol{x}_d \mid \boldsymbol{x}_s, z) = [o(\boldsymbol{x}_s - \boldsymbol{x}_d, z_0) \cdot h_{\text{exc}}(-\boldsymbol{x}_d, z - z_0)] * h_{\text{det}}(-\boldsymbol{x}_d, z - z_0) \tag{2.31}$$

After pixel reassignment (Fig. 2.4a), the micro-image is given by:

$$i(\boldsymbol{x}_d \mid \boldsymbol{x}_s, z) = \alpha(\boldsymbol{x}_s, z_0) \cdot [h_{\text{exc}}(-\boldsymbol{x}_d, z - z_0) * h_{\text{det}}(-\boldsymbol{x}_d, z - z_0)] \tag{2.32}$$

The fingerprint of an image generated by a 3D distribution of emitters $o(\boldsymbol{x}_s, z)$ is:

$$\begin{aligned}
f(\boldsymbol{x}_d, z) &= \int_{\mathbb{R}^2} i(\boldsymbol{x}_s \mid \boldsymbol{x}_d)\, d\boldsymbol{x}_s = \int_{\mathbb{R}^2} o(\boldsymbol{x}_s, z) * h(\boldsymbol{x}_s, z \mid \boldsymbol{x}_d)\, d\boldsymbol{x}_s \\
&= \int_{\mathbb{R}^2} o(\boldsymbol{x}_s, z) * \left[h_{\text{exc}}(-\boldsymbol{x}_s, z) \cdot h_{\text{det}}(\boldsymbol{x}_s - \boldsymbol{x}_d, z)\right] d\boldsymbol{x}_s \\
&= \int \left[\int_{\mathbb{R}^2} o(\boldsymbol{x}_s, z - z_0)\, d\boldsymbol{x}_s \int_{\mathbb{R}^2} h_{\text{exc}}(-\boldsymbol{x}_s, z_0) \cdot h_{\text{det}}(\boldsymbol{x}_s - \boldsymbol{x}_d, z_0)\, d\boldsymbol{x}_s \right] dz_0 \\
&= \int \gamma(z - z_0) \cdot \left[h_{\text{exc}}(-\boldsymbol{x}_d, z_0) * h_{\text{det}}(-\boldsymbol{x}_d, z_0)\right] dz_0 \\
&= \gamma(z) * \left[h_{\text{exc}}(-\boldsymbol{x}_d, z) * h_{\text{det}}(-\boldsymbol{x}_d, z)\right]
\end{aligned} \tag{2.33}$$

where $\gamma(z) = \int_{\mathbb{R}^2} o(\boldsymbol{x}_s, z)\, d\boldsymbol{x}_s$ is a weight function that depends only on the sample distribution.

The first convolution is calculated along the z-axis, while the second is along the $\boldsymbol{x}_s$-axis. Assuming the sample is localized on a single plane, $\gamma(z) = \gamma_0\, \delta(z - z_0)$, and we obtain:

$$f(\boldsymbol{x}_d, z - z_0) = \gamma_0\, h_{\text{exc}}(-\boldsymbol{x}_d, z - z_0) * h_{\text{det}}(-\boldsymbol{x}_d, z - z_0) \tag{2.34}$$

which is proportional to the micro-image after reassignment. Finally, for multiple emitters at different axial planes, the micro-image can be expressed as:

$$i(\boldsymbol{x}_d) = \sum_{k=1}^{K} \alpha_k\, f(\boldsymbol{x}_d, z_k) \tag{2.35}$$

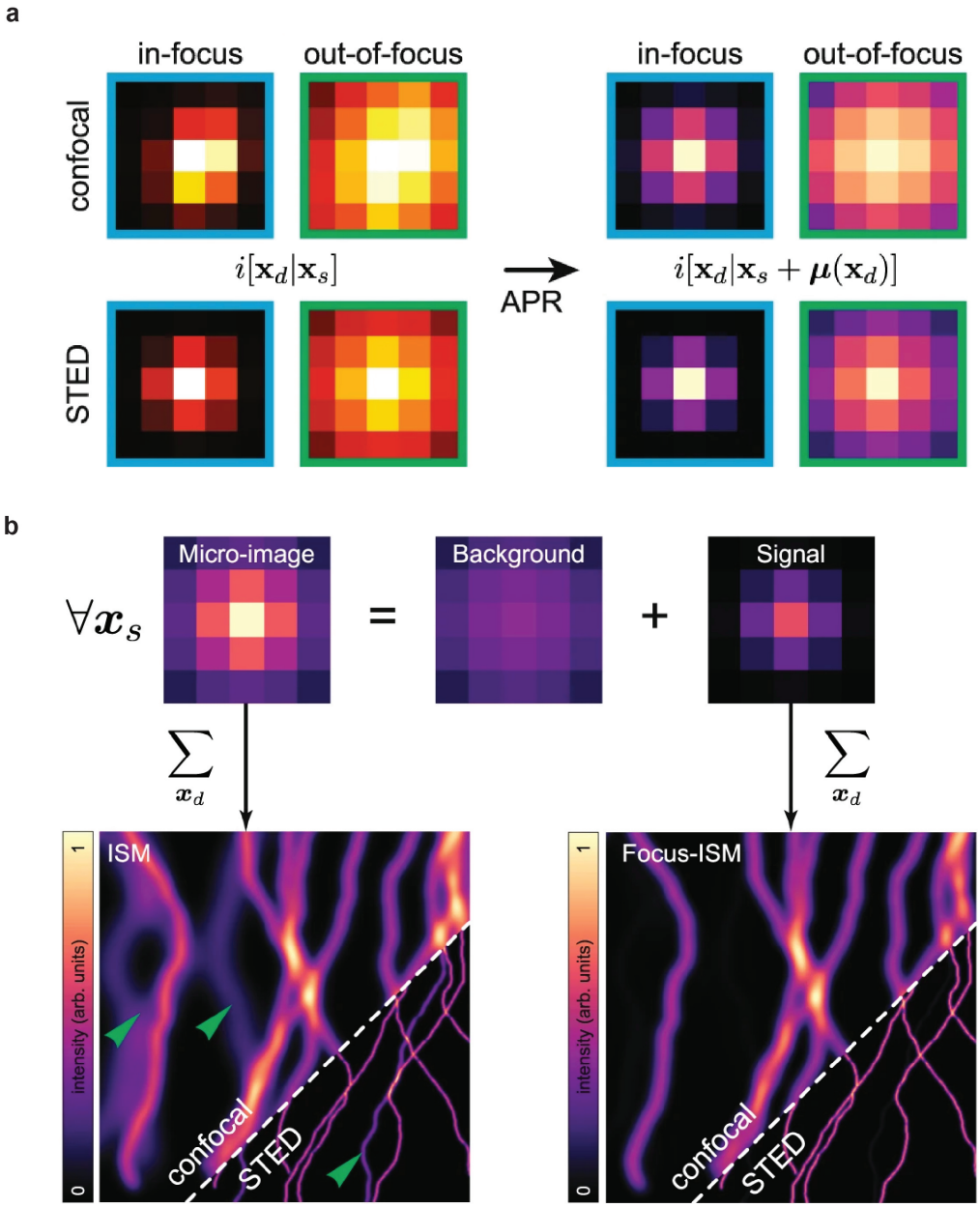

Figure 2.4: **APR and Focus-ISM working principle. a** APR registers a shift for each detector channel (in the first line, we see the encoding of the shift of the orange-highlighted SPAD array channel) using a correlation-based approach, thus subsequently reassigns each image onto the optical axis. Summing all reassigned images produces the APR reconstruction, which exhibits enhanced resolution and signal-to-noise ratio (SNR). Additionally, APR regularizes the micro-images, ensuring proper alignment even in cases where initial registration is imperfect. **b** Focus-ISM needs a region of the sample in which only in-focus emitters are present. In such a way, it encodes the in-focus micro-image, thus being able to reject the out-of-focus micro-image for each scanning point of the data, enhancing optical sectioning.

Then, instead of summing the result, the algorithm fits each reassigned and normalized micro-image to the following two-component Gaussian mixture model :

$$i(\boldsymbol{x}_d) = \alpha \cdot g(\boldsymbol{x}_d|\mathbf{0}, \sigma_{sig}) + (1-\alpha) \cdot g(\boldsymbol{x}_d|\mathbf{0}, \sigma_{bkg}) \tag{2.36}$$

where $g(\boldsymbol{x}_d|\mathbf{0}, \sigma)$ is a centred Gaussian function and σ_{sig} is kept fixed following calibration procedure. Finally, the map of weights $\alpha(\boldsymbol{x}_s)$ is applied to the APR image to remove the background (Fig. 2.4b). This post-processing tool has also been extended to the STED acquisition regime (Fig. 2.4), leading to great optical sectioning enhancement.

The total computation time is 1 hour and 30 minutes for a $2000 \times 2000 \times 25$ dataset on a computer equipped with an 8-core CPU ($3.6\,\mathrm{GHz}$) and 32 GB of RAM.

Conclusion of Chapter 2

Starting from image formation in confocal microscopy, the ISM forward model was derived, highlighting the role of detector architecture and the fingerprint function in encoding axial information. The analysis demonstrated that ISM inherently contains depth-dependent information that can be computationally exploited to achieve optical sectioning. By framing ISM within the structured illumination formalism, this chapter establishes the physical and mathematical foundations for the proposed reconstruction algorithms.

References

[1] Mauro Buttafava, Federica Villa, Marco Castello, Giorgio Tortarolo, Enrico Conca, Mirko Sanzaro, Simonluca Piazza, Paolo Bianchini, Alberto Diaspro, Franco Zappa, Giuseppe Vicidomini, and Alberto Tosi. Spad-based asynchronous-readout array detectors for image-scanning microscopy. *Optica*, 7:755, 7 2020.

[2] Marco Castello, Giorgio Tortarolo, Mauro Buttafava, Takahiro Deguchi, Federica Villa, Sami Koho, Luca Pesce, Michele Oneto, Simone Pelicci, Luca Lanzanó, Paolo Bianchini, Colin J. R. Sheppard, Alberto Diaspro, Alberto Tosi, and Giuseppe Vicidomini. A robust and versatile platform for image scanning microscopy enabling super-resolution flim. *Nature Methods*, 16:175–178, 2 2019.

[3] Giorgio Tortarolo, Alessandro Zunino, Francesco Fersini, Marco Castello, Simonluca Piazza, Colin J.R. Sheppard, Paolo Bianchini, Alberto Diaspro, Sami Koho, and Giuseppe Vicidomini. Focus image scanning microscopy for sharp and gentle super-resolved microscopy. *Nature Communications*, 13, 12 2022.

3

The Super-resolution Sectioning Image Scanning Algorithm (s^2ISM)

Optical sectioning is the ability of a fluorescence microscopy system to selectively visualize a thin axial slice of a three-dimensional specimen while suppressing signal contributions originating from out-of-focus planes. This capability is essential in fluorescence microscopy, where fluorescence emitted by regions outside the focal plane degrades image contrast, spatial resolution, and quantitative accuracy. Effective optical sectioning enables high-contrast imaging in thick and densely labeled samples, facilitates volumetric reconstruction, and is a prerequisite for the reliable interpretation of spatially resolved biological processes, particularly in tissue and small-animal imaging.

Optical sectioning can be categorized into three non-mutually exclusive strategies: (i) avoiding excitation of out-of-focus planes, (ii) removing out-of-focus fluorescence before detection, and (iii) computationally removing out-of-focus light. Category (i) comprises techniques that do not excite the sample outside the focal region. A representative example is SPIM, which only excites a thin sample slice [1]. Another technique that falls in the same category is multi-photon excitation microscopy (MPEM), where the non-linear interaction with the sample occurs exclusively within the focal volume [2, 3]. Thanks to long wavelengths, MPEM is robust to scattering and enables deep penetration into tissues. However, MPEM requires high optical power, which increases both costs and the risk of photo-damaging to the sample. Category (ii) involves blocking light before reaching the detector. This goal can be achieved by physically filtering the out-of-focus light with a pinhole, as in CLSM [4]. However, complete rejection of out-of-focus light inevitably compromises the detection of in-focus photons and, consequently, the SNR. Another strategy employs destructive interference to remove out-of-focus light, as in 4Pi [5–7] and I^5M microscopy [8, 9]. These techniques provide axial super-resolution and superior optical sectioning, but they increase experimental complexity, making them cumbersome for live applications. Category (iii)

comprises techniques that eliminate out-of-focus light during post-processing by encoding axial information into the data. The most important representative of this category is structured illumination microscopy (SIM), which can be used to either achieve optical sectioning (OS-SIM) [10, 11] or super-resolution (SR-SIM) [12], depending on the periodicity of the illumination pattern [13]. The two effects can be combined using a three-dimensional structured illumination (3D-SIM) [14]. Even single-plane SIM acquisitions can achieve both SR and OS using multi-layer [15, 16] or two-layer [17] reconstruction techniques. Thus, SIM enables high-contrast super-resolution imaging. However, its capability to decode additional information from the images depends on the contrast of the illumination pattern, which may be severely reduced in thick and crowded samples [18]. Finally, some authors devised strategies to reduce the background content even from conventional images using purely computational approaches [19]. However, the physics of wide-field image formation prevents the complete separation of in-focus light from out-of-focus light. This limitation is due to the phenomenon of the *missing cone* in the 3D optical transfer function. As a result, all these strategies depend – either directly or indirectly – on some prior knowledge of the out-of-focus content. Commonly, out-of-focus background is assumed to have low intensity and low spatial frequencies [20]. In specific scenarios, these assumptions are appropriate and lead to an effective rejection of out-of-focus light. However, these approaches lack generality and may result in the unintended rejection of in-focus light or the introduction of artefacts.

In the original implementation of ISM, optical sectioning is achieved by preserving confocality through the use of a detector array whose effective size is chosen to be on the order of one Airy unit. This configuration enables partial rejection of out-of-focus fluorescence; however, it inevitably discards a non-negligible fraction of in-focus photons. Such photon loss directly contradicts one of the core motivations of ISM, namely the efficient collection of all available fluorescence signal to maximize the SNR. In Chapter 2, we described how Focus-ISM addresses this limitation by exploiting the fact that the axial position of emitters is intrinsically encoded in the raw ISM dataset. While conceptually effective, Focus-ISM relies on a two-step reconstruction procedure, which limits its applicability in low-photon-count regimes. Moreover, the sequential nature of the algorithm results in increased computational complexity and extended processing times. Building upon the same underlying principle—that axial information is encoded in the raw ISM data—we here derive a fundamentally different reconstruction strategy. By formulating a unified inverse problem, we introduce an algorithm that simultaneously provides optical super-resolution, optical sectioning, and SNR enhancement in a single reconstruction step. This approach overcomes the intrinsic limitations of state-of-the-art ISM reconstruction methods for optical sectioning and, more broadly, establishes a comprehensive and physically grounded framework for ISM data processing.

Thereafter, we show the step-by-step derivation of our novel method s^2ISM. It is able to surpass the state-of-the-art inherent trade-offs by facing the problem in a completely different way, namely by modeling the image formation of the optical setup, and inverting it through a Maximum Likelihood Estimation (MLE) approach. It comes embedded with a data-driven procedure to extract fundamental parameters needed for the reconstruction directly from the raw ISM dataset, alleviating the burden on the user and optimizing performance. Finally, we characterize s^2ISM capabilities quantitatively, comparing the latter with ISM-specific and non-

specific algorithms, showing its superiority in performance, generality, and versatility. Owing to its mathematical–physical foundations, the method exhibits strong robustness against artifact creation and aberration uncertainty.

3.1 Mathematical derivation of s²ISM

As introduced in the previous chapters, we assume an incoherent image formation process, leading to the complete forward model for an image scanning microscope:

$$i(\boldsymbol{x}_s|\boldsymbol{x}_d) = o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s|\boldsymbol{x}_d) \tag{3.1}$$

where $\boldsymbol{x}_s = (x_s, y_s, z_s)$ and $\boldsymbol{x}_d = (x_d, y_d)$ are the coordinates of the sample plane and detector plane, respectively. The operator $*$ is the 3D convolution with respect to the coordinates $\boldsymbol{x}_s$, $o(\boldsymbol{x}_s)$ is the 3D distribution of fluorescence emitters, $h(\boldsymbol{x}_s|\boldsymbol{x}_d)$ is the set of 3D PSFs for each detector element, and $i(\boldsymbol{x}_s|\boldsymbol{x}_d)$ is the set of 3D images composing the ISM dataset. Evaluating the above equation at $z_s = 0$ we obtain the forward model for single-plane imaging

$$i(\boldsymbol{x}_s|\boldsymbol{x}_d) = \int o(\boldsymbol{x}_s, z) * h(\boldsymbol{x}_s, z|\boldsymbol{x}_d)\,\mathrm{d}z \tag{3.2}$$

where we redefined $\boldsymbol{x}_s = (x_s, y_s)$ and $*$ as the 2D convolution operator. The PSFs – and, correspondingly, the fingerprint – evolve along the axial direction on a scale of the order of the Depth of Field (DOF). However, emitters that are more out-of-focus than a few DOFs are so weak that they provide a negligible contribution. Therefore, we discretize the integral along the axial coordinate, assuming that the detected light stems from a finite number N of planes

$$i(\boldsymbol{x}_s|\boldsymbol{x}_d) = \sum_{k=1}^{N} o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d) \tag{3.3}$$

where k indexes the axial position.

We consider only two axial planes ($N = 2$) to account for the in-focus and out-of-focus sections of the sample. Our goal is to estimate the distribution of emitters located at the focal plane. To this end, we infer the full vector of planar distributions $\mathbf{o} = (o_1, o_2)$. According to Bayes' theorem, the posterior probability is

$$P[\mathbf{o}(\boldsymbol{x}_s)|i(\boldsymbol{x}_s|\boldsymbol{x}_d)] = \frac{P[i(\boldsymbol{x}_s|\boldsymbol{x}_d)|\mathbf{o}(\boldsymbol{x}_s)]P[\mathbf{o}(\boldsymbol{x}_s)]}{P[i(\boldsymbol{x}_s|\boldsymbol{x}_d)]} \tag{3.4}$$

An unbiased estimator of $\mathbf{o}(\boldsymbol{x}_s)$ is found by maximizing the posterior probability. Since we have no prior information on the specimen, the aforementioned task is equivalent to maximizing the likelihood probability. In this work, we used a single-photon avalanche diode (SPAD) array detector with no read-out noise, low dark count rates, and negligible cross-talk. Then, we assume that the signal is corrupted only by shot noise [21]. Therefore, we can see the photon counts for each pixel as random variables following a Poisson distribution. Neglecting cross-talk among

different detector elements, we assume independent realizations of noise. Thus, the likelihood is the product of individual probabilities for each scan and detector point

$$P[i(\boldsymbol{x}_s|\boldsymbol{x}_d)|\mathbf{o}(\boldsymbol{x}_s)] = \prod_{x_d}\prod_{x_s} \frac{\sum_k [o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)]^{i(\boldsymbol{x}_s|\boldsymbol{x}_d)} e^{-\sum_k [o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)]}}{i(\boldsymbol{x}_s|\boldsymbol{x}_d)!} \tag{3.5}$$

The corresponding negative log-likelihood functional is

$$\mathcal{L}[\mathbf{o}(\boldsymbol{x}_s)] = -\log\{P[i(\boldsymbol{x}_s|\boldsymbol{x}_d)|\mathbf{o}(\boldsymbol{x}_s)]\} = \int l[\mathbf{o}(\boldsymbol{x}_s)|\boldsymbol{x}_d]\,\mathrm{d}\boldsymbol{x}_d \tag{3.6}$$

where we defined

$$l[\mathbf{o}(\boldsymbol{x}_s)|\boldsymbol{x}_d] = \int \sum_k o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d) - i(\boldsymbol{x}_s|\boldsymbol{x}_d) \cdot \log\left\{\sum_k o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)\right\} \mathrm{d}\boldsymbol{x}_s \tag{3.7}$$

and discarded the constant terms.

The solution to our inverse problem is given by the vector $\hat{\boldsymbol{o}}$ that minimizes the log-likelihood

$$\hat{\boldsymbol{o}}(\boldsymbol{x}_s) = \underset{\boldsymbol{o}(\boldsymbol{x}_s)}{\arg\min}\, \mathcal{L}[\mathbf{o}(\boldsymbol{x}_s)] \tag{3.8}$$

Let us minimize this functional by imposing the object gradient to zero. Slightly below we will show that the Hessian is non-strictly positive to achieve the convexity of the operator $\mathcal{L}$ and so the uniqueness of the minimizer solution.

Here we present the calculations of the gradient for a two-axial plane object ($N = 2$). Supposing good smoothness of the integrand functions and exploiting the linearity of the functional derivative, we have that

$$\begin{aligned}
&\frac{\delta}{\delta o_1}\int \sum_k o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d) d\boldsymbol{x}_s = \frac{\delta}{\delta o_1}\int o_1(\boldsymbol{x}_s) * h_1(\boldsymbol{x}_s|\boldsymbol{x}_d) + o_2(\boldsymbol{x}_s) * h_2(\boldsymbol{x}_s|\boldsymbol{x}_d)\, d\boldsymbol{x}_s \\
&= \frac{\delta}{\delta o_1}\int o_1(\boldsymbol{x}_s) * h_1(\boldsymbol{x}_s|\boldsymbol{x}_d) d\boldsymbol{x}_s = \frac{\delta}{\delta o_1}\iint o_1(x) h_1(x - \boldsymbol{x}_s|\boldsymbol{x}_d) dx\, d\boldsymbol{x}_s \\
&= \int \frac{\delta}{\delta o_1}\int o_1(x) h_1(x - \boldsymbol{x}_s) dx\, d\boldsymbol{x}_s = \int h_1(x - \boldsymbol{x}_s) 1(x) d\boldsymbol{x}_s \\
&= \int \overline{h_1}(\boldsymbol{x}_s - x) 1(x) d\boldsymbol{x}_s = \overline{h_1} * 1
\end{aligned} \tag{A15}$$

where to compute the last nontrivial equality, we used the fact that the integrand function does not present dependence on derivatives of the object o_1. Repeating the same calculations for the object o_2 holds true that

$$\frac{\delta}{\delta o_2}\int \sum_k o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d) d\boldsymbol{x}_s = \overline{h_2} * 1 \tag{A16}$$

The derivative of the second addend of the functional $\mathcal{L}$ follows the previously presented path of calculation, but taking into account the derivative of the logarithm, and so the associated chain rule of derivation. The minimum is found by setting to zero the functional derivative of the log-likelihood. Using the fact that the adjoint operator of the convolution is the convolution with the flipped kernel, we find

$$\frac{\delta l}{\delta o_j} = h_j(-\boldsymbol{x}_s, \boldsymbol{x}_d) * \left[1 - \frac{i(\boldsymbol{x}_s|\boldsymbol{x}_d)}{\sum_k [o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)]}\right] \tag{3.9}$$

Returning to the log-likelihood, for $j = 1, 2$ holds true:

$$\frac{\delta \mathcal{L}}{\delta o_j} = \int h_j(-\boldsymbol{x}_s, \boldsymbol{x}_d) * \left[1 - \frac{i(\boldsymbol{x}_s|\boldsymbol{x}_d)}{\sum_{k=1}^{2} [o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)]}\right] d\boldsymbol{x}_d =$$

$$= \iint h_j(-\boldsymbol{x}_s|\boldsymbol{x}_d) d\boldsymbol{x}_s d\boldsymbol{x}_d - \int h_j(-\boldsymbol{x}_s|\boldsymbol{x}_d) * \frac{i(\boldsymbol{x}_s|\boldsymbol{x}_d)}{\sum_{k=1}^{2} [o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)]} d\boldsymbol{x}_d$$

Thus, we impose the PSFs to be normalized as follows:

$$\iint h_j(\boldsymbol{x}_s|\boldsymbol{x}_d)\, \mathrm{d}\boldsymbol{x}_s\, \mathrm{d}\boldsymbol{x}_d = 1 \qquad \forall j \in \{1, \ldots, N\} \tag{3.10}$$

While other normalization choices could be made, the one above preserves the fingerprint information and allows for interpreting the PSFs as probability distributions. Therefore, we obtain

$$\frac{\delta \mathcal{L}}{\delta o_j} = 1 - \int h_j(-\boldsymbol{x}_s|\boldsymbol{x}_d) * \frac{i(\boldsymbol{x}_s|\boldsymbol{x}_d)}{\sum_k [o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)]}\, \mathrm{d}\boldsymbol{x}_d \tag{3.11}$$

We minimize the log-likelihood with an iterative gradient descent method

$$o_j^{(m+1)} = o_j^{(m)} - \gamma_k^{(m)} o_j^{(m)} \frac{\delta \mathcal{L}}{\delta o_j} \tag{3.12}$$

where m is the iteration index and $\gamma_j^{(m)}$ is the step of the descent. Choosing $\gamma_j^{(m)} = 1\ \forall\ j, m$, we obtain a multiplicative iteration rule

$$o_j^{(m+1)}(\boldsymbol{x}_s) = o_j^{(m)}(\boldsymbol{x}_s) \int h_j(-\boldsymbol{x}_s|\boldsymbol{x}_d) * \frac{i(\boldsymbol{x}_s|\boldsymbol{x}_d)}{\sum_k o_k^{(m)}(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)}\, \mathrm{d}\boldsymbol{x}_d \tag{3.13}$$

where all the planes are updated at the same iteration. As a consequence of the multiplicative structure of the above equation, the solution is constrained to have non-negative values if initialized with a positive starting point.

The Hessian matrix of the log-likelihood is the following

$$\frac{\delta^2 \mathcal{L}}{\delta o_i \delta o_j} = \int h_i(-\boldsymbol{x}_s, \boldsymbol{x}_d) * \frac{i(\boldsymbol{x}_s|\boldsymbol{x}_d)}{[\sum_k o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)]^2} * h_j(\boldsymbol{x}_s|\boldsymbol{x}_d)\, \mathrm{d}\boldsymbol{x}_d \tag{3.14}$$

Since we work with light intensities (photon counts), all the quantities in the above equation are non-negative. Therefore, the Hessian matrix is positive and the likelihood functional convex. As a result, the algorithm is guaranteed to converge to a unique solution.

Finally, the s²ISM algorithm is completed by choosing $N = 2$ with the two sets of PSFs evaluated at the focal plane and an out-of-focus plane. We chose the starting point $\boldsymbol{o}^{(0)}$ as constant and strictly positive, whose integral equals that of the raw dataset to be reconstructed.

Under the aforementioned assumptions, we can write an explicit Poisson likelihood functional (App. Note 3), whose maximization leads to the following iterative solution

$$\begin{aligned} o_1^{(m+1)}(\boldsymbol{x}_s) &= o_1^{(m)}(\boldsymbol{x}_s) \int h_1(-\boldsymbol{x}_s|\boldsymbol{x}_d) * \frac{i(\boldsymbol{x}_s|\boldsymbol{x}_d)}{\sum_{k=1}^{2} o_k^{(m)}(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)} \,\mathrm{d}\boldsymbol{x}_d \\ o_2^{(m+1)}(\boldsymbol{x}_s) &= o_2^{(m)}(\boldsymbol{x}_s) \int h_2(-\boldsymbol{x}_s|\boldsymbol{x}_d) * \frac{i(\boldsymbol{x}_s|\boldsymbol{x}_d)}{\sum_{k=1}^{2} o_k^{(m)}(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)} \,\mathrm{d}\boldsymbol{x}_d \end{aligned} \tag{3.15}$$

where m and k are the iteration and axial index, respectively. Namely, one estimated image contains the projection of the out-of-focus light and is discarded. The other contains only the signal originating from the focal plane and is built by fusing and deconvolving the twenty-five images of the ISM dataset. Indeed, the proposed algorithm can be considered a generalization of the multi-image deconvolution method [22], which can be recovered simply by putting $N = 1$ in Eq. 3.3 (Fig.3.1).

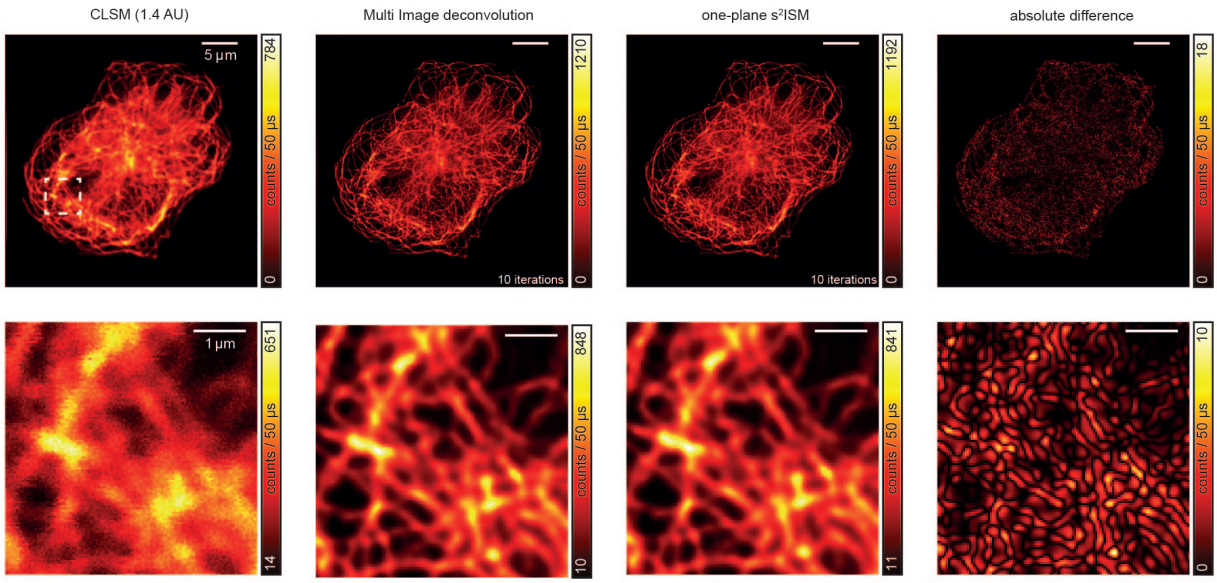

Figure 3.1: **Comparison between multi-Image deconvolution and single-plane s²ISM** The s²ISM algorithm reconstructing a single plane is equivalent to multi-image deconvolution. We quantified the absolute difference of the reconstructions, which is due to numerical errors. Pixel size: $40\,\mathrm{nm}$. Format of the images: 875×875 pixels. Iterations: 10 for both algorithms.

Furthermore, dropping the detector dimension $\boldsymbol{x}_d$ also recovers the conventional Richardson-Lucy deconvolution equation [23, 24]. The result of s²ISM is an image with enhanced resolution and optical sectioning compared to its confocal counterpart. Importantly, we achieve both benefits without discarding in-focus light, enhancing the SNR [25].

3.2 Data-driven PSFs estimation

To apply the proposed algorithm (Eq. 3.15), a set of PSFs for each axial plane involved in the reconstruction is required. This is straightforward for *in silico* experiments, allowing easy validation of the method. We first tested s²ISM in a simplified scenario where the image contains fluorescence from a single defocused plane at a known axial position. Using vectorial diffraction theory, we simulated the PSFs at the relevant axial positions and applied iterative reconstruction. s²ISM effectively removes out-of-focus light, producing images with enhanced resolution and optical sectioning.

In realistic samples, out-of-focus light arises from multiple planes. While Eq. 2.15 can be generalized to include N axial contributions, doing so increases computational cost and worsens the ill-posed nature of the problem. Using only two axial planes effectively asks the algorithm whether detected light originates from the in-focus or out-of-focus plane. This approach simplifies the inverse problem, improves conditioning, and enhances optical sectioning, regardless of sample thickness.

In an ideal, aberration-free optical setup, the PSFs are z-independent. Consequently, s²ISM is insensitive to the symmetric placement of the background plane relative to the focal plane. Symmetrically positioning two reconstruction planes would generate identical out-of-focus reconstructions, adding computational complexity without benefit. As s²ISM addresses an ill-posed inverse problem, more prior knowledge improves performance. For instance, for a single pixel (a scanning coordinate in the raw ISM dataset), s²ISM distributes photons between in-focus and out-of-focus planes based on a linear combination of their PSFs.

When reconstruction planes are separated by distances comparable to the microscope's depth of field (DOF), their PSFs are distinct, facilitating accurate classification. Conversely, using multiple planes within the DOF produces highly similar PSFs, increasing misclassifications and degrading the in-focus reconstruction (Fig. 3.2). Therefore, in the ideal case, two reconstruction planes along the z-axis are sufficient.

The axial position of the second plane, namely the out-of-focus plane, is not well-defined, so we select it to maximize PSF diversity relative to the focal plane. Similarity between PSFs at different axial positions is measured using Pearson correlation or Kullback-Leibler divergence, choosing the out-of-focus plane as the minimum or maximum, respectively. Both metrics yield the same position (Fig. 3.3).

Using this approach, s²ISM was successfully applied to simulated data.

We first applied it plane-by-plane to a 3D phantom dataset of a point-like source (Fig. 3.4b). The results show that the algorithm effectively rejects defocused light while preserving 100 of the in-focus signal within a few iterations. Additionally, it performs joint deconvolution across planes within the focal volume, fusing the raw ISM images into a single image with enhanced resolution

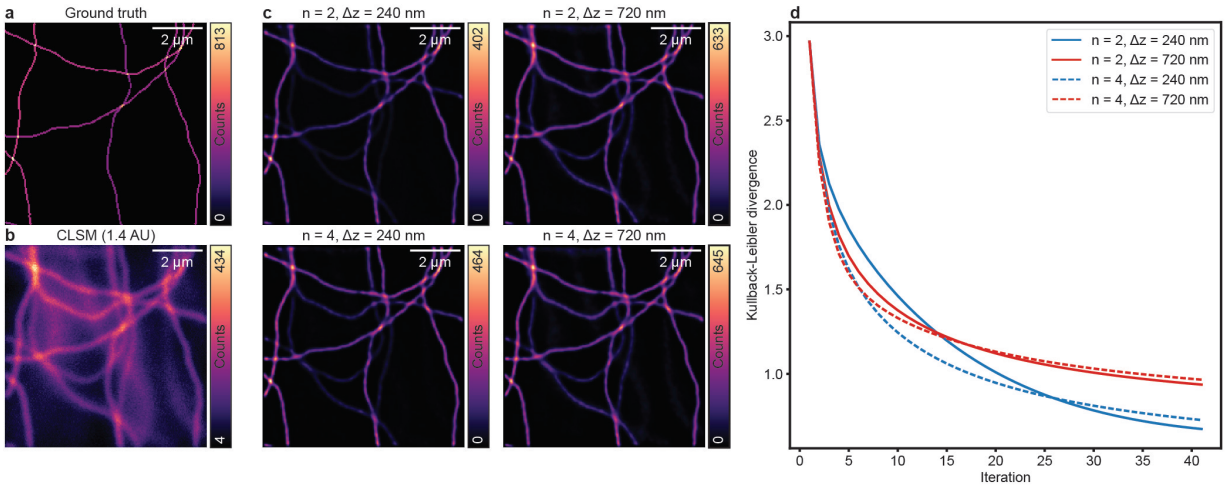

Figure 3.2: **Effect of the number of axial planes.** Simulation of an ISM dataset generated by a thick sample (9 planes axially separated by $240\,\mathrm{nm}$) reconstructed using s²ISM. **a**, ground-truth section at $z = 0\,\mathrm{nm}$. **b**, confocal image generated by summing ISM dataset images. **c**, s²ISM reconstructions using two (top) or four (bottom) axial planes separated by $\Delta z = 240\,\mathrm{nm}$ (left) or $\Delta z = 720\,\mathrm{nm}$ (right). Reconstruction stopped at 20 iterations. **d**, Kullback-Leibler divergence of images in **c** against the ground truth (**a**). More than a single out-of-focus plane increases convergence rate but worsens ill-posedness. After sufficient iterations, the two-plane approach yields better reconstruction, demonstrating robustness.

and optical sectioning.

We then applied the same analysis to an experimental PSF—a 3D stack of sub-diffraction gold bead images (Fig. 3.4a).

Results are qualitatively similar to the simulated case: s²ISM rejects 80 % of out-of-focus light in fewer than five iterations. As convergence is rapid, we fixed the iteration number to 10 and investigated the effect of the out-of-focus plane choice in alternative methods, selecting the out-of-focus plane while optimizing other physical parameters (we describe in depth in App. B.5). Reconstructions of the simulated PSF stack with varying defocus positions of the background PSFs (Fig. 3.7) show that correctly choosing the out-of-focus plane allows near-complete rejection of defocused light while preserving 100 % of in-focus signal. Underestimating this position reduces in-focus preservation and optical sectioning, whereas overestimation mainly reduces optical sectioning without significant in-focus loss.

Experimental datasets replicate these trends (Fig. 3.8) with minor differences: in-focus signal saturates at 80 %, and optical sectioning is less sensitive to the exact out-of-focus position. These differences arise from PSF mismatches caused by slight spherical aberrations, yet s²ISM still effectively rejects nearly all defocused light.

This analysis also clarifies the definition of the out-of-focus plane: we look for the axial position that maximizes both optical sectioning and in-focus preservation. While rigorous determination is computationally expensive, a convenient approximation is half of the axial diffraction limit of an ideal confocal microscope. Overestimating this position has minimal impact on in-focus preservation and only slight effects on optical sectioning (Figs. 3.6,3.8f).

In practice, direct access to experimental PSFs is challenging, and reconstruction quality de-

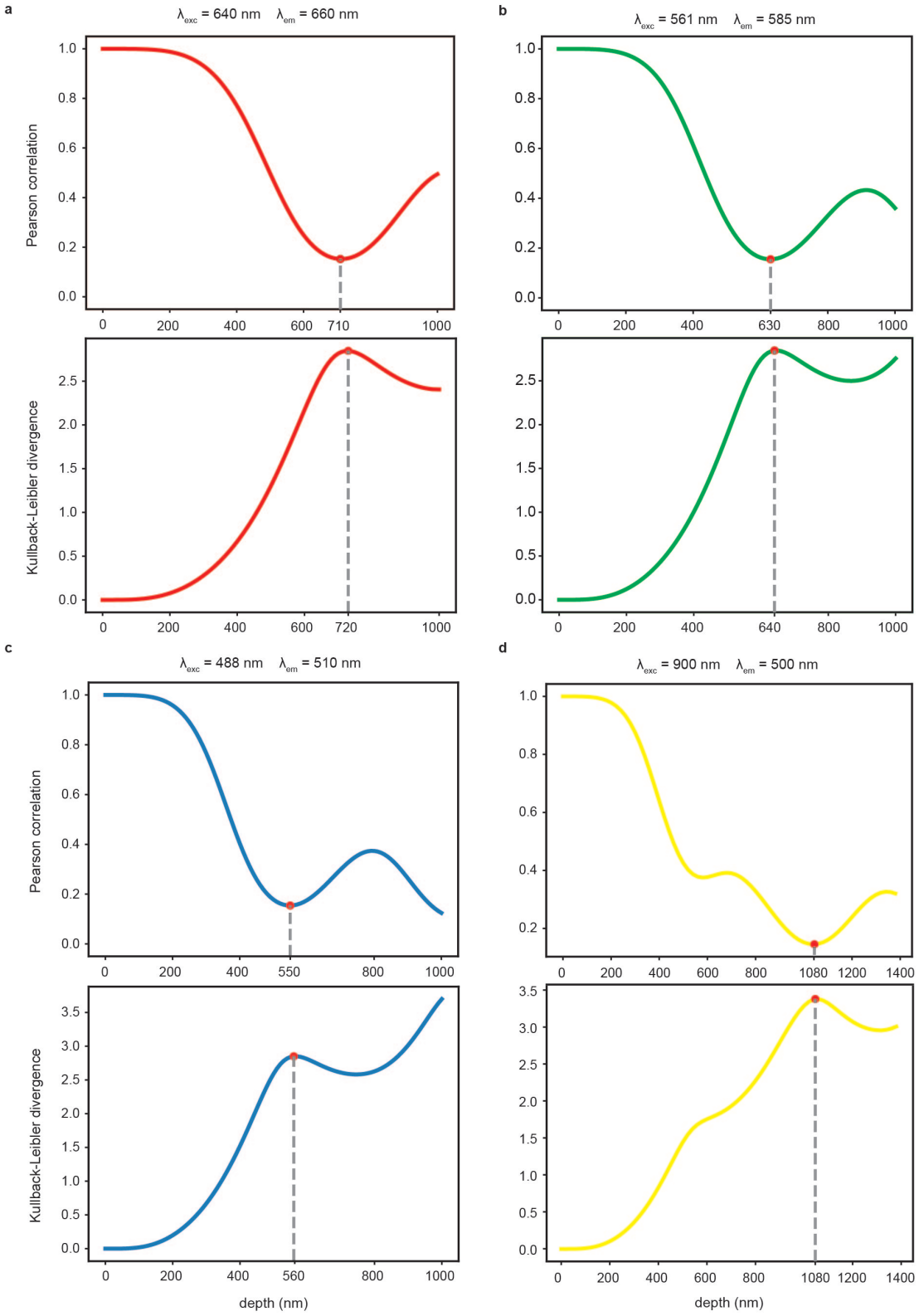

Figure 3.3: **Calculation of the out-of-focus position.** Optimal out-of-focus depth estimated for red **a**, green **b**, blue **c**, and two-photon **d** excitation. Pearson correlation and Kullback-Leibler divergence between in-focus and defocused PSFs are reported.

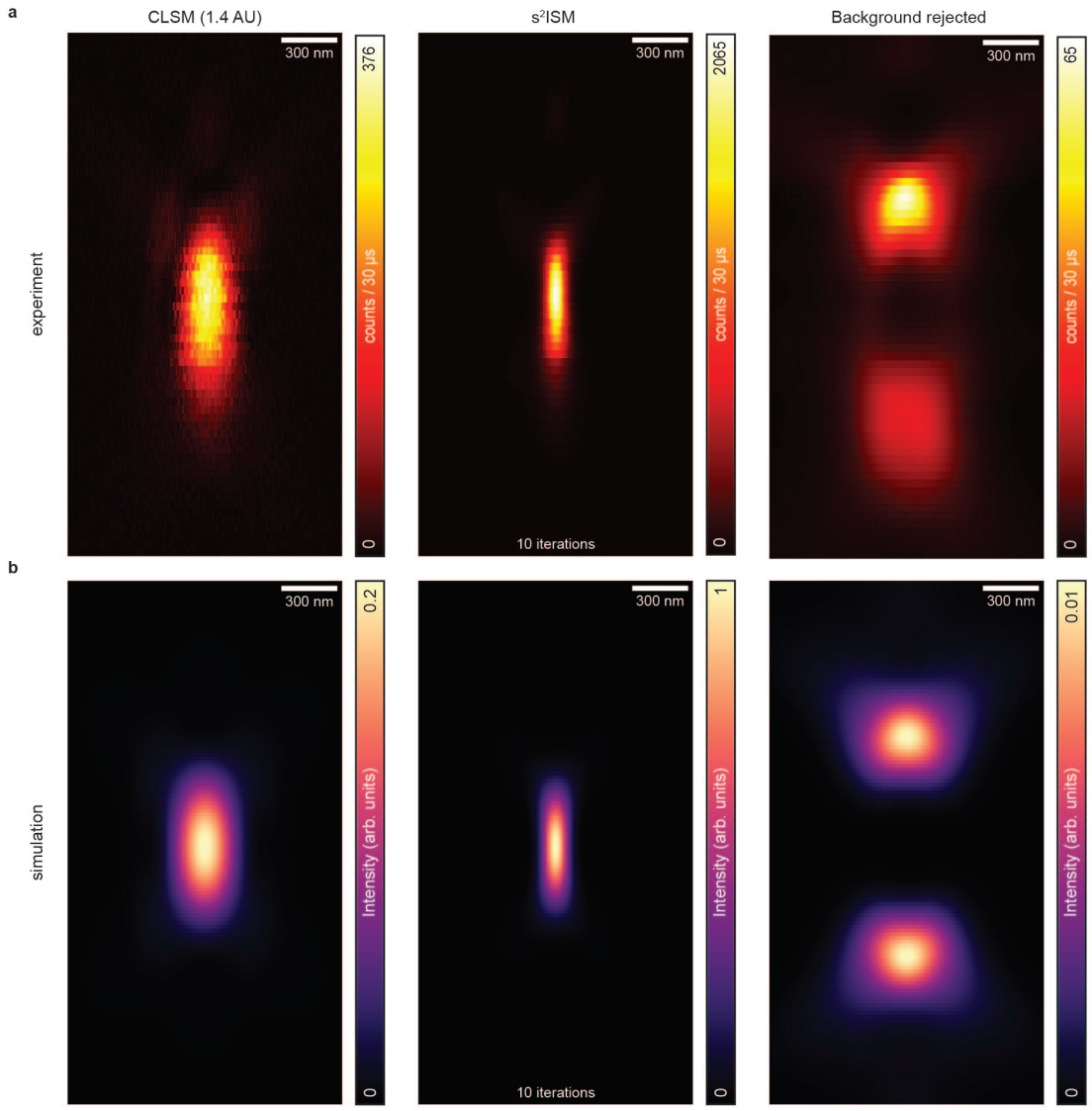

Figure 3.4: Reconstruction with s^{2}ISM using synthetic (**b**) simulated in the red excitation regime, and experimental sub-diffraction gold beads (**a**). Both datasets span the *xyz* space. We apply the s^{2}ISM algorithm for 10 iterations, for each z-plane. In the experimental scenario, some spherical aberration can be observed in the xz profile. However, s^{2}ISM effectively suppresses out-of-focus light at each lateral plane. By inspecting the intensity bar in the synthetic scenario, we can see that the normalized photon counts indicate a higher number of photons in the in-focus s^{2}ISM reconstruction compared to the raw dataset, even if flux conservation still holds true, and some photons have been discarded in the out-of-focus reconstruction. That is the deconvolution effect; reassigning the photons during the iterative process can lead to a higher PSNR in the in-focus reconstruction. Indeed, for 10 iterations of reconstruction, this behavior is also happening in the experimental scenario.

pends on PSF accuracy, which may be degraded by noise, misalignment, or other imperfections. We therefore exploit information encoded in the ISM dataset to estimate PSF parameters reliably. In particular, PSFs corresponding to the inner detector elements (3×3) are nearly identical in shape but spatially shifted. These shift vectors, related to the array geometry and microscope magnification, can be measured without direct PSF imaging.

Through a tailored data-driven optimization framework, we identify and estimate a set of parameters that are both necessary and sufficient for the reliable reconstruction of the system's PSFs. The first two parameters are the rotation and orientation of the single detector pixels.

First, we calculate the phase correlation among the images of the experimental ISM dataset

$$\mathcal{R}(\boldsymbol{x}_s|\boldsymbol{x}_d) = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{i(\boldsymbol{x}_s|\boldsymbol{x}_d)\}\overline{\mathcal{F}\{i(\boldsymbol{x}_s|\boldsymbol{0})\}}}{|\mathcal{F}\{i(\boldsymbol{x}_s|\boldsymbol{x}_d)\}\overline{\mathcal{F}\{i(\boldsymbol{x}_s|\boldsymbol{0})\}}|}\right\} \tag{3.16}$$

where $\mathcal{F}$ is the Fourier transform operator, and the overline stands for complex conjugate. Then, we find the shift vectors as the position of the maximum of each correlogram

$$\boldsymbol{\mu}^{(\text{exp})}(\boldsymbol{x}_d) = \arg\max_{\boldsymbol{x}_s} \mathcal{R}(\boldsymbol{x}_s|\boldsymbol{x}_d) \tag{3.17}$$

Since the images with the highest SNR are those close to the centre of the detector array, we select the shift-vectors from the inner 3×3 array. Using the discretized detector coordinates, we have

$$\left(\frac{x_d}{\Delta x_d}, \frac{y_d}{\Delta y_d}\right) = (m_x, m_y) = \boldsymbol{m} \in \{-1, 0, 1\}^2 \tag{3.18}$$

We flatten the detector dimension into a single index $j \in [1, 9]$ and indicate the position coordinates using the index $i \in [1, 2]$. Thus, the detector coordinates m_{ij} and experimental shift-vectors $\mu_{ij}^{(\text{exp})}$ are represented as 2×9 (Fig. 3.5a).

Finally, we assume that the transformation from the detector coordinates to the experimental shift-vectors is given by three operations: mirroring, rotation, and dilation. These transformations are described by the matrix

$$\boldsymbol{T}(\rho, \theta, \alpha) = \boldsymbol{A}(\alpha)\boldsymbol{R}(\theta)\boldsymbol{M}(\rho) \tag{3.19}$$

The mirroring matrix $\boldsymbol{M}(\rho)$ accounts for the orientation of the detector

$$\boldsymbol{M}(\rho) = \begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix} \quad \text{with } \rho \in \{-1, 1\}; \tag{3.20}$$

The rotation matrix $\boldsymbol{R}(\theta)$ accounts for the rotation of the detector

$$\boldsymbol{R}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \quad \text{with } \theta \in [-\pi, \pi]; \tag{3.21}$$

The dilatation matrix $\boldsymbol{A}(\alpha)$ accounts for the magnification of the microscope

$$\boldsymbol{A}(\alpha) = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \quad \text{with } \alpha \in \mathbb{R}^+; \tag{3.22}$$

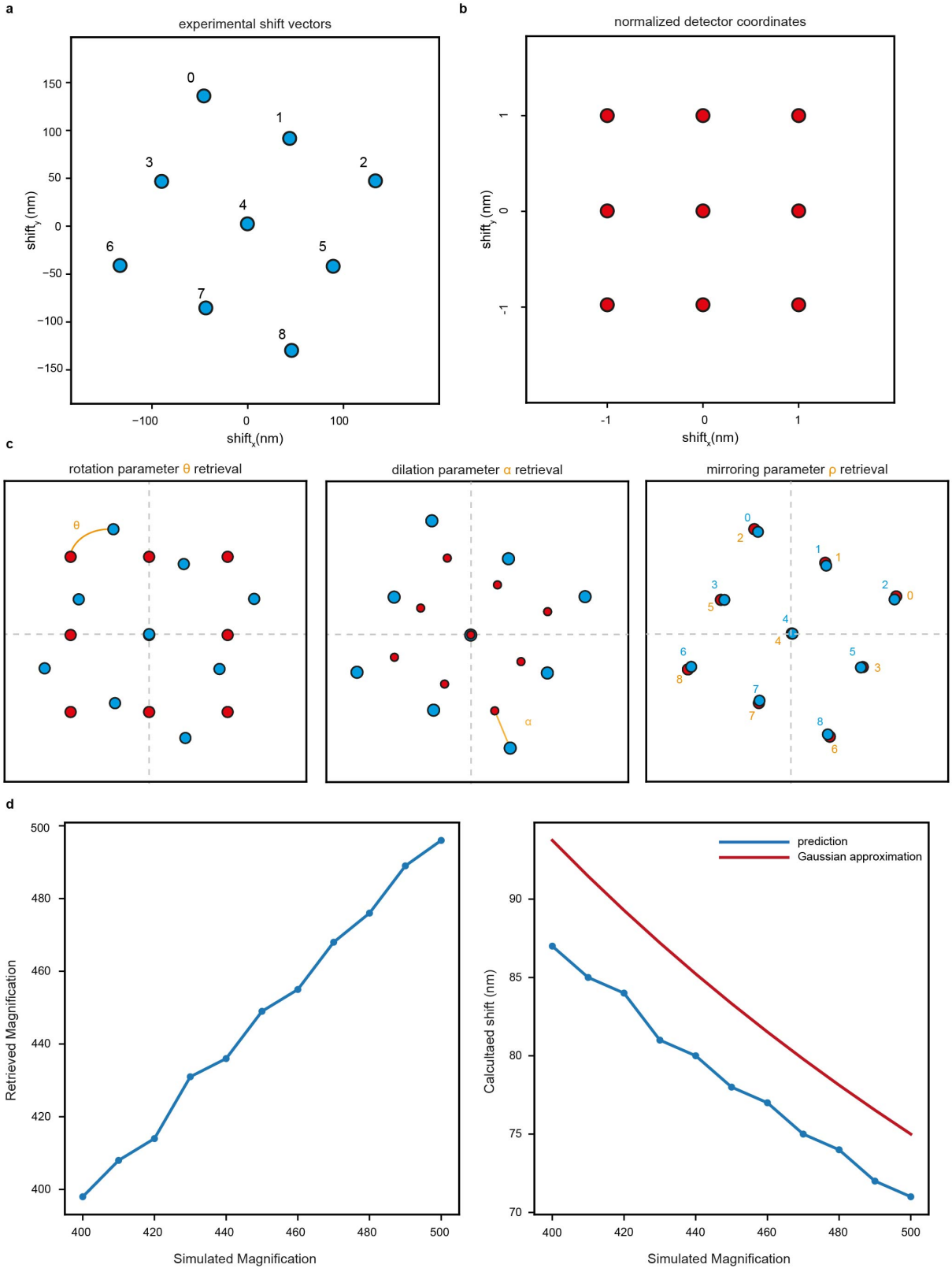

Figure 3.5: **Workflow of parameter estimation.** We fit the experimental shift-vectors from the 3×3 inner array, **a**, to the normalized detector coordinates, **b**, transformed by a rotation, dilation, and mirroring operations, **c**. Finally, we convert the dilation parameter into a magnification value, minimizing the difference with the shift-vectors value calculated using the scalar approximation of the PSF, **d**. Comparing the values calculated using the vectorial model, we show that the error is negligible (left). Then, we invert the model to retrieve the magnification (right). The simple law derived under the Gaussian approximation of PSF is not accurate enough to be used.

We show a graphical depiction of the presented method in Fig. 3.5. We find the parameters $\hat{\rho}$, $\hat{\theta}$, and $\hat{\alpha}$ that describe the microscope used for the acquisition of the dataset by numerically solving the following minimization problem

$$\hat{\rho}, \hat{\theta}, \hat{\alpha} = \underset{\rho,\theta,\alpha}{\arg\min} \left\| \boldsymbol{\mu}^{(\text{exp})} - \boldsymbol{T}(\gamma, \theta, \alpha)\boldsymbol{m} \right\|_F^2 \tag{3.23}$$

where we used the Frobenius norm. While the loss function is not convex in general, the constraints on the minimization values uniquely define a single solution.

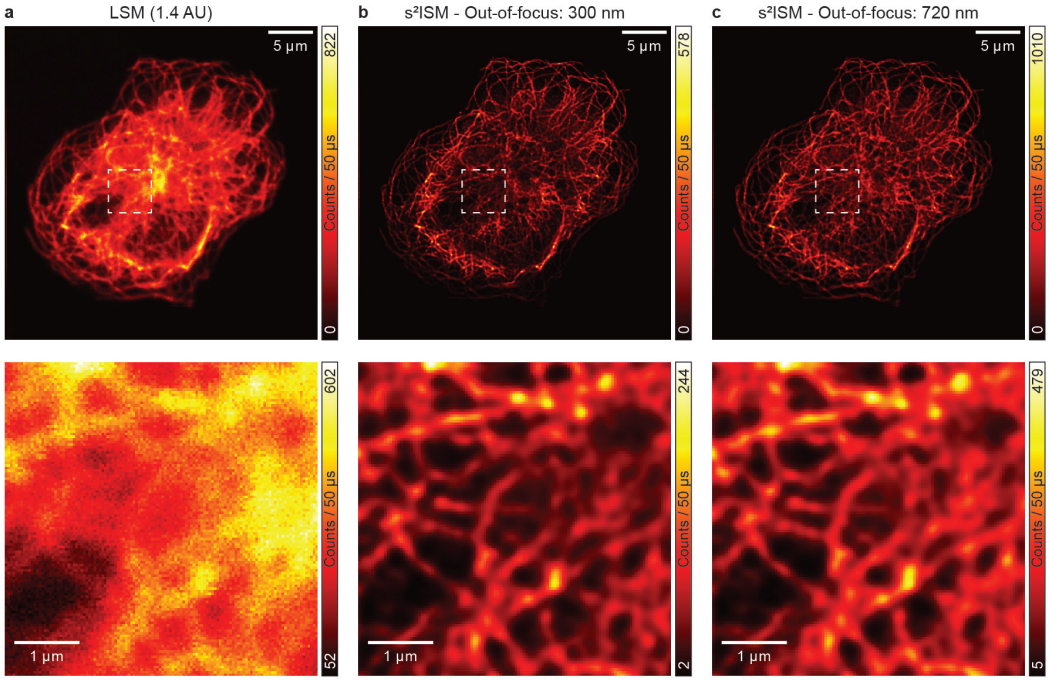

Figure 3.6: **Reconstruction with different out-of-focus positions.** Experimental ISM reconstruction. **a**, confocal image from summed raw images. **b**, s²ISM reconstruction at 20 iterations with out-of-focus at $300\,\text{nm}$. **c**, reconstruction at 20 iterations with out-of-focus at $720\,\text{nm}$. The reconstructions are qualitatively similar, demonstrating robustness to defocus parameter choice.

The third parameter of the set is the dilation parameter $\hat{\alpha}$ is the absolute value of the shift-vectors for a detector element located at one pixel pitch from the centre of the detector. Approximating the PSFs as Gaussian functions, we can calculate the shift-vectors as $\mu(M) = \frac{\Lambda}{2M}$, where Λ is the pixel pitch of the detector and M the magnification of the microscope. However, the Gaussian approximation is too crude and would lead to a poor estimation of the magnification. Therefore, we used the more accurate scalar model, which requires also the following variables: the pixel size of the detector, the wavelength of the excitation and fluorescence light, and the numerical aperture of the objective lens. These latter are typically known from the experimental setup and can be used to convert $\hat{\alpha}$ into a magnification value M. To speed up the calculation,

we performed a 1D calculation of the following scalar model of the in-focus PSF

$$h(x|\lambda) = \left| \frac{J_1(k\mathrm{NA}x)}{k\mathrm{NA}x} \right|^2 \tag{3.24}$$

where $k = 2\pi/\lambda$ and J_1 is the first order Bessel function of the first kind. We define the pinhole function as

$$p(x|M) = \begin{cases} 1 & \text{if } |x| \leq \frac{\Delta}{2M} \\ 0 & \text{otherwise} \end{cases} \tag{3.25}$$

where Δ is the pixel size of the detector. We calculate the theoretical shift as

$$\mu(M) = \arg\max_x h(x|\lambda_{\mathrm{exc}})[h(x|\lambda_{\mathrm{em}}) * p(x - \Lambda/M|M)] \tag{3.26}$$

Finally, we estimate the magnification by numerically solving the following minimization problem

$$\hat{M} = \arg\min_M \|\hat{\alpha} - \mu(M)\|_2^2 \tag{3.27}$$

As reported in Fig. 3.5d, the above procedure correctly estimates the magnification value. We could also have used a vectorial model to calculate the PSFs. Despite being more rigorous, it is more computationally expensive and leads to a negligible improvement in accuracy in estimating the magnification.

We highlight the fact that the detector rotation and element orientation can be determined for any array geometry (Fig. 6.2). Assuming negligible aberrations, these parameters allow simulation of PSFs with a vectorial diffraction model [26, 27], fully capturing the high-NA system. Imaging of a tubulin network demonstrates effective background removal and resolution enhancement using this approach (Fig. C.3).

In the final step, we estimate the minimal pixel support necessary to generate synthetic PSFs that fully fit within the simulation field of view (FOV). This ensures that all physically relevant features of the PSFs are contained within the simulation box. Consequently, any increase in the simulation domain can be performed via zero-padding, which does not introduce additional physical information, thus providing an efficient optimization of the simulation procedure.

We find the optimal field of view by leveraging the generalized divergence law of Gaussian beams

$$w(z) = w_0 \sqrt{1 + \left(\frac{M^2 \cdot z}{z_R} \right)^2} \tag{3.28}$$

where $w(z)$ is the beam waist and at a distance z from the focus, z_R is the Rayleigh range, and M^2 is a factor to correct for the PSF's non-Gaussian shape. We empirically found the optimal value of this latter to be $M^2 = 3$.

We define the in-focus beam size as the radius of the Airy disk

$$w_0 = 0.61 \frac{\lambda_{\mathrm{exc}}}{\mathrm{NA}} \tag{3.29}$$

and the Rayleigh range as

$$z_R = \frac{\pi w_0^2 n}{\lambda_{\text{exc}}}. \tag{3.30}$$

where n is the refractive index of the immersion medium.

We also need to consider the shift induced by the off-axis detector elements. Therefore, we evaluate the size of the field of view FOV as

$$\mathrm{FOV} = 2[w(z_2) + N_d \cdot \Delta x_d] \tag{3.31}$$

where z_2 is the position of the out-of-focus plane, and N_d is the number of elements per axis of the detector array. Thus, $N_d \cdot \Delta x_d$ is the lateral size of the detector.

Finally, reconstructions using synthetic PSFs were compared with s²ISM applied to experimental PSFs (App. Note A.3). Focal planes were defined by the maximal integral of the modulation transfer function, and out-of-focus planes by maximal discrepancy (App. Fig. A.2). Both approaches yield qualitatively similar reconstructions (App. Fig. A.3), although experimental PSFs produce slightly lower resolution due to noise. Given the ease of simulating PSFs and their reliability, synthetic PSFs were used for all reconstructions in this manuscript unless otherwise specified.

3.3 Validation of s²ISM on synthetic data

The first challenge, now that we have all the ingredients to effectively launch the algorithm, is to demonstrate its reliability as a reconstruction method—specifically, its ability to produce accurate results across a wide range of applications without generating artifacts or non-physical solutions. To this end, we constructed a thick phantom sample and used simulated PSFs to model its passage through the imaging system, thereby creating a synthetic ISM dataset. The ground truth—the actual object, which is inherently unrecoverable in a real scenario—allows us to assess the performance of our s²ISM reconstruction after processing. Within this framework, we quantitatively evaluate the improvements that s²ISM offers compared to state-of-the-art algorithms.

The first validation step is imposing the object to be a point-like source, namely, we assess the performance of s²ISM on simulated ISM PSFs. As depicted in Fig. 3.10, we perform s²ISM reconstruction on both red, green, and blue excitation, eventually evaluating the optical sectioning performed by s²ISM.

Then, we simulated a set of ISM data with both in-focus and out-of-focus spatial structures arranged in a geometrically regular pattern (Fig. 3.9). This design enables precise knowledge of emitter positions within the volume and allows for accurate characterization of the algorithms' optical sectioning capabilities. Additionally, we introduced varying noise levels and increased disk radii into the raw data to explore conditions under which s²ISM 's performance may be compromised due to inversion boundary limitations. Comparing s²ISM results with the state-of-the-art algorithm for non-ISM (App. Note D) and SIM frameworks, like rolling-ball, sectioningMRA, BF-SIM, and focus-ISM, we appreciate unbeaten optical sectioning capabilities while still obtaining

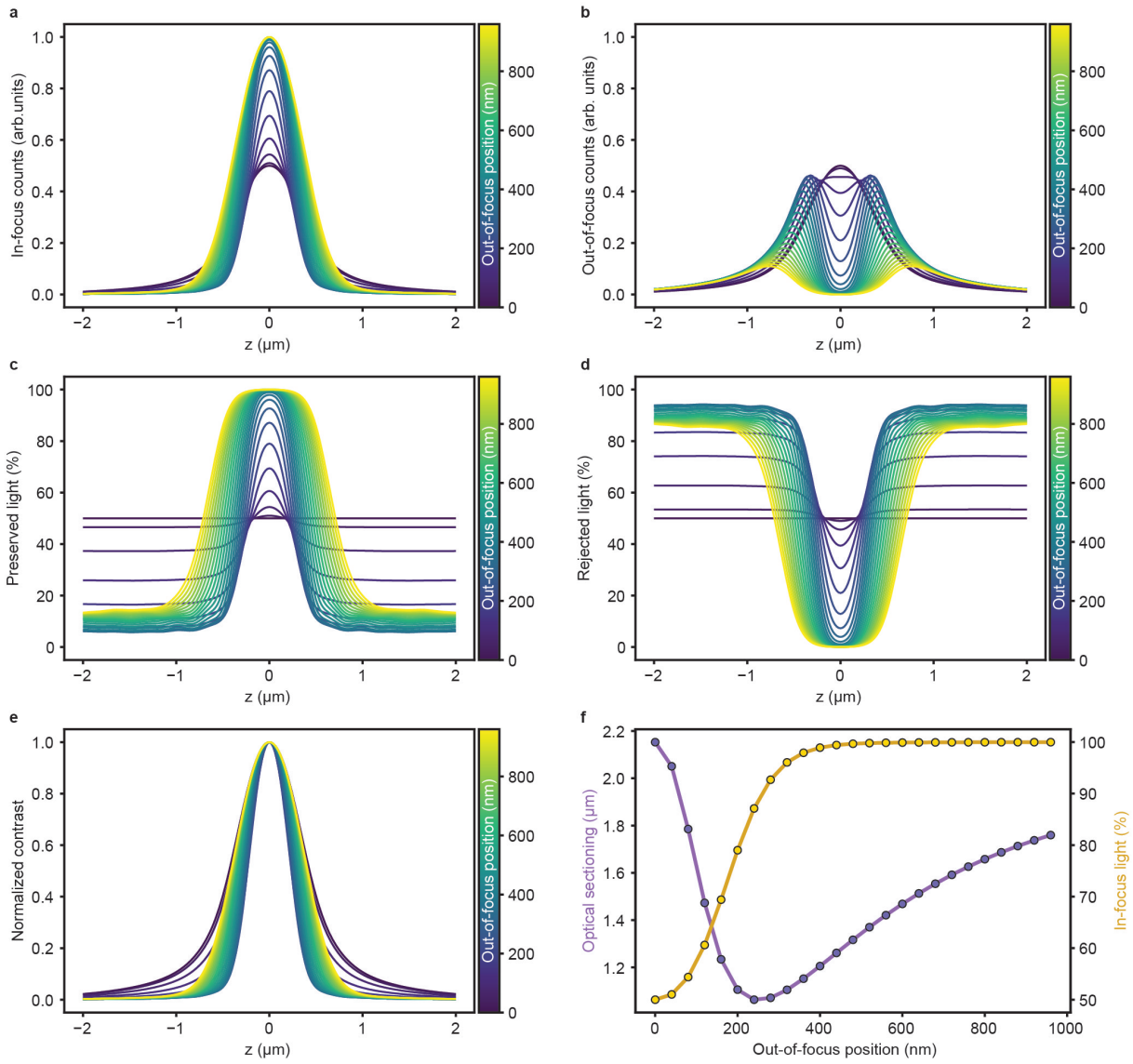

Figure 3.7: **Optical sectioning analysis: simulation** Plane-by-plane s^{2}ISM reconstruction of the simulated PSF dataset shown in Fig. 3.4b, using 10 iterations and varying the value of the out-of-focus position from $0\,\mathrm{nm}$ to $980\,\mathrm{nm}$ with steps of $40\,\mathrm{nm}$. Plot of the total photon counts of the in-focus (**a**) and out-of-focus (**b**) reconstructions, summed over the xy plane and normalized the total photon counts at $z = 0\,\mathrm{nm}$ of the raw data. Plot of the fraction of preserved (**c**) and rejected (**d**) light, obtained by dividing the curves in **a** and **b** by the total amount of photon counts per plane in the raw data. **e**, contrast curves obtained by normalizing the curves in **a** with the corresponding maximum value. **f**, plot of optical sectioning (violet, left axis) and percentage of in-focus light preserved (yellow, right axis). The former is obtained by calculating the integral of the curves in **e**, the latter is calculated as the value of the curves in **c** at $z = 0\,\mathrm{nm}$.

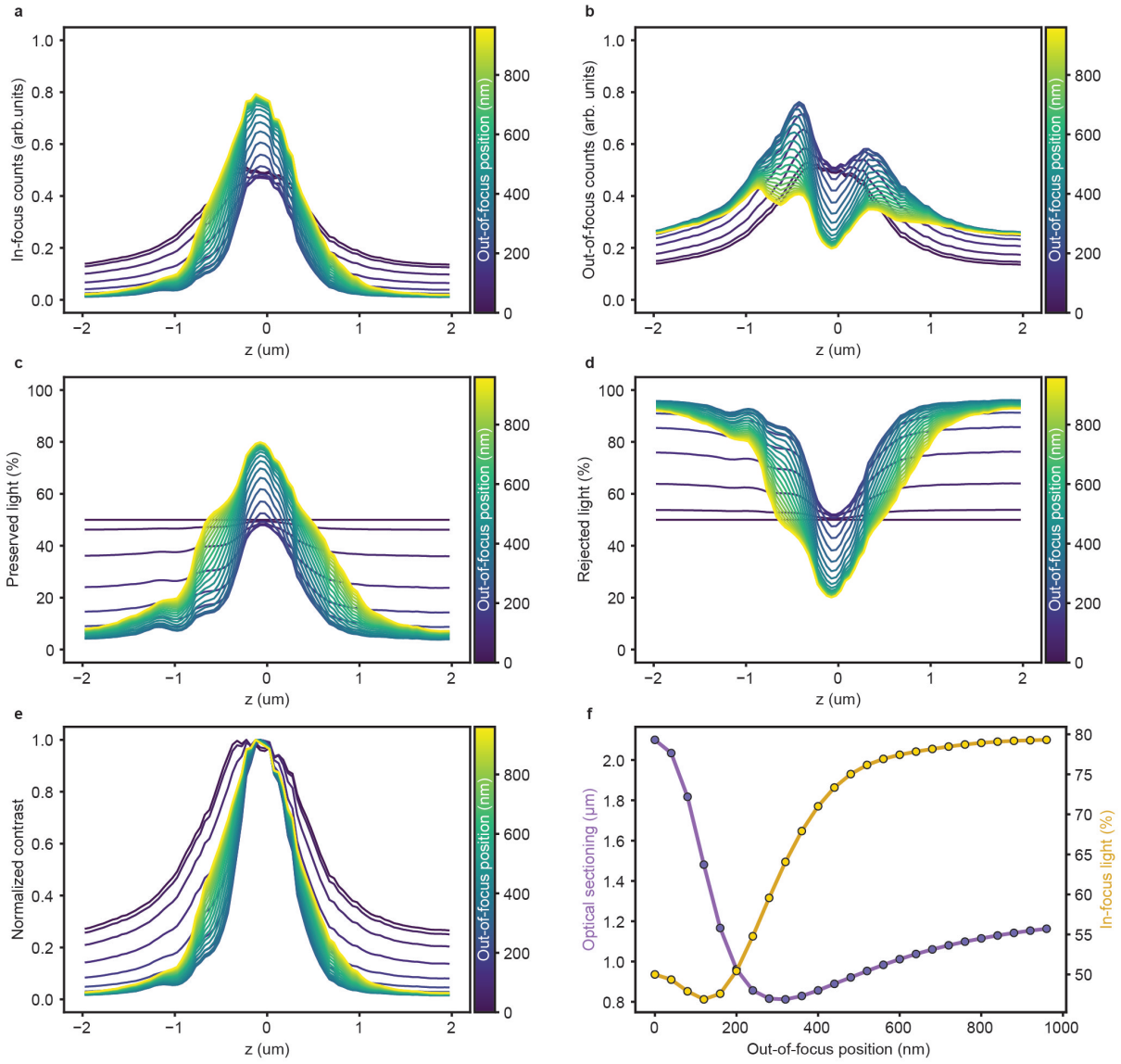

Figure 3.8: **Optical sectioning analysis: experiment** Plane-by-plane s^{2}ISM reconstruction of the raw experimental dataset of a sub-diffraction bead shown in Fig. 3.4a, using 10 iterations and varying the value of the out-of-focus position from $0\,\mathrm{nm}$ to $980\,\mathrm{nm}$ with steps of $40\,\mathrm{nm}$. Plot of the total photon counts of the in-focus (**a**) and out-of-focus (**b**) reconstructions, summed over the xy plane and normalized the the total photon counts of the raw data at the focal plane, found as show in Supp. Fig. A.2a. Plot of the fraction of preserved (**c**) and rejected (**d**) light, obtained by dividing the curves in **a** and **b** by the total amount of photon counts per plane in the raw data. **e**, contrast curves obtained by normalizing the curves in **a** with the corresponding maximum value. **f**, plot of optical sectioning (violet, left axis) and percentage of in-focus light preserved (yellow, right axis). The former is obtained by calculating the integral of the curves in **e**, the latter is calculated as the maximum value of the curves in **c**.

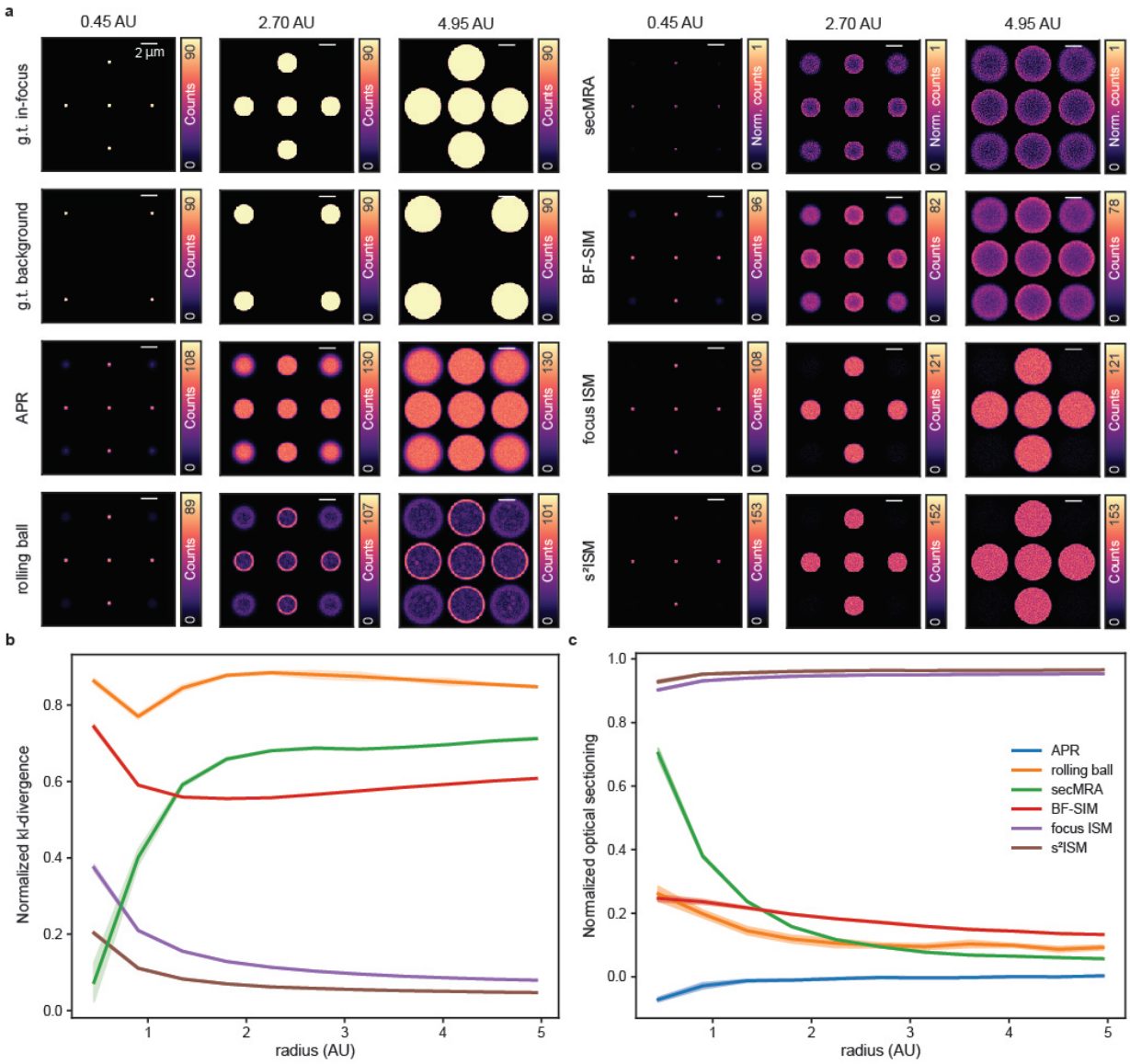

Figure 3.9: **Comparison of background removal algorithms on simulated data.** Simulation of in-focus and out-of-focus ($z = 550\,\mathrm{nm}$) identical disks with increasing radius, reported in Airy units (AU). We used $\lambda_{\mathrm{exc}} = 488\,\mathrm{nm}$, $\lambda_{\mathrm{det}} = 510\,\mathrm{nm}$, $\mathrm{NA} = 1.4$, and $M = 450$. **a**, Ground truths and images reconstructed using multiple algorithms: APR, rolling ball (ball radius $= 240\,\mathrm{nm}$), secMRA (automatically optimized parameters), BF-SIM (default parameters), Focus-ISM, and s^{2}ISM (20 iterations). General-purpose algorithms are applied to the APR image. **b**, Kullback Leibler (KL) divergence of the reconstructions to the in-focus ground truth for each radius. **c**, normalized optical sectioning curve, obtained using the expression $1 - 1/\mathrm{SBR}$, where SBR is the signal-to-background ratio. The SBR is calculated as the ratio of the total counts of an in-focus and out-of-focus disk. The curve is bound between 0 (no sectioning) and 1 (perfect sectioning). Dashed areas represent the uncertainty, calculated as the standard deviation of the corresponding curve calculated over multiple noise realizations of the same simulation.

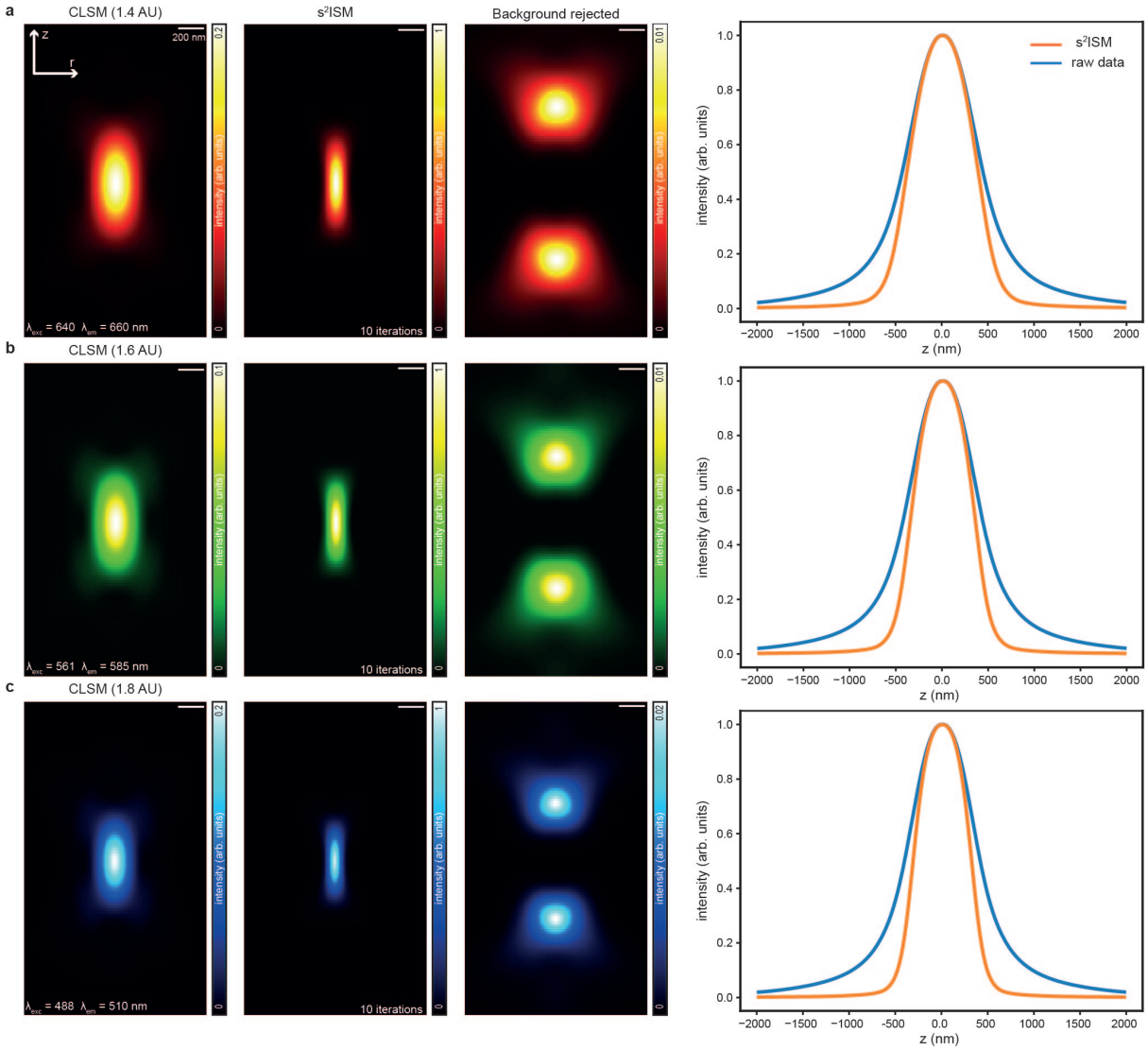

Figure 3.10: **s²ISM reconstruction of synthetic PSFs.** 3D PSFs for red **a**, green **b**, and blue **c** light. We depict the open-pinhole confocal (left) and s²ISM (middle) images. We calculate the integral of the PSF along the lateral dimension to quantify optical sectioning (right). For each color, the PSFs are simulated in a volume of $4.04\,\mu\text{m} \times 4.53\,\mu\text{m} \times 4.53\,\mu\text{m}$ with $202 \times 453 \times 453$ voxels (zxy). We reconstruct the stacks with 10 s²ISM iterations for each axial plane.

the best distribution reconstruction for all disk radii.
The development of an algorithm capable of enhancing any desired benefits for an optical microscope has been possible thanks to the great properties of the SPAD array detector. His pixelated nature inherently encodes the axial position of the fluorescent emitters in the raw dataset. Optical sectioning can be pushed just by leveraging that fact, hidden in the fingerprint, namely a quantity that shows dependence only on the detector coordinates. Old-fashioned algorithms that rely on non-physical assumptions on the dataset will fail while analyzing such types of samples (Fig. 3.9a).

Moreover, even if the algorithm arises from a well-built physical model, if the raw data does not include the axial information in itself, it will fail the reconstruction.

General-purpose background removal algorithms are designed to work on single images. Thus, they cannot distinguish between in-focus and out-of-focus emitters, and they need to rely on prior assumptions of the structure of the specimen. Typically, the out-of-focus image is assumed to be low-intensity and dominated by low spatial frequencies. Indeed, general-purpose algorithms are typically based on high-pass filters. Therefore, they suppress intensity in flat and homogeneous regions of the image. Therefore, they cannot provide true optical sectioning (Fig. 3.9), as demonstrated by the negligible effect of removing out-of-focus light by rolling ball and more advanced algorithms [28, 29]. Furthermore, they might generate structural artefacts trying to match their prior assumptions on the reconstructed image (Fig. C.7).

Here, we show a small proof-of-concept. We simulate a very broad and thick sample (Fig. 3.11). Since objects are much larger than the system's PSF, they carry mainly low-frequency content and cannot be discriminated with a high-pass frequency filter. Then we apply the s²ISM algorithm to the dataset integrated on the detector coordinates (Fig. 3.11a) and on the ISM data (Fig. 3.11b). As we can see, if the algorithm cannot access the detector information, background light cannot be properly discarded from the in-focus reconstruction (Fig. 3.11c). This difference can also be appreciated by looking at the Kullback-Leibler divergences shown in the bottom plots (Fig. 3.11f). Along the iteration routine, the red line, namely the loss function associated with the s²ISM algorithm, continuously decreases, given that out-of-focus light gets more and more discarded from the in-focus plane, the error of the in-focus reconstruction with respect to the respective ground truth decreases as well. While the blue line, associated with the normal Richardson-Lucy algorithm extended on the axial dimension, is failing in classifying the position of the emitters along the z-axis, indeed, the Kullback-Leibler divergence remains constant along the iteration. Thus, our method detects the emitters' axial position without discarding the low-frequency content of the in-focus plane. As discussed at the beginning of this chapter, one of the key strengths of s²ISM lies in its ability to retrieve an optically sectioned reconstruction from a planar acquisition that exhibits no explicit dependence on the axial coordinate. In the following, we compare the performance of s²ISM with that of MID-3D, noting that the availability of prior information, and therefore the conditioning of the inverse problem, is improved in the three-dimensional case.To this end, we simulated a three-dimensional sample of tubulin filaments for a corresponding five-dimensional (3D + 2D) ISM dataset.

Simulation size: $41\,\text{voxels} \times 401\,\text{voxels} \times 401\,\text{voxels}$ with a voxel size of $100\,\text{nm} \times 40\,\text{nm} \times 40\,\text{nm}$ (zxy). Simulation parameters: $\text{NA} = 1.4$, $n = 1.5$, $\lambda_{\text{exc}} = 640\,\text{nm}$, $\lambda_{\text{exc}} = 660\,\text{nm}$. On the latter, we applied 20 iterations of MID-2D, 20 iterations of MID-3D, and 20 iterations of

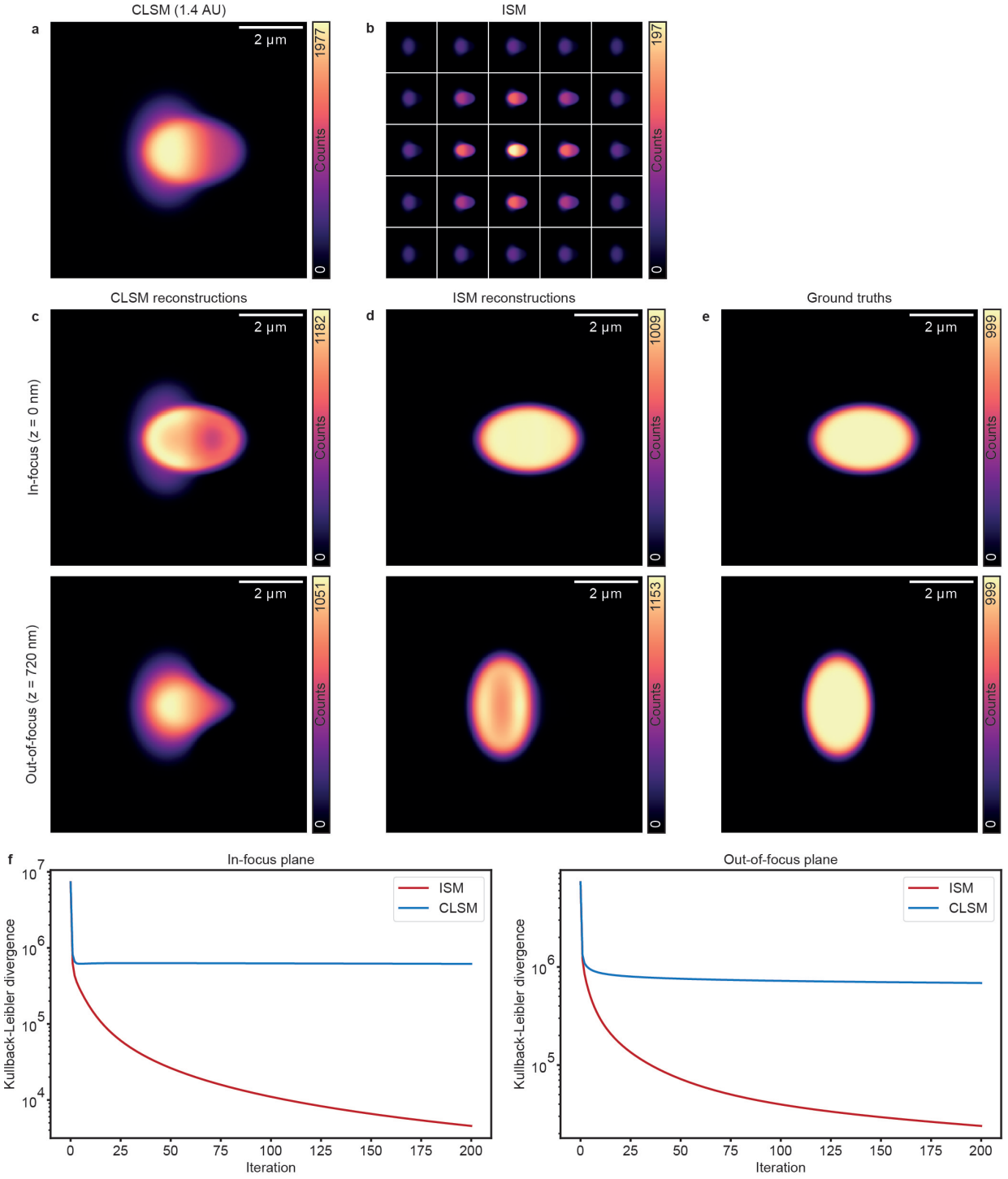

Figure 3.11: **Low-frequency sample analysis.** **a**, raw confocal image (open pinhole). **b**, raw ISM dataset. **c**, result of s^{2}ISM applied to the single confocal image. **d**, result of s^{2}ISM applied to the full ISM dataset. **e**, ground truth for each axial plane. **f**, Kullback-Leibler divergence of the reconstructions against the corresponding ground truth.

s²ISM for each plane of the phantom thick object. Then, we calculated the KL-divergence at each iteration, for each reconstruction, as we can see in Fig. 3.12f.

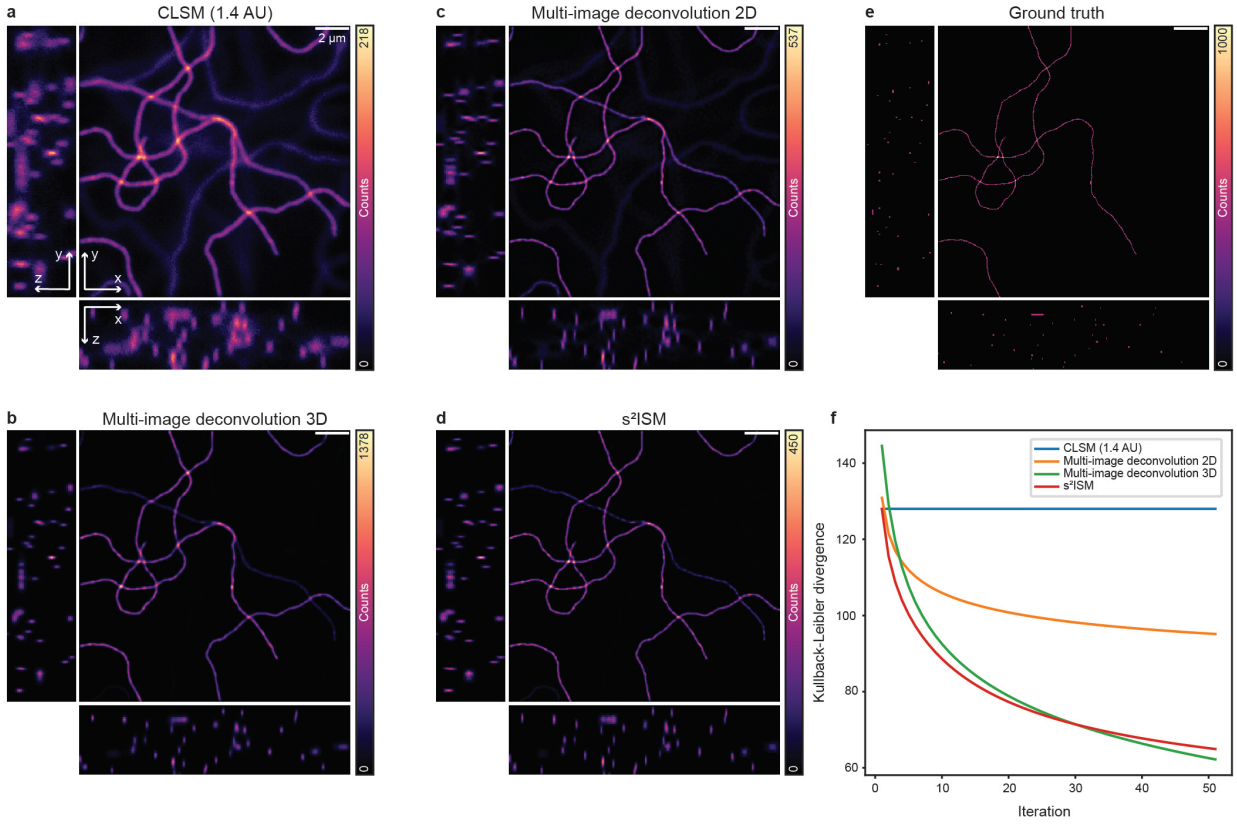

Figure 3.12: **Comparison of s²ISM and 3D multi-Image deconvolution.** **a**, open-pinhole confocal image. **b**, 3D multi-image deconvolution (20 iterations). **c**, 2D multi-image deconvolution applied plane-by-plane (20 iterations per plane). **d**, s²ISM applied plane-by-plane (20 iterations per plane) using as out-of-focus position $z_2 = 300\,\mathrm{nm}$. **e**, Ground truth. **f**, Kullback-Leibler divergence of the stacks to the ground truth, calculated with each slice normalized and integrated over the axial dimension. The three slices (xy, xz, yz) are taken from the center of the stack.

In this scenario, axial information is explicitly encoded in the images of the 3D stack, allowing each slice of the volume to be reconstructed with improved optical sectioning. However, this approach requires the experimental acquisition of multiple images at different axial positions of the sample. Although the resulting dataset contains richer information and enables a complete three-dimensional reconstruction of the specimen, it is time-consuming and increases the light dose delivered to the sample.

In contrast, s²ISM requires only a single planar dataset to achieve optical sectioning. The key difference is that out-of-focus light is discarded rather than reassigned to its corresponding axial plane. As a consequence, s²ISM reconstructions are typically less bright than those obtained through 3D deconvolution (Fig. 3.12b,d). Nevertheless, our method accurately reconstructs the specimen's structure, enabling rapid measurements while minimizing phototoxicity.

3.3.1 Optimal number of iterations

The iterative nature of the algorithm raises the issue of determining an appropriate stopping point.

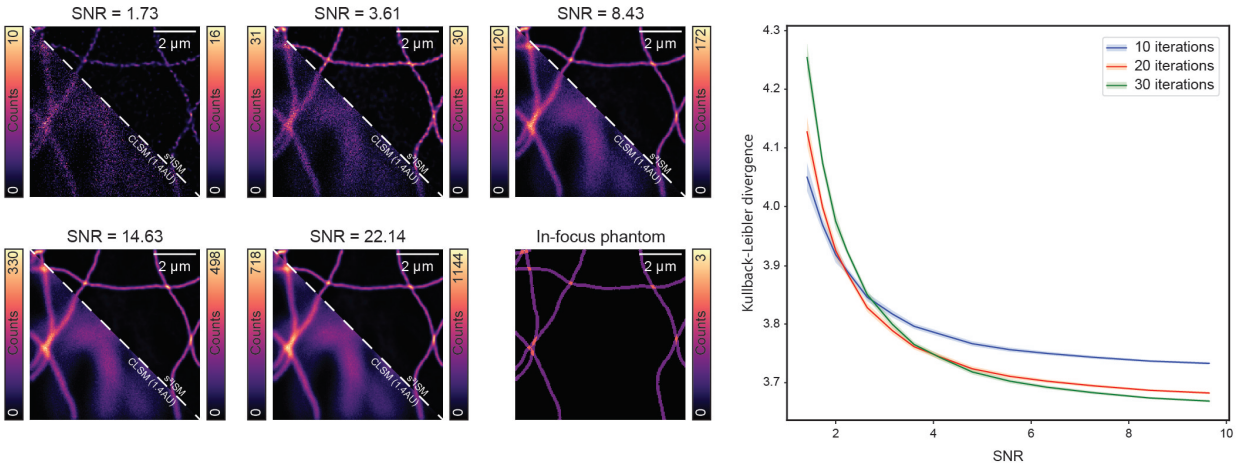

Figure 3.13: **Reconstruction fidelity at different signal-to-noise ratio levels.** We simulated the same image at different signal-to-noise ratio (SNR) levels by defining the ground-truth as the phantom multiplied by SNR^2. We applied s²ISM to multiple realizations of the same simulated dataset and compared them with the corresponding ground truth. In order to make images at different SNRs comparable, we normalized each image and ground truth by the total number of photon counts before calculating the Kullback-Leibler divergence. The solid line is the average value, and the shaded area represents the standard deviation, which is calculated over ten noise realizations at the same SNR. The results show that the higher the SNR, the more s²ISM can be iterated before noise amplification occurs. Note that the optimal number of iterations is highly sample-dependent and cannot be calculated in advance.

We compared the reconstructions with the ground truth at different iteration numbers, showing that the algorithm exhibits semiconvergent behavior, similarly to conventional deconvolution methods. An early stopping criterion should be adopted to prevent noise amplification [30]. In the absence of GT, the optimal number of iterations cannot be determined a priori and typically depends on the specimen and on the SNR of the images. In general, lower SNR values require earlier termination of the reconstruction (Fig. 3.13). Nonetheless, we can estimate the optimal iteration range by exploiting the photon conservation property of our algorithm. The total number of photons in the raw dataset equals the sum of the two reconstructed images, regardless of the iteration. The photon counts of individual images $o_k^{(m)}$ may change with each iteration. However, the overall quantity of photons is unchanged after each iteration. The proof is the

following:

$$
\begin{aligned}
&\sum_k \int o_k^{(m+1)}(\boldsymbol{x}_s)\,\mathrm{d}\boldsymbol{x}_s = \\
&= \sum_k \iint o_k^{(m)}(\boldsymbol{x}_s)\left[h_k(-\boldsymbol{x}_s|\boldsymbol{x}_d) * \frac{i(\boldsymbol{x}_s|\boldsymbol{x}_d)}{\sum_j [o_j^{(m)}(\boldsymbol{x}_s) * h_j(\boldsymbol{x}_s|\boldsymbol{x}_d)]}\right]\mathrm{d}\boldsymbol{x}_s\,\mathrm{d}\boldsymbol{x}_d = \\
&= \sum_k \iiint o_k^{(m)}(\boldsymbol{x}_s) h_k(\boldsymbol{x}-\boldsymbol{x}_s|\boldsymbol{x}_d)\frac{i(\boldsymbol{x}|\boldsymbol{x}_d)}{\sum_j [o_j^{(m)}(\boldsymbol{x}) * h_j(\boldsymbol{x}|\boldsymbol{x}_d)]}\,\mathrm{d}\boldsymbol{x}_s\,\mathrm{d}\boldsymbol{x}_d\,\mathrm{d}\boldsymbol{x} = \\
&= \sum_k \iint o_k^{(m)}(\boldsymbol{x}) * h_k(\boldsymbol{x}|\boldsymbol{x}_d)\frac{i(\boldsymbol{x}|\boldsymbol{x}_d)}{\sum_j [o_j^{(m)}(\boldsymbol{x}) * h_j(\boldsymbol{x}|\boldsymbol{x}_d)]}\,\mathrm{d}\boldsymbol{x}_d\,\mathrm{d}\boldsymbol{x} = \\
&= \iint \sum_k [o_k^{(m)}(\boldsymbol{x}) * h_k(\boldsymbol{x}|\boldsymbol{x}_d)]\frac{i(\boldsymbol{x}|\boldsymbol{x}_d)}{\sum_j [o_j^{(m)}(\boldsymbol{x}) * h_j(\boldsymbol{x}|\boldsymbol{x}_d)]}\,\mathrm{d}\boldsymbol{x}_d\,\mathrm{d}\boldsymbol{x} = \\
&= \iint i(\boldsymbol{x}_s|\boldsymbol{x}_d)\,\mathrm{d}\boldsymbol{x}_s\,\mathrm{d}\boldsymbol{x}_d.
\end{aligned}
\tag{3.32}
$$

Indeed, s²ISM works by reassigning photons to their correct axial plane while preserving the total photon count. As a result, s²ISM does not create or destroy light, but only redistributes photons between the in-focus and out-of-focus planes. Most of the photon exchange between the two planes occurs during the first iterations and then gradually reaches a plateau (Fig. 3.14), indicating that convergence has been achieved and that the algorithm should be stopped. If the iterations are not properly terminated, noise starts to overfit the estimation, making the introduction of a regularization term necessary, as discussed in Chapter 6.

We also observe that flux conservation, together with the positivity constraint on the solution, effectively imposes an implicit L_1 constraint. If the reconstructed object $o^{(m)}$ satisfie $o^{(m)}(\boldsymbol{x}_s) \geq 0 \quad \forall\, \boldsymbol{x}_s$, and photon flux conservation holds true

$$
\sum_{\boldsymbol{x}_s} o^{(m)}(\boldsymbol{x}_s) = \sum_{\boldsymbol{x}_s,\boldsymbol{x}_d} i(\boldsymbol{x}_s|\boldsymbol{x}_d) = c \in \mathbb{R} > 0, \tag{3.33}
$$

then its L_1 norm is fixed by construction:

$$
\|o^{(m)}\|_1 = \sum_{\boldsymbol{x}_s} |o^{(m)}(\boldsymbol{x}_s)| = \sum_{\boldsymbol{x}_s} o^{(m)}(\boldsymbol{x}_s) = c. \tag{3.34}
$$

Thus, every reconstruction lies in the set of non-negative functions with fixed norm. This corresponds to an implicit L_1 constraint: no explicit penalty is introduced in the loss function, but the solution is still constrained to a fixed total energy.

3.3.2 Aberration robustness assessment

To assess the robustness of s²ISM to aberrations, we generated a synthetic phantom and simulated its acquisition under an aberrated imaging system by convolving it with aberrated point spread functions (PSFs) (Fig. 3.15). In particular, we introduced vertical astigmatism and vertical

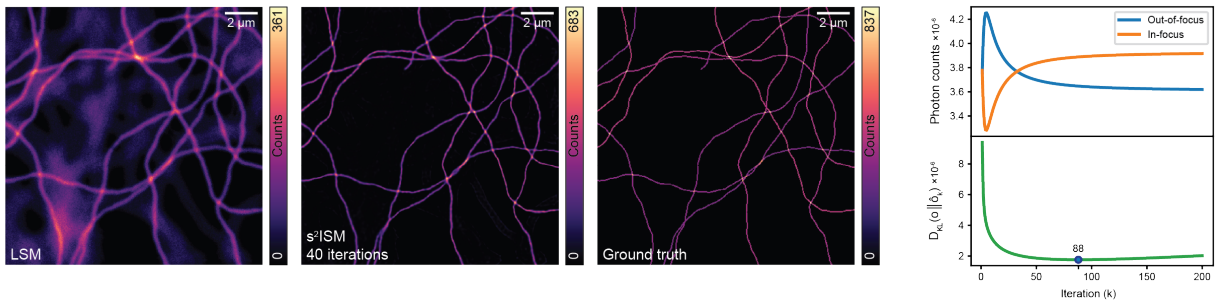

Figure 3.14: Picking the best object estimator during the iterative routine in a synthetic scenario. Ground-truth is at our disposal. We choose the iteration producing the object estimator that minimize the error function w.r.t. the imposed in-focus ground-truth.

trefoil. This procedure was repeated for multiple aberration amplitudes. Each dataset was then processed using the standard s²ISM algorithm (Fig. 3.15e) and, for comparison, using s²ISM with prior knowledge of the aberrations to be corrected (Fig. 3.15f). For each aberration type and strength, we quantified performance by computing the KL-divergence between the corrected and uncorrected reconstructions and the in-focus ground truth. As shown in Fig. 3.15c, the error metric remains largely insensitive to aberration strength up to 1 radian for vertical trefoil and up to 1.25 radians for astigmatism. This level of robustness enables the extension of s²ISM to more challenging imaging conditions, including deep imaging and nonlinear modalities such as 2PE.

3.3.3 Generalization to upsampling and its constraints

If the pixel size $\Delta \boldsymbol{x}_s$ of an ISM acquisition is identical to the detector pitch $\Delta \boldsymbol{x}_d$, the ISM dataset contains enough redundancy to enable the reconstruction of an image with twice the pixels per axis. Here, we demonstrate that the s²ISM algorithm can achieve this goal.

First, we define the coordinates of the discretized scanning space as

$$\left(\frac{x_s}{\Delta x_s}, \frac{y_s}{\Delta y_s}\right) = (n_x, n_y) = \boldsymbol{n} \in \mathbb{N}^2 \tag{3.35}$$

We model the downsampling process using a space-variant excitation PSF

$$\tilde{h}_{\text{exc}} = \begin{cases} h_{\text{exc}}(\boldsymbol{x}_s) & \text{if } \boldsymbol{x}_s \in \chi \\ 0 & \text{otherwise} \end{cases} \tag{3.36}$$

where

$$\chi := \left\{ \boldsymbol{x}_s \mid (n_x, n_y) \in (2\mathbb{N}+1)^2 \right\} \tag{3.37}$$

is the subset of pixels with odd indices. Therefore, depending on the position, we can factorize the likelihood probability of Eq. 3.5 in two terms

$$P[i(\boldsymbol{x}_s|\boldsymbol{x}_d)|o(\boldsymbol{x}_s)] = \prod_{x_d} \prod_{x_s \in \chi} P_1[i(\boldsymbol{x}_s|\boldsymbol{x}_d)|o(\boldsymbol{x}_s)] \prod_{x_s \notin \chi} P_2[i(\boldsymbol{x}_s|\boldsymbol{x}_d)|o(\boldsymbol{x}_s)] \tag{3.38}$$

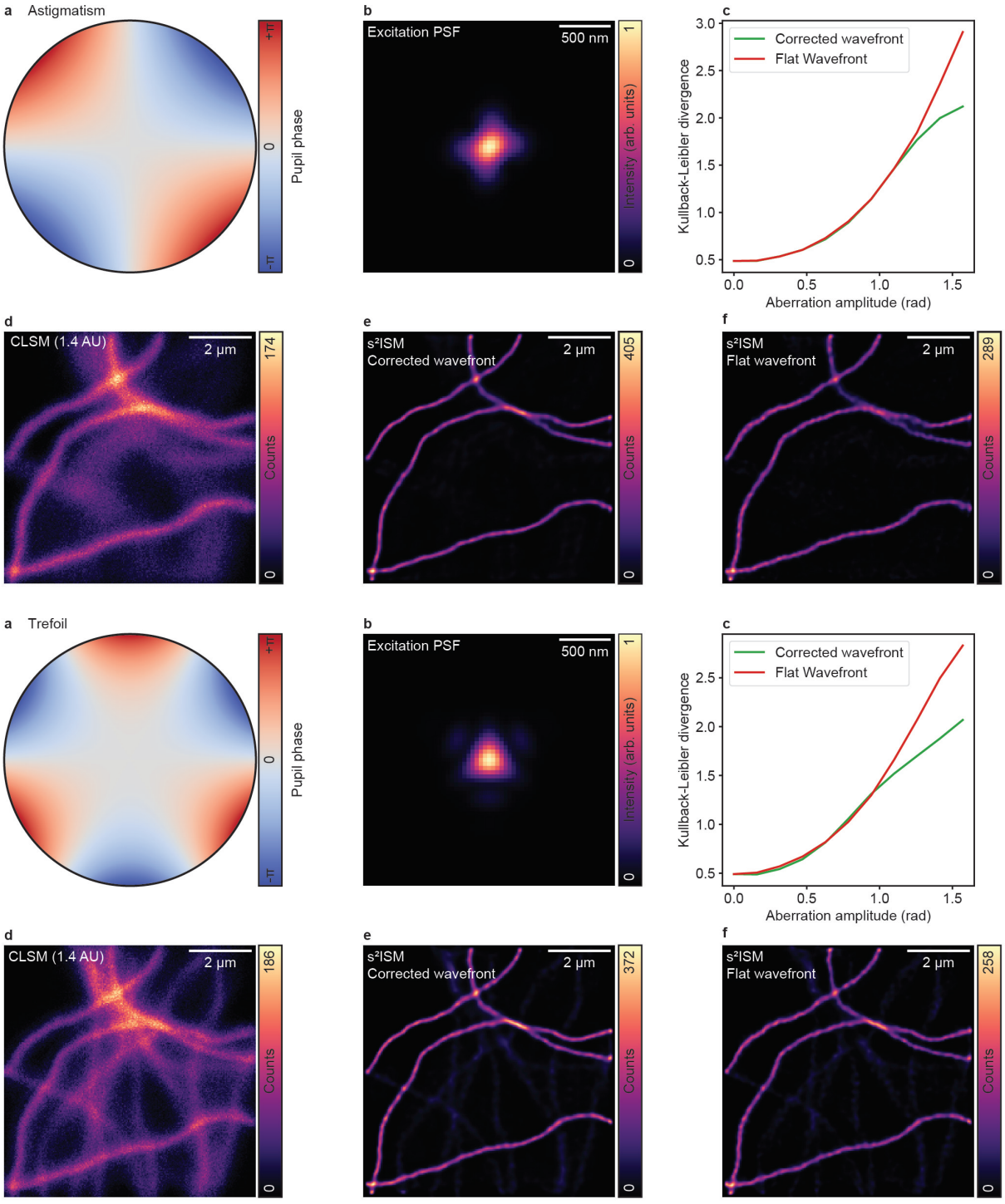

Figure 3.15: **Effect of PSF mismatch.** Simulation of ISM data with optical aberrations (vertical astigmatism, top; vertical trefoil, bottom). **a**, wavefront error. **b**, in-focus PSF with aberration. **c**, Kullback-Leibler divergence of s^{2}ISM reconstruction vs. in-focus ground truth using aberrated or ideal PSFs. **d**, confocal image with aberration. **e**, s^{2}ISM reconstruction using aberrated PSFs. **f**, s^{2}ISM reconstruction using ideal PSFs. Reconstructions at 40 iterations show s^{2}ISM is robust to small or moderate PSF mismatches.

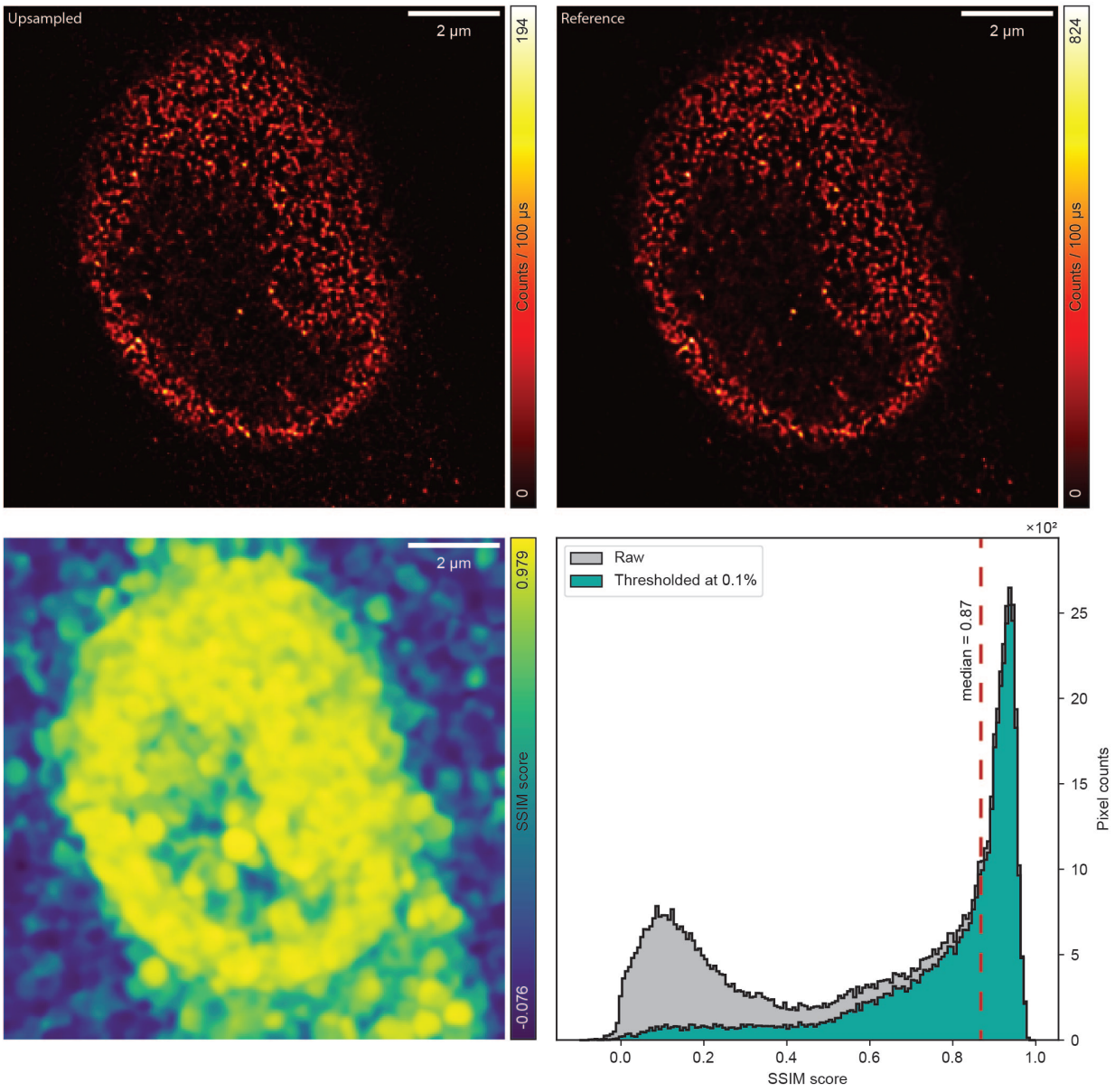

Figure 3.16: **Upsampling fidelity.** We calculated the structural similarity index measure (SSIM) score between the upsampled and reference images from Fig. 3.19b. The SSIM score is locally high in the regions of the sample rich with signal. Unsurprisingly, the SSIM score decreases in the regions with little or no signal. Therefore, we removed the pixels with less than 0.1% of the maximum intensity of the reference images to calculate the thresholded histogram of the SSIM score. The median score calculated on the thresholded histogram is 0.87.

We observe that the PSFs and the images are null when no excitation occurs. Therefore, we have that

$$P_2[i(\boldsymbol{x}_s|\boldsymbol{x}_d)|o(\boldsymbol{x}_s)] = \left.\frac{\sum_k[o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)]^{i(\boldsymbol{x}_s|\boldsymbol{x}_d)} e^{-\sum_k[o_k(\boldsymbol{x}_s)*h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)]}}{i(\boldsymbol{x}_s|\boldsymbol{x}_d)!}\right|_{\boldsymbol{x}_s\notin\chi} = \\ = \frac{0^0 \cdot e^0}{0!} \tag{3.39}$$

The expression above contains the indeterminate form 0^0, which is generally undefined. Nonetheless, in the case of fluorescence imaging, the relation between excitation and signal is given by the power law

$$y = \alpha x^n \tag{3.40}$$

where y is the fluorescence flux, x is the excitation flux, $n \geq 1$ is the order of the excitation process, and $\alpha > 0$ is an efficiency factor. Therefore, our context allows for the definition of 0^0 as the result of the following limit

$$\lim_{x\to 0} x^y = \lim_{x\to 0} x^{\alpha x^n} = \exp\left(\alpha \lim_{x\to 0} x^n \log x\right) = 1 \tag{3.41}$$

which is readily calculated using L'Hôpital's rule.

Therefore, we obtain that

$$\prod_{x_s\notin\chi} P_2[i(\boldsymbol{x}_s|\boldsymbol{x}_d)|o(\boldsymbol{x}_s)] = 1 \tag{3.42}$$

which means that the total likelihood is unaffected by the lack of signal in the pixels with an even index. As a result, we can reconstruct an upsampled image by maximizing the likelihood of Eq. 3.5 without modifications. Consequently, the iterative reconstruction rule remains the same as Eq. 3.12, even if the raw dataset contains signal only on the pixels with an odd index

$$\tilde{i}(\boldsymbol{x}_s|\boldsymbol{x}_d) = \begin{cases} i(\boldsymbol{x}_s|\boldsymbol{x}_d) & \text{if } \boldsymbol{x}_s \in \chi \\ 0 & \text{otherwise} \end{cases} \tag{3.43}$$

To demonstrate that the reconstruction fills the voids in the dataset by exploiting the redundancy in the ISM dataset, we develop a simplified toy-model. Without loss of generality, we consider 1D scanning and detector coordinates. Furthermore, we approximate the ISM PSFs as identical but shifted.

$$h(x_s|x_d) \approx h(x_s - \mu(x_d)) \tag{3.44}$$

In the ideal case of Gaussian PSFs, point-like detector elements, and no Stokes-shift, we have an explicit equation for the shift-vector

$$h(x_s|x_d) \approx h(x_s - \frac{x_d}{2}) = h\left(n\Delta x_s - \frac{m\Delta x_d}{2}\right) \tag{3.45}$$

where $x_s = n\Delta x_s$ and $x_d = m\Delta x_d$ are the discretized scanning and detector coordinate, respectively. We assume that the dataset was acquired respecting the upsampling condition $\Delta x_s = \Delta x_d$

and that the reconstruction is taking place on a grid twofold finer than the original. Thus, we have that on the reconstruction grid $2\Delta x_s = \Delta x_d$. Substituting the latter identity in Eq. 3.45 and dropping the pixel size, we have

$$h(n|m) \approx h(n-m) = h(n) * \delta(n-m) \tag{3.46}$$

The iterative reconstruction rule of s²ISM becomes

$$o_j^{(l+1)}(n) = o_j^{(l)}(n) \sum_m h_j(-n|m) * \frac{\tilde{i}(n|m)}{\sum_k o_k^{(l)}(n) * h_k(n|m)}$$

For the sake of simplicity, we consider only the detector elements with indexes $m \in \{0,1\}$. Expanding the terms of the summation and exploiting the associative property of the convolution operation, we have

$$\begin{aligned} o_j^{(l+1)}(n) = o_j^{(l)}(n) \cdot h_j(-n) * \delta(n) * \frac{\tilde{i}(n|0)}{\sum_k o_k^{(l)}(n) * h_k(n|0)} + \\ + o_j^{(l)}(n) \cdot h_j(-n) * \delta(n-1) * \frac{\tilde{i}(n|1)}{\sum_k o_k^{(l)}(n) * h_k(n|1)} = \\ = o_j^{(l)}(n) \cdot h_j(-n) * \frac{\tilde{i}(n|0)}{\sum_k o_k^{(l)}(n) * h_k(n|0)} + \\ + o_j^{(l)}(n) \cdot h_j(-n) * \frac{\tilde{i}(n-1|1)}{\sum_k o_k^{(l)}(n) * h_k(n|1)} \end{aligned} \tag{3.47}$$

Thus, the reconstructions at pixel n are given by the contributions coming from the dataset $\tilde{i}$ evaluated at pixels with opposite parity, namely n and $n-1$. Consequently, even if the original dataset did not contain signal in the even pixels, the algorithm fills them exploiting the properties of ISM. Thus, s²ISM can generate an upsampled image, relaxing Nyquist's criterion by a factor of two. The above argument easily generalizes to the 2D case with any number of detector elements, even if some residual dependency of the detector position remains in the shape of the PSFs.

3.4 Experimental results

We compared the performances on experimental samples of s²ISM with conventional methods (Ext. Fig. C.5). The simplest approach consists of summing all images in the ISM dataset, which yields an image equivalent to that produced by a confocal microscope with a pinhole size equal to that of the detector array, namely 1.4 Airy units (AU). Similarly, extracting only the image built by the central element of the detector array corresponds to generating a confocal image with a closed pinhole (0.3 AU). We know that closing the pinhole improves lateral resolution and optical sectioning but irremediably compromises the SNR. Conversely, by opening the pinhole, the photon collection efficiency improves, but the resolution is limited by diffraction, and optical

sectioning deteriorates according to the size of the detector. The aforementioned reconstruction methods completely disregard the spatial information that the SPAD array detector provided, and we use them only as a reference towards CLSM. The results should not be considered ISM images. A smarter reconstruction method is APR, an algorithm that first estimates the shifts between the images of the raw datasets, later registers them, and finally performs the summation [31]. As a result, the image preserves the super-resolution enabled by the confocal effect while exploiting all the photons collected and achieving an excellent SNR. However, APR is designed assuming all the images of the ISM dataset are identical but shifted and rescaled. A more rigorous approach to ISM reconstruction is multi-image deconvolution [22], which takes into account the unique structure of the PSFs of the imaging system. However, APR and multi-image deconvolution cannot reject out-of-focus light. Thus, their feasibility is limited in the context of thin samples or using detector arrays with a limited size – compromising SNR. A solution to reject defocused light is focus-ISM [32], which builds upon APR. After reassignment, each micro-image – namely, the distribution of light on the detector space – is fitted to a two-component Gaussian mixture. The broadest distribution is interpreted as background and discarded, generating an optically-sectioned image pixel-by-pixel. A recent re-implementation, named spatial phasor ISM [33], follows the same working principle but classifies micro-images in the frequency domain following a phasor transformation along the detector dimension. While effective, focus-ISM stands on an approximated model of ISM image formation, and its local nature makes it very sensitive to noise. It is incompatible with multi-image deconvolution and cannot benefit from it. Finally, s^2ISM is the most comprehensive reconstruction algorithm designed for ISM, enabling all the benefits provided by the previously described methods without sacrificing any image feature.

We also carried out the comparison in Fourier space by calculating the radial spectra of the images from App. Note B.4 and App. Fig. B.1. The results confirm that s^2ISM outperforms the conventional reconstruction also in experimental frameworks. Indeed, the high-frequency contrast uniquely correlates with superior lateral resolution and suppression of low-frequency background. Furthermore, the spectrum of the reconstructed image indicates a lower noise level, confirming the improvement of the SNR suited for live-cells and time-lapse imaging.

3.4.1 Quantitative characterization of experimental capabilities

In order to rigorously quantify the performance of our ISM reconstruction method, we performed imaging on a fluorescent resolution target [34]. First, we estimated the resolution gain provided by s^2ISM using a pattern of gradually spaced lines. The target comprises fourteen sets of line quartets. Each set includes two external lines spaced at a fixed distance and two internal lines with decreasing spacing, ranging from $390\,\mathrm{nm}$ to $0\,\mathrm{nm}$ in steps of $30\,\mathrm{nm}$. This target allows for a reliable assessment of image resolution based on its fundamental definition: the ability to distinguish two separate objects. Our algorithm does not incorporate explicit regularizers. Therefore, it does not favour a specific structure, and it is completely sample-agnostic. Consequently, any observed improvement in resolution is attributable solely to the capability of s^2ISM to decode the information embedded in the raw dataset. Indeed, a couple of lines appear separated only if the reconstruction from the data matches the forward model from Eq. 2.15.

More in detail, we reconstructed using established ISM processing algorithms for super-

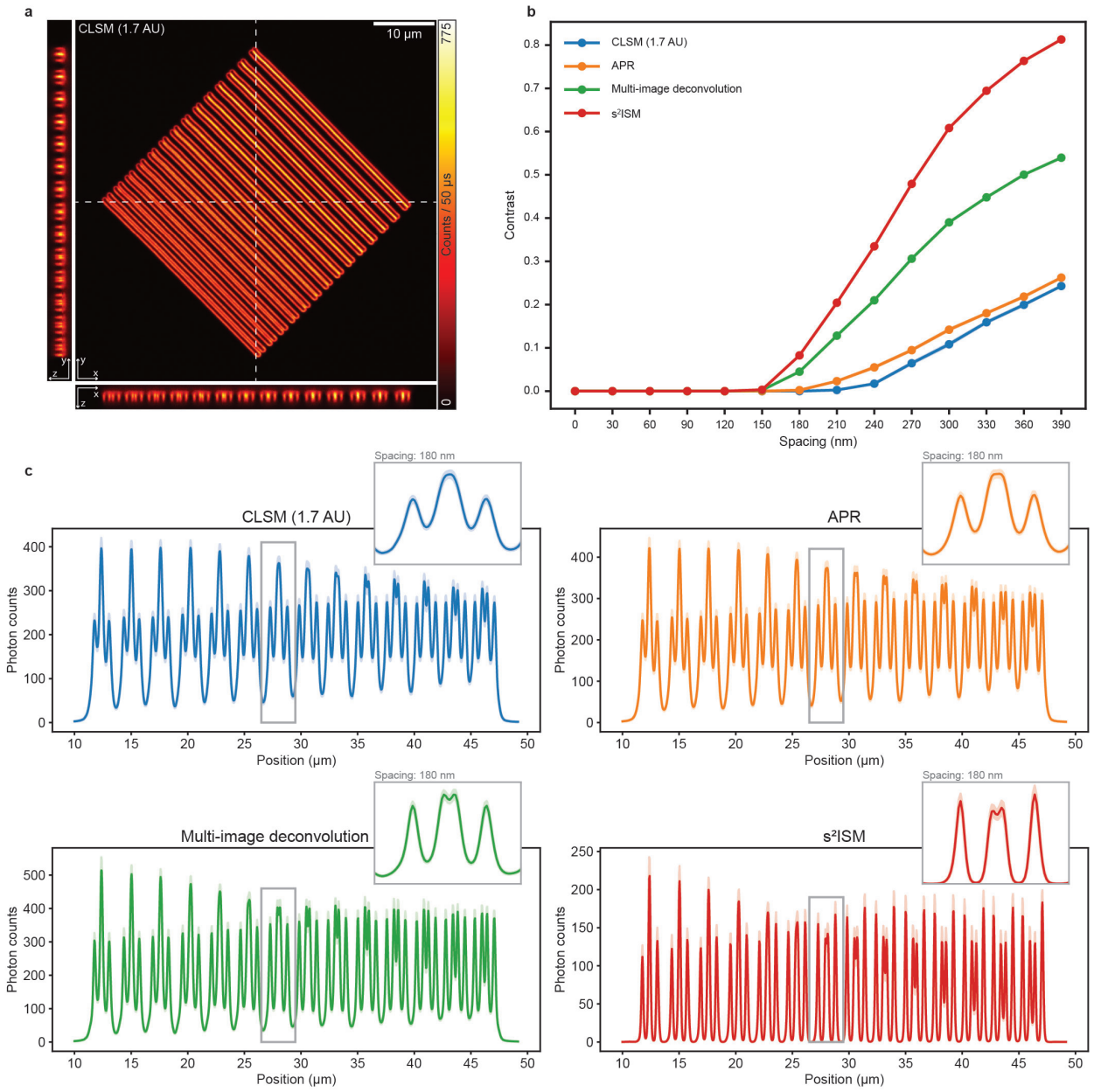

Figure 3.17: **Resolution enhancement of ISM reconstruction algorithms.** **a**, 3D stack of confocal images of the resolution target from Argolight, consisting of a grid a gradually spaced parallel lines. **b**, Analysis of the contrast of the gradually spaced lines shown in Fig. 6.3a, calculated for each conventional ISM reconstruction algorithm and s^{2}ISM. The shaded area represents the standard error of the mean. **c**, Line profile of the central image of **a**, reconstructed with different algorithms and averaged along the direction parallel to lines. The shaded area represents the standard deviation.

resolution (App. Fig. C.6). On one hand, we summed the ISM dataset over all the detector elements – obtaining the corresponding open-pinhole confocal image for comparison. On the other hand, we applied APR and multi-image deconvolution. Also, we applied s²ISM using the procedure estimating the PSFs directly from the data, as described in the previous chapter. For each image, we measured the visibility of the dip between adjacent lines at each spacing. The contrast value in App. Fig. C.6b at each spacing is calculated as:

$$c = \frac{I_{max} - I_{min}}{I_{max}} \tag{3.48}$$

with

$$I_{max} = \frac{I_{max}(0) + I_{max}(1)}{2} \tag{3.49}$$

where $I_{\max}(0)$ denotes the value of the integrated line profile at the left line of the sample. This value is known *a priori* because the sample is structured. Similarly, the value at the second line, $I_{\max}(1)$, is also known *a priori*. The quantity $I_{\min}$ corresponds to the value of the line profile at the central position between the two lateral lines and is likewise known *a priori*. The results shown in Fig. 3.17b demonstrate that our reconstruction method enhances the lateral resolution of the final image, beating the diffraction limit. Indeed, the cut-off spacing for the confocal image is $210\,\mathrm{nm}$, which is pushed to $150\,\mathrm{nm}$ thanks to s²ISM. The theoretical cut-off of ISM is twice as good as that of diffraction-limited imaging (App. Note 2.2.3). Although we do not reach the theoretical limit, we increase lateral resolution by 40%. Furthermore, the target sample has a non-negligible thickness. Therefore, the enhanced optical sectioning of s²ISM also improves the contrast at any spacing, outperforming existing reconstruction procedures.

Then, we quantified the capability of s²ISM to reject out-of-focus light by imaging a 3D stair composed of ring-shaped targets uniformly spaced along the axial coordinate (Ext. Fig. C.4). We reconstructed a single-plane ISM dataset taken from the center of the stairs using APR – which acts as a reference and has no additional sectioning compared to the open-pinhole confocal image. Then, we applied the popular background removal algorithm known as *rolling ball* [35] to the APR image. Finally, we applied ISM-specific algorithms for optical sectioning, namely Focus-ISM and s²ISM. Our results (Fig. 3.18) demonstrate that s²ISM provides the highest degree of optical sectioning, doubled compared to conventional ISM implementations.

Indeed, s²ISM leverages the unique information provided by structured detection – namely, the modulation of intensity on the detector plane – to decode the axial information embedded into the raw dataset. The fingerprint modulation is independent of the lateral structure of the specimen. Thus, s²ISM allows for sample-agnostic optical sectioning without relying on properties of the structure on the image coordinate (such as brightness or sharpness). Without the additional information contained in the multiple images of the raw dataset, computational optical sectioning is not possible.

Finally, we estimated the SNR enhancement achieved by s²ISM by comparing the reconstruction of multiple noisy realizations of the same ISM dataset (App. Note A.4). Thanks to the SPAD's temporal encoding capabilities, we can unravel the time dimension if we are not interested in analyzing the temporal behaviour, to extract a few i.i.d. noise realizations on the same

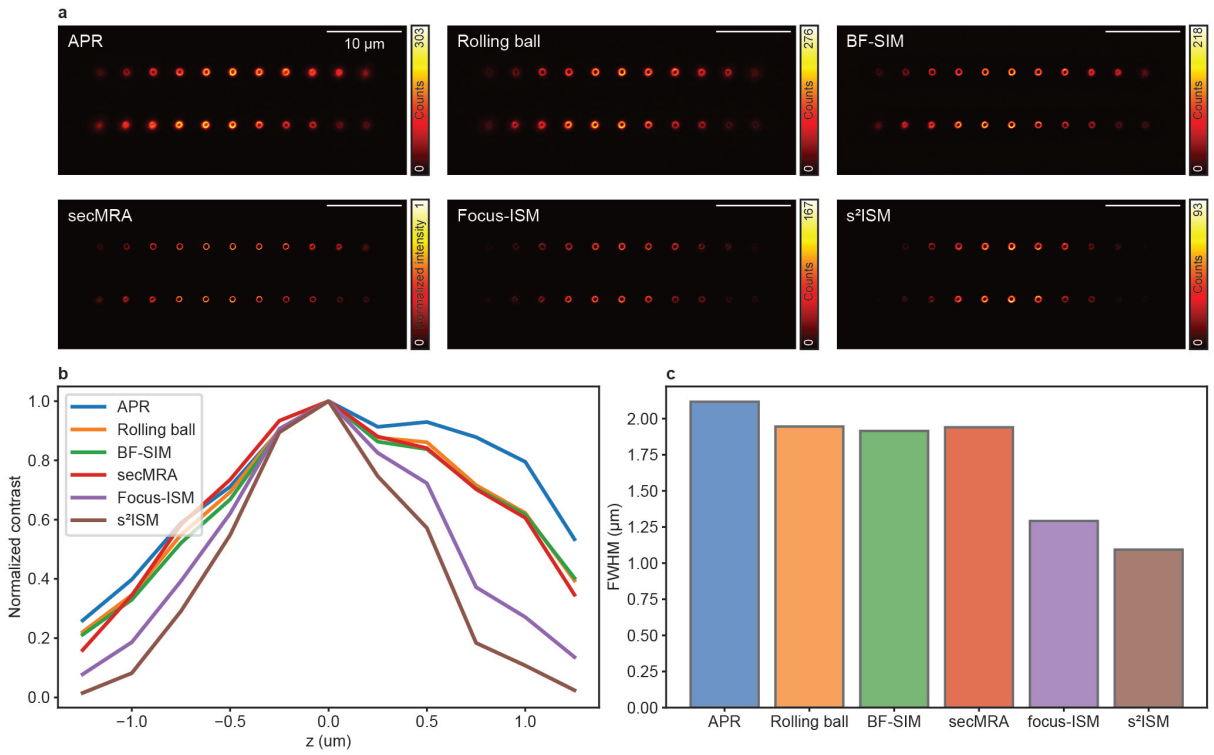

Figure 3.18: **Background removal algorithm comparison.** **a**, comparison of adaptive pixel reassignment (APR) with the result of different background removal algorithms. Focus-ISM and s²ISM are ISM-specific algorithms, while the remaining are general-purpose. **b**, contrast curves, obtained by calculating the relative photon counts associated with the rings at different axial positions, relative to the presented processing algorithms. **c**, full width at half maximum (FWHM) of the curves presented in **b**. The experimental data is the same as in App. Fig. C.4 and B.3.

underlying structure. Specifically, we calculated the SNR as the ratio of the pixel-wise mean to the standard deviation of the s²ISM reconstructions at successive iterations. This was done relative to the confocal image, which served as the baseline. The results demonstrate a significant SNR improvement with just a few iterations of the s²ISM algorithm when compared to the initial confocal image. On the other hand, if the reconstruction is applied for too many iterations, noise amplification might happen, degrading the SNR. This result underlines once again the importance of an early stop, or a regularizer, which we will explore in depth in the next chapters.

3.5 Versatility of s²ISM

As recently demonstrated [22], the ISM dataset contains enough redundancy to allow for a two-fold lateral upsampling. Indeed, the raw images of the ISM dataset are inherently shifted by a quantity – the shift-vector – which depends on the $\boldsymbol{x}_d$ coordinate. If the pixel size $\Delta\boldsymbol{x}_s$ is chosen to equal the pitch of the array detector $\Delta\boldsymbol{x}_d$ projected in the sample plane, then the images $i(\boldsymbol{x}_s|\boldsymbol{x}_d)$ are mutually shifted by half the pixel size. Since the coordinates between two pixels of

an image are sampled by another image of the ISM dataset, fusing the images allows for doubling the number of scanned points in the final reconstruction (Fig. 3.19a). An important implication is that ISM enables super-resolution also in the digital sense. Indeed, it is possible to recover an image with a pixel size respecting the Nyquist sampling criterion, even though this criterion was not respected by the raw images. In other words – if the upsampling condition $\Delta \boldsymbol{x}_s = \Delta \boldsymbol{x}_d$ is respected – ISM enables surpassing both Nyquist's and Abbe's limits.

The s²ISM reconstruction method can also leverage the extra information encoded into the ISM dataset to achieve image upsampling without sacrificing the benefits demonstrated so far in 3.3.3. To demonstrate it, we acquired a dataset of nuclear pore complexes on the nuclear membrane of a HeLa cell (Ext. Fig. C.8). We used two different values of pixel size, $160\,\mathrm{nm}$ and $80\,\mathrm{nm}$, the first not respecting the Nyquist criterion but respecting the upsampling criterion. Then, we reconstructed an image using the s²ISM method on both datasets, reaching a target pixel size of $80\,\mathrm{nm}$. Despite having roughly one-fourth of the photon counts of the reference data, the upsampled image is correctly reconstructed, as demonstrated by the similarity with the reference image. More quantitatively, the structural similarity index measure (SSIM) (App. Note B.3) and calculated on the upsampled and reference images reaches local values as high as 0.98, with a median value of 0.87, Fig. 3.16. Thus, the s²ISM method enables high-fidelity reconstruction of undersampled raw data, paving the way for gentler and faster imaging.

As described previously, s²ISM requires only a single plane dataset, improves the SNR, and enables faster acquisition. Such benefits are especially useful in the context of live cell imaging over a long period of time. We demonstrate the feasibility of s²ISM on a time series of live mitochondria (Fig. 3.19c and Ext. Fig. C.9) acquisitions. We performed the reconstruction frame-by-frame, obtaining a sequence of images with high resolution and optical sectioning. Our results demonstrate how high-quality continuous imaging of live cells can be performed at moderate pixel dwell times (Supp. movie). Similarly, we extend s²ISM to multi-colour imaging by applying the reconstruction algorithm to each channel.

Multi-target imaging is necessary to decode and understand biological behaviors, like the organization of organelles inside the cell, and is required in any microscope for biological investigation. In more detail, we modelled the image formation of each channel using the corresponding excitation and emission wavelengths. The results demonstrate that s²ISM can easily be applied to multi-channel datasets with a simple sequential reconstruction to obtain high-resolution and high optical sectioning multi-colour images (Fig. 3.19d and Ext. Fig. C.10).

Also, we are able to image both the lamina envelop of the nucleus with the mitochondria arrangement inside the cell. Naturally, the mitochondria surround the nucleus and stay outside the nucleus.

The benefits of s²ISM are also particularly useful while investigating thick samples, namely those whose axial structure extends over a range much larger than the microscope's depth of field. In this case, the images are typically corrupted by out-of-focus background, given the inherent 3D structure of the specimen. We demonstrate the advantage of s²ISM by acquiring a volumetric dataset of a cell (Ext. Fig. C.2). Then, we moved to an even more challenging sample, namely a cleared cerebellum slice of a mouse, which we imaged up to a depth of $10\,\mu\mathrm{m}$ (Ext. Fig. C.1). Thanks to the improved optical sectioning, each slice of the 3D stack is free from defocused blur. At the same time, the size of the detector array ensures a good collection efficiency and a

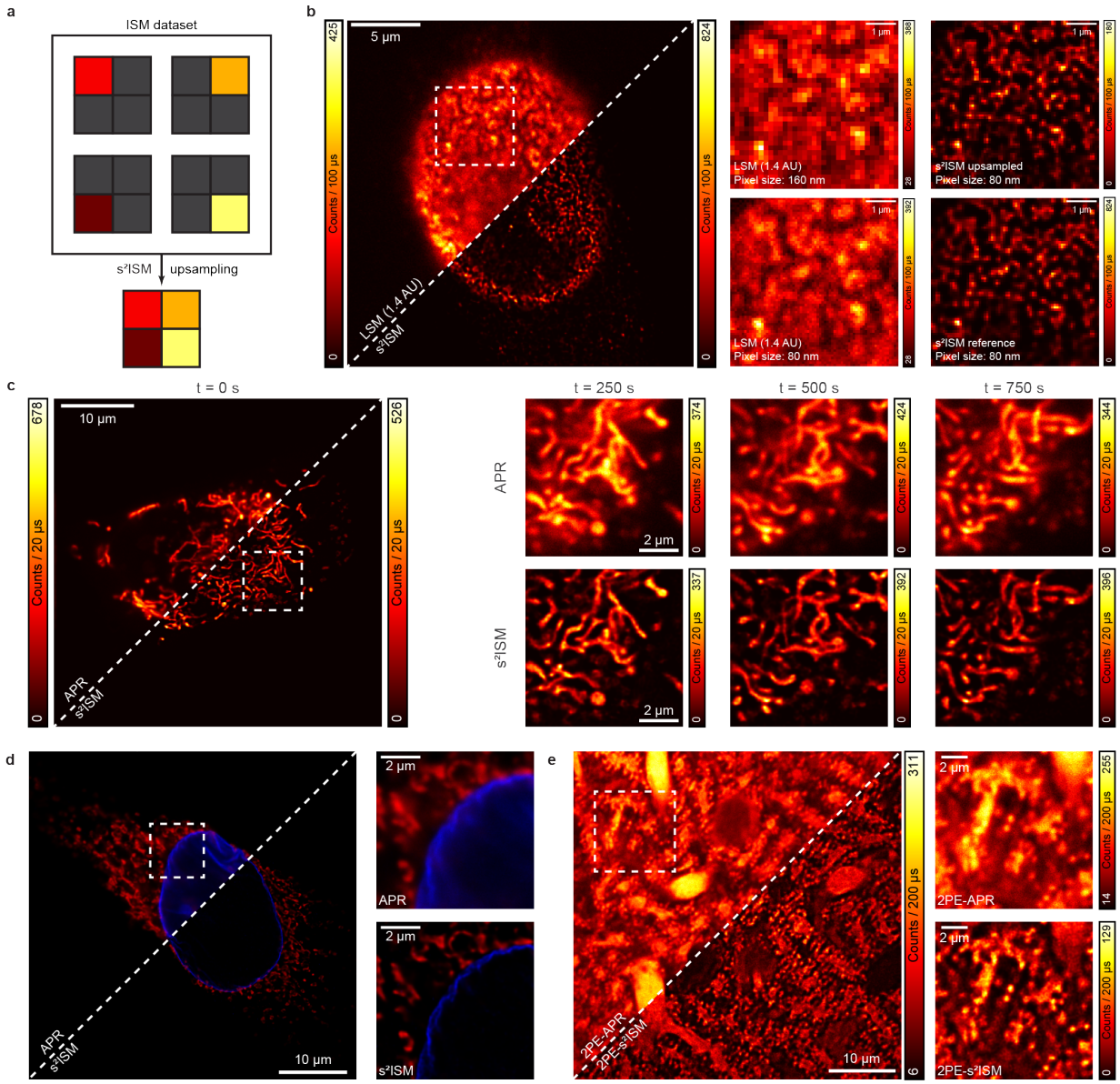

Figure 3.19: **Generalization of s²ISM. a**, Sketch of the upsampling working principle. The images of the ISM dataset are shifted, and the mutual redundancy can be used to fill the gaps and reconstruct an image on a finer grid than that generated by the acquisition process. **b**, Results of s²ISM reconstruction with and without upsampling at the same target pixel size of $80\,\mathrm{nm}$. The sample is an immunostained HeLa cell for nuclear pore complexes on the nucleus surface (rabbit anti-Nup-153 combined with anti-rabbit Abberior STAR 635P) **c**, time series of a live HeLa cell with stained mitochondria (Mito Tracker Orange). **d**, Multicolor imaging of a fixed HeLa cell with mitochondria and nuclear membrane immunostained with two different fluorophores (mouse anti-ATP Synthase combined with anti-Mouse Alexa647 and rabbit anti-lamin B1 combined with anti-rabbit Alexa488, respectively). **e**, two-photon excitation imaging of Purkinje cells in a slice of a mouse's cerebellum at a depth of roughly $10\,\mu\mathrm{m}$.

correspondingly good SNR.

To further push the capability to explore deep samples, we demonstrate the feasibility of the combination of two-photon excitation (2PE) with s^2ISM. Indeed, our reconstruction method is general and – in principle – can be used with any laser scanning microscope equipped with a detector array. More in detail, 2PE exploits near-infrared light – which is less affected by scattering, compared to visible light – and the non-linear excitation process further improves the optical sectioning of the microscope. We imaged the same mouse cerebellum slice described previously, demonstrating an improvement in the image quality in the more general case of multi-photon excitation (Ext. Fig. C.11). Finally, we tested the robustness of s^2ISM in a deep uncleared sample. It is well-known that the imaging quality at depths larger than a few tens of microns is typically compromised by specimen-induced aberrations and scattering [36]. The shorter the wavelength of light, the more severe the problem. Therefore, in such crowded samples, optical sectioning is a requirement. For short-to-mid depths, CLSM is one of the methods of choice. In this scenario, conventional ISM fails to provide substantial advantages compared to CLSM, since the improvement of lateral resolution becomes hard to appreciate at low signal-to-background levels. Conversely, s^2ISM enables both optical sectioning and lateral resolution enhancement, which are key features for imaging at moderate depths. We demonstrate such capabilities by imaging a Zebrafish embryo (Fig. 3.20) using linear excitation at $488\,\mathrm{nm}$. Despite the image quality degrading with increasing depth, s^2ISM provides a robust and substantial advantage over confocal imaging for the whole depth range we explored, up to almost $40\,\mu\mathrm{m}$. Since conventional ISM reconstruction methods provide little or no advantage, s^2ISM can be regarded as an enabling technology which expands the versatility and feasibility of ISM-based imaging over a much wider range of experimental scenarios.

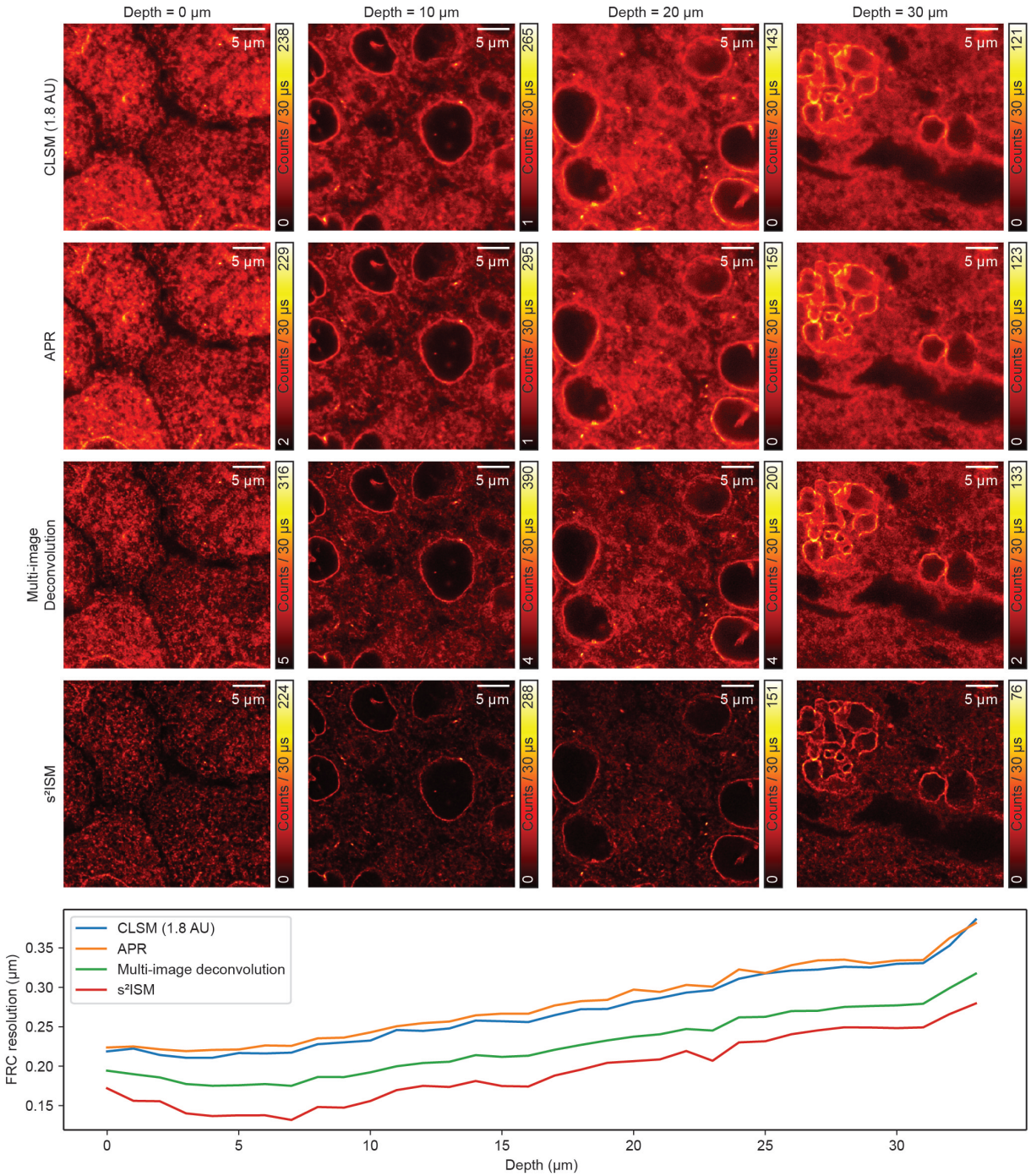

Figure 3.20: **Deep imaging performance.** Images of a Zebrafish embryo at high stage in which lamin is labelled with eGFP, excited at $\lambda = 488\,\mathrm{nm}$ (CW, average power at sample plane: $16.7\,\mu\mathrm{W}$), acquired at increasing depth starting from the coverslip inner surface. We reconstructed a pair of images at each depth with different algorithms and calculated the corresponding Fourier ring correlation (FRC). We estimated the effective image resolution at each depth using the $1/7$ fixed threshold. The strong background hinders the resolution improvement of APR and multi-image deconvolution, completely or partially, respectively. Instead, s^{2}ISM maximizes the performance by simultaneously removing out-of-focus light and leveraging the super-resolution effect. Despite the image quality monotonically decreases with increasing depth, due to scattering and aberrations, the improvement of s^{2}ISM over conventional methods remains constant.

Conclusion of Chapter 3

In this chapter, the s^2ISM reconstruction algorithm was derived as a maximum-likelihood inversion of the ISM forward model. The formulation explicitly accounts for in-focus and out-of-focus contributions, enabling simultaneous super-resolution and optical sectioning from a single-plane ISM acquisition. Data-driven strategies for PSFs estimation and algorithm initialization were introduced, ensuring robustness against experimental misalignments and aberrations. Theoretical and numerical characterization demonstrated the algorithm's stability, robustness, and capability for digital upsampling, validating the theoretical advantages of the proposed framework under controlled conditions. Additionally, we validated the s^2ISM framework using experimental data such as structured and biological samples. Quantitative analysis confirmed the simultaneous enhancement of lateral resolution, SNR, and optical sectioning, without the need for hardware modifications or shrinking detector size. The versatility of the method was demonstrated across different imaging conditions, sample types, and noise regimes. These results establish s^2ISM as a practical and robust reconstruction strategy for real-world ISM applications.

References

[1] Alessandro Zunino, Francesco Garzella, Alberta Trianni, Peter Saggau, Paolo Bianchini, Alberto Diaspro, and Martí Duocastella. Multiplane encoded light-sheet microscopy for enhanced 3d imaging. *ACS Photonics*, 8:3385–3393, 11 2021.

[2] Winfried Denk, James H. Strickler, and Watt W. Webb. Two-photon laser scanning fluorescence microscopy. *Science*, 248:73–76, 4 1990.

[3] Alberto Diaspro, Paolo Bianchini, Giuseppe Vicidomini, Mario Faretta, Paola Ramoino, and Cesare Usai. Multi-photon excitation microscopy. *BioMedical Engineering OnLine*, 5:36, 12 2006.

[4] Peter O. Bayguinov, Dennis M. Oakley, Chien-Cheng Shih, Daniel J. Geanon, Matthew S. Joens, and James A. J. Fitzpatrick. Modern laser scanning confocal microscopy. *Current Protocols in Cytometry*, 85:e39, 7 2018.

[5] Colin J R Sheppard and Yunrui Gong. Improvement in axial resolution by interference confocal microscopy. *Optik*, 87:129–132, 3 1991.

[6] Stefan Hell and Ernst H. K. Stelzer. Properties of a 4pi confocal fluorescence microscope. *Journal of the Optical Society of America A*, 9:2159, 12 1992.

[7] Xiang Hao, Yiming Li, Shuang Fu, Yanghui Li, Yingke Xu, Cuifang Kuang, and Xu Liu. Review of 4pi fluorescence nanoscopy. *Engineering*, 11:146–153, 4 2022.

[8] Gustafsson, Agard, and Sedat. I5m: 3d widefield light microscopy with better than 100 nm axial resolution. *Journal of Microscopy*, 195:10–16, 7 1999.

[9] J. Bewersdorf, R. Schmidt, and S. W. Hell. Comparison of i5m and 4pi-microscopy. *Journal of Microscopy*, 222:105–117, 5 2006.

[10] M A A Neil, R Juškaitis, and T Wilson. Method of obtaining optical sectioning by using structured light in a conventional microscope, 1997.

[11] M A A Neil, R Juskaitis, and T Wilson. Real time 3d fluorescence microscopy by two beam interference illumination, 1998.

[12] M. G. L. Gustafsson. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of Microscopy*, 198(2):82–87, 2000.

[13] Xin Chen, Suyi Zhong, Yiwei Hou, Ruijie Cao, Wenyi Wang, Dong Li, Qionghai Dai, Donghyun Kim, and Peng Xi. Superresolution structured illumination microscopy reconstruction algorithms: a review, 12 2023.

[14] Mats G.L. Gustafsson, Lin Shao, Peter M. Carlton, C. J.Rachel Wang, Inna N. Golubovskaya, W. Zacheus Cande, David A. Agard, and John W. Sedat. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophysical Journal*, 94:4957–4970, 6 2008.

[15] Aurélie Jost, Elen Tolstik, Polina Feldmann, Kai Wicker, Anne Sentenac, and Rainer Heintzmann. Optical sectioning and high resolution in single-slice structured illumination microscopy by thick slice blind-sim reconstruction. *PLoS ONE*, 10, 7 2015.

[16] Emmanuel Soubies and Michael Unser. Computational super-sectioning for single-slice structured-illumination microscopy. *IEEE Transactions on Computational Imaging*, 5:240–250, 12 2018.

[17] Yann Lai-Tim, Laurent M. Mugnier, François Orieux, Roberto Baena-Gallé, Michel Paques, and Serge Meimon. Jointly super-resolved and optically sectioned bayesian reconstruction method for structured illumination microscopy. *Opt. Express*, 27(23):33251–33267, Nov 2019.

[18] James D. Manton. Answering some questions about structured illumination microscopy. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 380, 4 2022.

[19] Awoke Negash, Thomas Mangeat, Patrick C. Chaumet, Kamal Belkebir, Hugues Giovannini, and Anne Sentenac. Numerical approach for reducing out-of-focus light in bright-field fluorescence microscopy and superresolution speckle microscopy. *J. Opt. Soc. Am. A*, 36(12):2025–2029, Dec 2019.

[20] Ruijie Cao, Yaning Li, Wenyi Wang, Guoxun Zhang, Gang Wang, Yu Sun, Wei Ren, Jing Sun, Yiwei Hou, Xinzhu Xu, Jiakui Hu, Yanye Lu, Changhui Li, Jiamin Wu, Meiqi Li, Junle Qu, and Peng Xi. Dark-based optical sectioning assists background removal in fluorescence microscopy. *bioRxiv*, 2024.

[21] M Bertero, P Boccacci, G Desiderà, and G Vicidomini. Image deblurring with poisson data: from cells to galaxies. *Inverse Problems*, 25:123006, 12 2009.

[22] Alessandro Zunino, Marco Castello, and Giuseppe Vicidomini. Reconstructing the image scanning microscopy dataset: An inverse problem. *Inverse Problems*, 39, 6 2023.

[23] William Hadley Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62:55, 1 1972.

[24] L. B. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79:745, 6 1974.

[25] Stephan Roth, Colin J. R. Sheppard, and Rainer Heintzmann. Superconcentration of light: circumventing the classical limit to achievable irradiance. *Optics Letters*, 41:2109, 5 2016.

[26] Fernando Caprile, Luciano A. Masullo, and Fernando D. Stefani. Pyfocus – a python package for vectorial calculations of focused optical fields under realistic conditions. application to toroidal foci. *Computer Physics Communications*, 275, 6 2022.

[27] Alessandro Zunino, Eli Slenders, Francesco Fersini, Andrea Bucci, Mattia Donato, and Giuseppe Vicidomini. Open-source tools enable accessible and advanced image scanning microscopy data analysis. *Nature Photonics*, 17:457–458, 6 2023.

[28] Yanquan Mo, Kunhao Wang, Liuju Li, Shijia Xing, Shouhua Ye, Jiayuan Wen, Xinxin Duan, Ziying Luo, Wen Gou, Tongsheng Chen, Yu-Hui Zhang, Changliang Guo, Junchao Fan, and Liangyi Chen. Quantitative structured illumination microscopy via a physical model-based background filtering algorithm reveals actin dynamics. *Nature Communications*, 14:3089, 5 2023.

[29] Yiwei Hou, Wenyi Wang, Yunzhe Fu, Xichuan Ge, Meiqi Li, and Peng Xi. Multi-resolution analysis enables fidelity-ensured deconvolution for fluorescence microscopy. *eLight*, 4:14, 2024.

[30] Yiming Liu, Spozmai Panezai, Yutong Wang, and Sjoerd Stallinga. Noise amplification and ill-convergence of richardson-lucy deconvolution. *Nature Communications*, 16:911, 1 2025.

[31] Marco Castello, Giorgio Tortarolo, Mauro Buttafava, Takahiro Deguchi, Federica Villa, Sami Koho, Luca Pesce, Michele Oneto, Simone Pelicci, Luca Lanzanó, Paolo Bianchini, Colin J. R. Sheppard, Alberto Diaspro, Alberto Tosi, and Giuseppe Vicidomini. A robust and versatile platform for image scanning microscopy enabling super-resolution flim. *Nature Methods*, 16:175–178, 2 2019.

[32] Giorgio Tortarolo, Alessandro Zunino, Francesco Fersini, Marco Castello, Simonluca Piazza, Colin J.R. Sheppard, Paolo Bianchini, Alberto Diaspro, Sami Koho, and Giuseppe Vicidomini. Focus image scanning microscopy for sharp and gentle super-resolved microscopy. *Nature Communications*, 13, 12 2022.

[33] Yuran Huang, Zitong Ye, Zhimin Zhang, Hanchu Ye, Liang Xu, Xiang Hao, Yubing Han, Cuifang Kuang, and Xu Liu. Spatial phasor analysis for optical sectioning nanoscopy. *ACS Photonics*, 1 2025. doi: 10.1021/acsphotonics.4c02336.

[34] Argolight. Argolight help center.

[35] Stanley R. Sternberg. Biomedical image processing. *Computer*, 16:22–34, 1 1983.

[36] Jacopo Bertolotti and Ori Katz. Imaging in complex media. *Nature Physics*, 18:1008–1017, 9 2022.

4

s^2ISM for Fluorescence Lifetime Imaging

Single-photon avalanche diode (SPAD) array detectors combine single-photon sensitivity with spatial and temporal resolution. By virtue of their array architecture, SPAD detectors are capable of recording the spatial position at which each photon is detected in the plane where the detector is placed, which, in the context of ISM, corresponds to the microscope image plane. In addition to spatial information, SPAD arrays can measure the arrival time of individual photons with a temporal precision on the order of a few hundred picoseconds (App. A.2). This timing capability enables the determination of photon arrival times relative to reference signals, such as the start of the acquisition or the excitation event generated by a pulsed laser. Such time-correlated single-photon counting (TCSPC) measurements form the basis of fluorescence lifetime imaging microscopy (FLIM). Fluorescence lifetime, defined as the average time a fluorophore remains in the excited state before emitting a photon, is an intrinsic photophysical parameter that is largely independent of fluorophore concentration and excitation intensity. Crucially, its value is highly sensitive to the local nano-environment of the molecule, including factors such as oxygen concentration, pH, and the presence of quenching species. Moreover, fluorescence lifetime is modulated by molecular interactions, energy transfer processes, and conformational changes, making it a powerful functional contrast mechanism for probing biochemical states and dynamic processes in biological systems. SPAD detectors allow recording a photon arrival–time histogram for each detector element, and, when combined with point scanning, this results in a spatio-temporal ISM dataset in which each pixel contains both spatial and temporal information. By extending the forward model introduced in Chapter 3 to include this temporal dimension and the associated instrument response function (IRF), we unlock a range of highly informative imaging modalities that jointly leverage spatial super-resolution, optical sectioning, and fluorescence lifetime contrast.

We first extend the s^2ISM algorithm to the context of FLIM. Specifically, we derive a time-resolved version of s^2ISM that reconstructs super-resolved images with optical sectioning while preserving the temporal fluorescence decay dynamics at each pixel. The output of this algorithm

can then be used as input to conventional fluorescence lifetime estimation methods, such as phasor analysis or model-based fitting approaches, to reconstruct fluorescence lifetime maps. Such maps allow combining structural and functional imaging. Indeed, the fluorescence lifetime signature allows understanding the interaction or nano-environment properties of the tagged biomolecule.

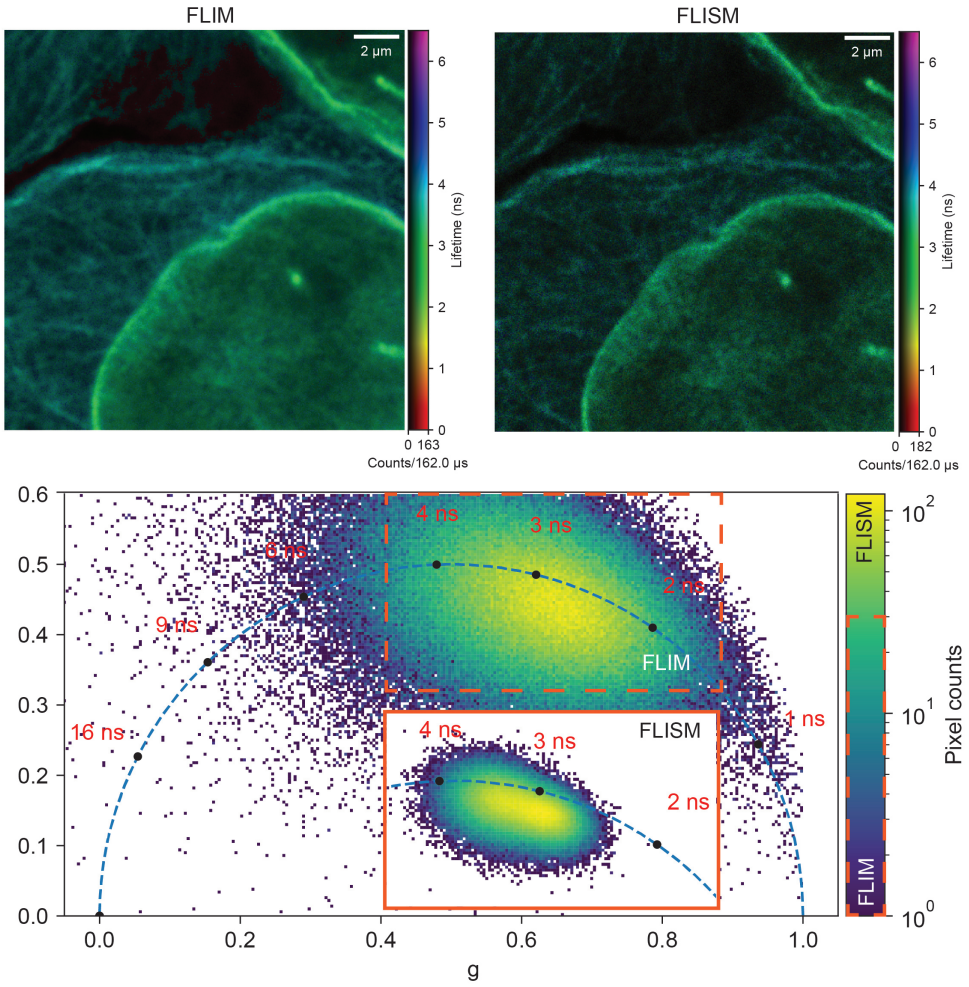

Figure 4.1: **FLIM vs FLISM reconstructions and corresponding phasors.** FLIM (top left) and FLISM (top right) images of HeLa cells, with tubulin stained with STAR RED ($\tau = 3.4\,\mathrm{ns}$) and lamin-A on the nuclear membrane stained with STAR 635 ($\tau = 2.8\,\mathrm{ns}$). Both fluorophores are excited using the same source at $\lambda = 640\,\mathrm{nm}$. The corresponding phasor plots demonstrate that the APR step improves the accuracy of FLIM lifetime estimation compared to the open-pinhole reconstruction. The dashed inset indicates the region in which the corresponding FLISM data are also displayed. The threshold for pixel intensities included in the analysis was set to 0.15% of the corresponding maximum image intensity.

We then address the problem of multiplexed, or multi-target, imaging based on fluorescence lifetime contrast. Fluorescence lifetime signatures provide an attractive alternative to spectral discrimination, as they enable the identification and spatial mapping of distinct biomolecular species even in the presence of strong spectral overlap. The compatibility of ISM with lifetime-based multiplexing has previously been demonstrated by sequentially applying pixel reassignment

to each temporal bin of a TCSPC dataset, followed by phasor-based lifetime unmixing [1]. While this approach provides a useful proof of principle, it suffers from inherent limitations. Pixel reassignment allows fast visualization but is outperformed by multi-image deconvolution in terms of resolution, contrast, and noise suppression. Moreover, phasor analysis, although intuitive and computationally efficient, compresses the TCSPC data into a small number of coefficients, discarding a substantial fraction of the available temporal information and limiting unmixing performance. An alternative strategy consists of first applying the time-resolved s²ISM reconstruction to the original TCSPC dataset and subsequently performing fluorescence lifetime unmixing using MLE methods. However, the sequential nature of these workflows is prone to error propagation: applying maximum-likelihood estimators to reconstructed data may compromise statistical optimality, as reconstruction alters the original Poisson photon-counting statistics. These limitations motivate the development of a unified ISM framework that jointly performs multi-image deconvolution and fluorescence-lifetime unmixing in a single processing step. Finally, we demonstrate that these capabilities naturally extend to multi-pulse excitation architectures, in particular within the Pulse-Interleaved Excitation (PIE) framework, in which excitation pulses from different laser sources or wavelengths are temporally interleaved and separated by known delays. This temporal separation allows photons to be unambiguously associated with their corresponding excitation pulses based on their arrival times, enabling simultaneous multi-channel or multi-species imaging without increasing spectral crosstalk.

4.1 State of the art for fluorescence lifetime ISM

Existing ISM reconstruction algorithms exhibit a fundamental deficiency: they are inherently restricted to spatial processing and ignore the temporal dimension of the data altogether. In time-resolved ISM acquisitions, photon arrival times carry critical information about fluorescence dynamics and are indispensable for FLIM. However, this temporal information is not incorporated into current reconstruction frameworks. As a consequence, no dedicated reconstruction algorithm presently exists for fluorescence lifetime ISM that can jointly and coherently exploit the complete spatio-temporal information content of the measurements. Instead, the prevailing methodology adopts a sequential and decoupled processing strategy, in which spatial ISM reconstruction is first performed, and fluorescence lifetime analysis is subsequently applied as an independent post-processing step.

This separation fundamentally limits the full utilization of the rich information available in time-resolved ISM data. The TCSPC dataset is a five-dimensional dataset, which we denote as $i(\boldsymbol{x}_s, t|\boldsymbol{x}_d)$. The current state of the art for time-resolved ISM data relies on a sequential processing approach [2]. In this framework, APR is applied as an initial step to improve the SNR and spatial resolution, saturating the information carried by the detector coordinates. While this strategy is effective also at denoising the temporal decay curves — which are typically highly noisy — it does not optimally exploit the $\boldsymbol{x}_d$ coordinates. As a result, the intrinsic coupling between spatial and temporal information encoded in the raw data is lost, leading to an inefficient use of the full spatio-temporal structure of the dataset. The APR framework is the one described for the solo-spatial ISM dataset, in Sec. 2.2.4. The processing is exactly repeated, but leading to

the dataset depicted as

$$i_{APR}(\boldsymbol{x}_s, t) = \sum_{\boldsymbol{x}_d} i(\boldsymbol{x}_s + \boldsymbol{\mu}(\boldsymbol{x}_d), t|\boldsymbol{x}_d) \tag{4.1}$$

Subsequently, we processed the $i_{APR}(\boldsymbol{x}_s, t)$ dataset using conventional fluorescence lifetime analysis methods. It consists of fitting the fluorescence decay trace – i.e., the photon-arrival-time histogram – at each scanning coordinate $\boldsymbol{x}_s$ with a *a priori* model to estimate the fluorescence lifetime. Curve fitting is a widely used method for fluorescence lifetime analysis, offering several advantages. It allows accurate estimation of fluorescence lifetimes and can resolve multiple lifetimes, which may indicate the presence of different fluorophores or conformational states. Its simplicity and ease of implementation further contribute to its popularity. However, it also has notable limitations. It is sensitive to noise, such that small fluctuations can produce significant errors, and it is strongly model-dependent, meaning results are only as accurate as the chosen *a priori* decay model. The method provides limited information beyond the parameters explicitly included in the model and can be computationally intensive for complex models or large datasets, which may constrain its practical use.

In addition to exponential fitting, the Fluorescence Lifetime ISM (FLISM) dataset, namely $i_{APR}(\boldsymbol{x}_s, t)$, was also amenable to analysis using the phasor approach. Again, thanks to the higher SNR, the points in the FLISM phasor plots, known as the phasor coordinates (g, s), were located more precisely than those in the FLIM counterpart, namely the analysis performed on the raw dataset integrated over the $\boldsymbol{x}_d$, without reassignment (App. Fig. 4.1FLIM).

4.2 s²ISM Extension to Time-Correlating Single-Photon Counting

By extending the capabilities of s²ISM to TCSPC datasets, we demonstrate that the proposed algorithm—and, more generally, the spatial information provided by array-based detection—offers substantial advantages for fluorescence lifetime imaging. First, the improved SNR achieved through the joint spatio-temporal reconstruction directly translates into more accurate lifetime estimates, as all fluorescence lifetime estimation methods applied to the reconstructed data, such as decay fitting or phasor analysis, are inherently sensitive to noise. Second, the enhanced spatial resolution provided by s²ISM leads to more precise lifetime maps by reducing spatial averaging over heterogeneous structures. Finally, the effective rejection of out-of-focus fluorescence—often characterized by heterogeneous lifetime components that contribute background signal—further improves the accuracy and reliability of lifetime estimation for in-focus structures. Collectively, these effects establish s²ISM as a powerful framework not only for spatial super-resolution and optical sectioning, but also for robust and quantitative fluorescence lifetime imaging.

In conventional fluorescence microscopy, the spatial and temporal impulse response functions are unrelated. Therefore, the spatiotemporal response function of the ISM setup is given by the product of the spatial PSF with the temporal IRF:

$$h_k(t, \boldsymbol{x}_s|\boldsymbol{x}_d) = h_k(\boldsymbol{x}_s|\boldsymbol{x}_d) \cdot h(t|\boldsymbol{x}_d) \tag{4.2}$$

We can estimated the IRF of the system, $h(t, \boldsymbol{x}_d)$, by averaging multiple ($\sim 10^4$) experimental recordings of gold beads scattering.
Given the spatio-temporal kernel, the generalized discrete forward model reads:

$$i(\boldsymbol{x}_s, t|\boldsymbol{x}_d) = \sum_k o_k(t, \boldsymbol{x}_s) * h_k(t, \boldsymbol{x}_s|\boldsymbol{x}_d) \tag{4.3}$$

where $*$ is the convolution operator with respect to (t, x_s, y_s). Since time is effectively an additional scanning coordinate, all the results derived in Section 3 still hold if we normalize them such that:

$$\sum_t \sum_{\boldsymbol{x}_d} \sum_{\boldsymbol{x}_s} h_k(t, \boldsymbol{x}_s, \boldsymbol{x}_d) = 1 \tag{4.4}$$

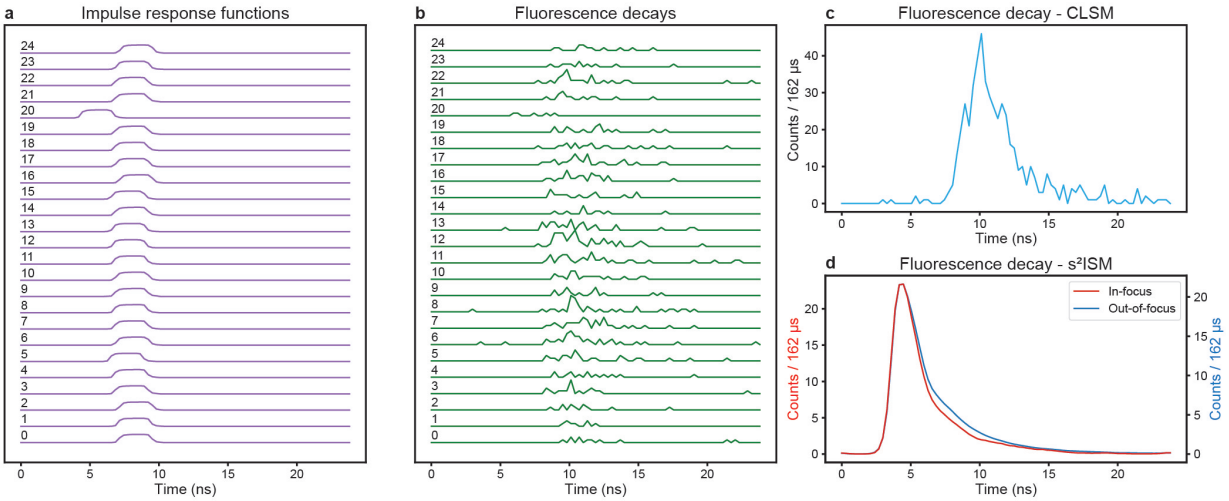

Figure 4.2: **Fluorescence decay reconstruction with s²ISM.** The data in this panel refer to a single pixel of the dataset shown in Fig. 4.4c. **a**, impulse response function (IRF) of each channel of our custom DFD acquisition system. **b**, fluorescence decays recorded by each channel of the SPAD array detector. **c**, fluorescence decay reconstructed by summing all the channels of the detector. **d**, fluorescence decays of each plane reconstructed by s²ISM, generalized for spatio-temporal data. The reconstructed decays are not aligned with the confocal curve because the algorithm compensates for the temporal shift caused by the IRFs.

where $k \in \{1, 2\}$ is the depth index. Thereby, we can derive the following new version of the s²ISM reconstruction algorithm:

$$o_k^{(m+1)}(t, \boldsymbol{x}_s) = o_k^{(m)}(t, \boldsymbol{x}_s) \sum_{\boldsymbol{x}_d} h_k(-t, -\boldsymbol{x}_s|\boldsymbol{x}_d) * \frac{i(t, \boldsymbol{x}_s|\boldsymbol{x}_d)}{\sum_k o_k^{(m)}(t, \boldsymbol{x}_s) * h_k(t, \boldsymbol{x}_s|\boldsymbol{x}_d)} \tag{4.5}$$

where m is the iteration index and $*$ is the convolution operator with respect to the coordinates t and $\boldsymbol{x}_s$. The out-of-focus reconstruction o_2 is discarded, and the in-focus image o_1 is the final result. We stop the algorithm at an arbitrary number of iterations. To avoid noise amplification, we iterate at most 20 times on the reconstruction of experimental images. As we can see in the

panel result Fig. 4.2d, the temporal profile of the noisy raw data has been well reconstructed and cleaned from high-frequency, namely noisy oscillations, and furthermore eliminating the temporal decay corresponding to the out-of-focus emitters, then enhancing spatio-temporal sectioning. We developed a unified ISM framework that would jointly perform lateral resolution enhancement, optical sectioning, and fluorescence lifetime unmixing in a single processing step.

4.2.1 s²ISM for Fluorescence Lifetime Imaging (s²FLISM)

To validate the s²ISM generalization to the temporal dimension encoding the fluorescence lifetime, we simulated the PSFs of the ISM microscope using the open-source Python package BrightEyes-ISM [3], based on the vectorial diffraction model provided by the package pyFocus [4]. We modelled the SPAD array detector as a 5×5 array of pinholes with a pitch of $75\,\mu\mathrm{m}$ and an individual size of $50\,\mu\mathrm{m}$. For all simulations, we set the magnification to $M = 450$, the numerical aperture to $\mathrm{NA} = 1.4$, and the refractive index to match the one of the immersion oil $n = 1.51$. We generate the synthetic datasets using an excitation and emission wavelength of $640\,\mathrm{nm}$ and $660\,\mathrm{nm}$, respectively. Finally, we applied Poisson noise to the generated images. In all cases, we assumed that we illuminated the back aperture of the objective lens with a uniform plane wave – namely, we did not apply any aberrations. We modelled the IRFs of the time-resolved synthetic data using a rectangular window smoothed by a Gaussian kernel ($w = 2\,\mathrm{ns}$ and $\sigma = 0.3\,\mathrm{ns}$). We simulated the fluorescence decay as a single exponential. We used 81 time bins separated by $\Delta t = 298\,\mathrm{ps}$ to match our DFD acquisition system. To better distinguish the simulations from the experimental data, we present them using the magma and hot colormap, respectively.

For simulations conducted at red wavelengths, we use an excitation wavelength λ_{exc} of $640\,\mathrm{nm}$ and an emission wavelength (λ_{em}) of $660\,\mathrm{nm}$. Green wavelength simulations adopt $\lambda_{exc} = 561\,\mathrm{nm}$ and $\lambda_{em} = 585\,\mathrm{nm}$, while blue wavelength simulations utilize $\lambda_{exc} = 488\,\mathrm{nm}$ and $\lambda_{em} = 510\,\mathrm{nm}$. For two-photon excitation, we choose λ_{exc} =$910\,\mathrm{nm}$ and $\lambda_{em} = 500\,\mathrm{nm}$. By defining the reference unit $AU = \frac{1.22\lambda_{em}}{NA}$, with the numerical aperture of the oil objective at 1.4, we derive the projected dimensions of the SPAD array detector on the specimen, dependent on the selected light source. For STED PSF simulations, we employ a STED fluorescence lifetime $\tau_F = 3.5\,\mathrm{ns}$, a maximum saturation factor $\xi = 15$, and a pulse duration of $1\,\mathrm{ns}$. Using GPU parallelization, the computation time is 1 second per iteration for a $2000 \times 2000 \times 81 \times 25$ spatiotemporal dataset on a computer equipped with an 8-core CPU ($3.6\,\mathrm{GHz}$), 32 GB of RAM, and a GPU with 8 GB of dedicated RAM. Without GPU parallelization on the same hardware, the computation time is 15 seconds per iteration.

Synthetic validation

We generate a spatiotemporal dependent raw phantom ISM dataset. It consists of filaments with a lifetime value of $\tau = 3\,\mathrm{ns}$ (in-focus) and a combination of $\tau = 3\,\mathrm{ns}$ and $\tau = 6\,\mathrm{ns}$ (out-of-focus).

We reconstructed the latter using APR, compensating for the shifts both in space and time, obtaining the FLISM image (Fig. 4.3a), [5]. Using the phasor approach [6], we calculated a lifetime value for each pixel. It is noticeable in the FLISM image that where the out-of-focus light superimposes with the in-focus structure, the lifetime value cannot be correctly estimated

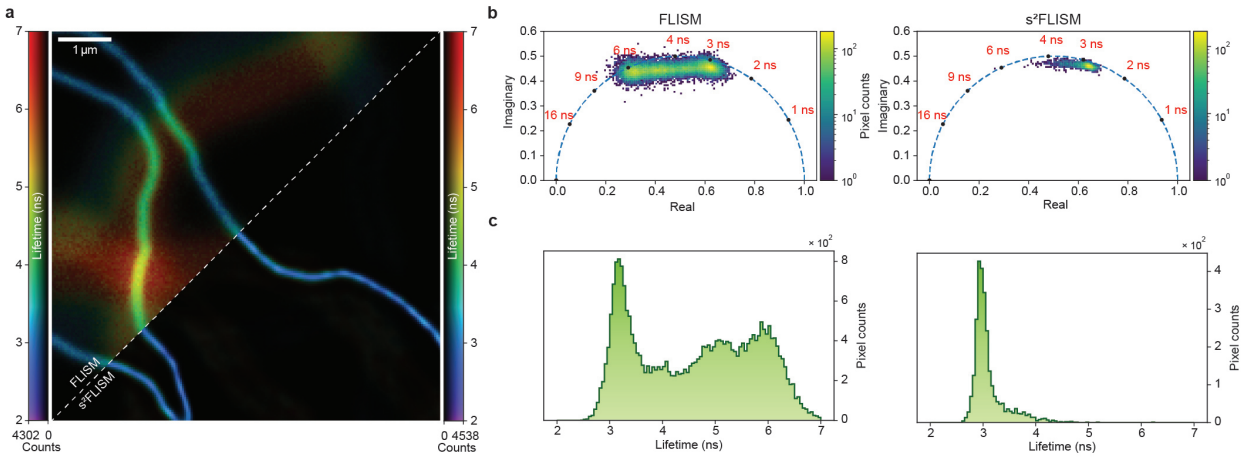

Figure 4.3: **Fluorescence lifetime imaging with s²ISM. a**, simulation of tubulin filaments with different lifetime values (NA $= 1.4$, $\lambda = 640\,\mathrm{nm}$). **b** The in-focus filaments have a lifetime value of $\tau = 3\,\mathrm{ns}$, while the out-of-focus ($z = 720\,\mathrm{nm}$) filaments have a lifetime value of either $\tau = 3\,\mathrm{ns}$ or $\tau = 6\,\mathrm{ns}$.

(Fig. 4.3a). The fractional components of the emitters in a thick sample depend on the $\boldsymbol{x}_d$ coordinate, following the fingerprint distribution. Such information can be exploited to remove the defocused decays from the temporal dimension. The SPAD array collects multiple fluorescence decays, one for each sensitive element. Each decay is convolved with a unique IRF and modulated by the fingerprint. The role of s²FLISM is to fuse the decays into one, compensating for the IRFs, and decomposing the decays into two components – in-focus and out-of-focus – according to the fingerprint modulation (Fig. 4.2). Furthermore, the spatial reconstruction takes place simultaneously, exploiting the redundancy between neighboring pixels, boosting the SNR of the fluorescence decay per pixel, and making lifetime estimation more robust.

A key point of our generalization is that we do not impose an exponential model, paving the way for broader applications where the fluorescence decay follows more complex dynamics – such as time-resolved STED microscopy [7, 8]. We validated the generalization of s²ISM to fluorescence lifetime on the simulations shown in Fig. 4.3. In the reconstructed image, the out-of-focus filaments have a highly suppressed intensity, enabling a more robust estimation of the lifetime of the in-focus fluorescence (Fig. 4.3a). Indeed, the phasor plot and the lifetime histogram of the FLISM image depict multiple components, while only a single component remains after the s²ISM reconstruction (Fig. 4.3b). Importantly, removing out-of-focus light corresponding to a different lifetime value using phasor decomposition [5] or related techniques would have required prior knowledge of the lifetime distribution of the sample, which is not needed with our method. Additionally, s²ISM is also capable of removing background fluorescence with an identical lifetime to that stemming from the focal plane, a task that would be infeasible with the sole measurement of the decay without array detection.

Experimental validation

We experimentally validated our approach by collecting the complete fluorescence decay for each scan point and each detector element using a multi-channel digital frequency domain (DFD) acquisition scheme, depicted before.

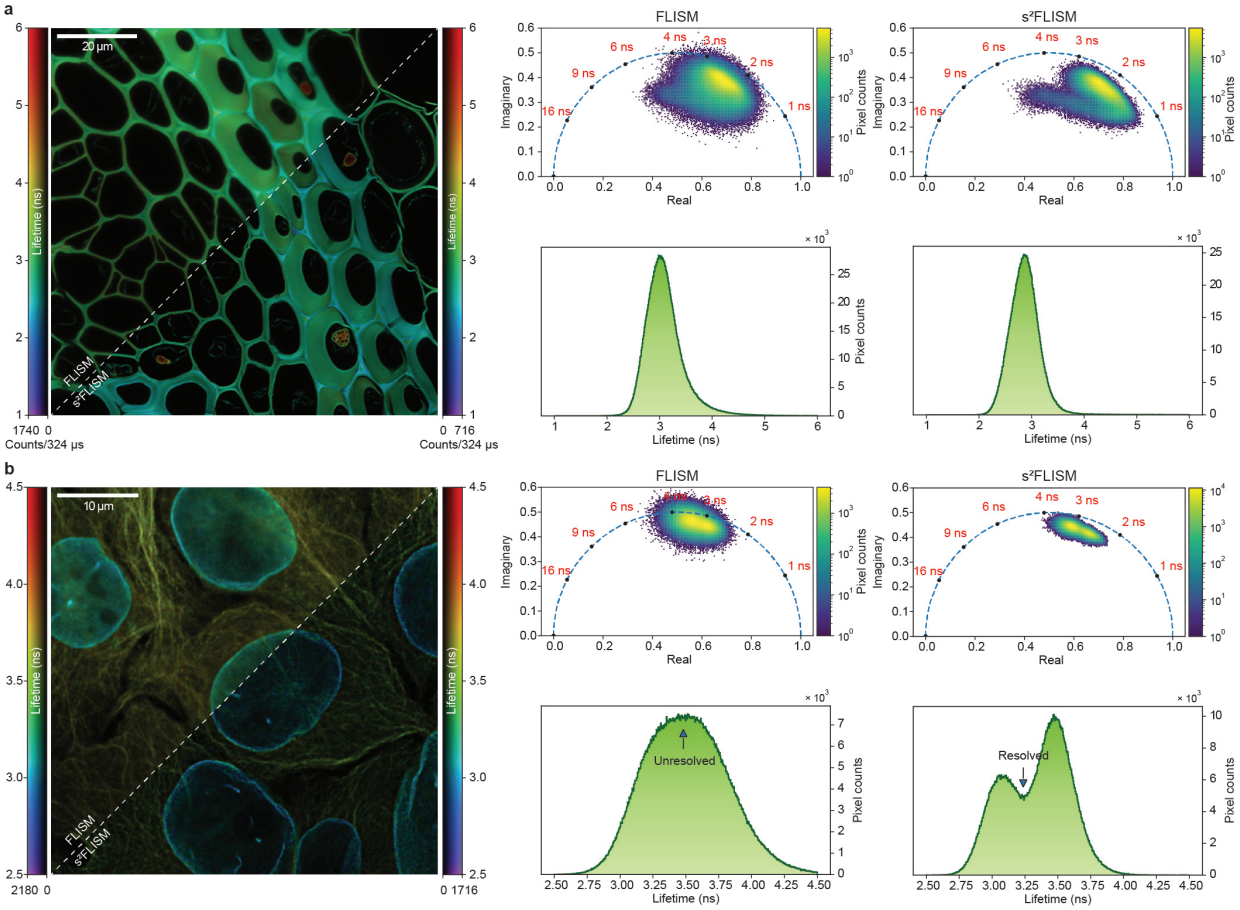

Figure 4.4: **Fluorescence lifetime imaging with s²ISM. a**, experimental image of a rhizome of Convallaria Majalis stained with acridine orange, excited with $\lambda = 488\,\mathrm{nm}$. **b**, experimental image of HeLa cells with tubulin stained with STAR RED ($\tau = 3.4\,\mathrm{ns}$) and lamin-A on the nuclear membrane stained with STAR 635 ($\tau = 2.8\,\mathrm{ns}$). Both fluorophores are excited with the same source at $\lambda = 640\,\mathrm{nm}$. The intensity of each image is normalized to its maximum. The phasor plots and the histograms are thresholded at 5% of the maximum intensity of the corresponding image. Lifetime values are calculated from the magnitude of the phasors [6].

In time-resolved measurements, it is crucial to maintain a high SNR at each pixel to ensure the accurate reconstruction of temporal decay profiles. At the same time, the pile-up effect and saturation might distort the fluorescence dynamics. Consequently, in the following experiments, we used low excitation power and longer pixel dwell times than in previous experiments (Fig. 4.4b). We acquired the dataset of a rhizome of Convallaria Majalis stained with a single fluorophore, whose lifetime is highly sensitive to the environment. We compared the conventional FLISM and

s^2FLISM reconstruction. As expected, this latter shows all the previously demonstrated benefits – such as improved resolution and optical sectioning – and also enables a more robust phasor analysis and estimation of the lifetime. Indeed, the phasor plot and lifetime histogram of Fig. 4.4b (Fig. C.12) have a narrower spread, indicating a more accurate analysis of the fluorescence dynamics due to both the improved SNR and the reduced axial cross-talk.

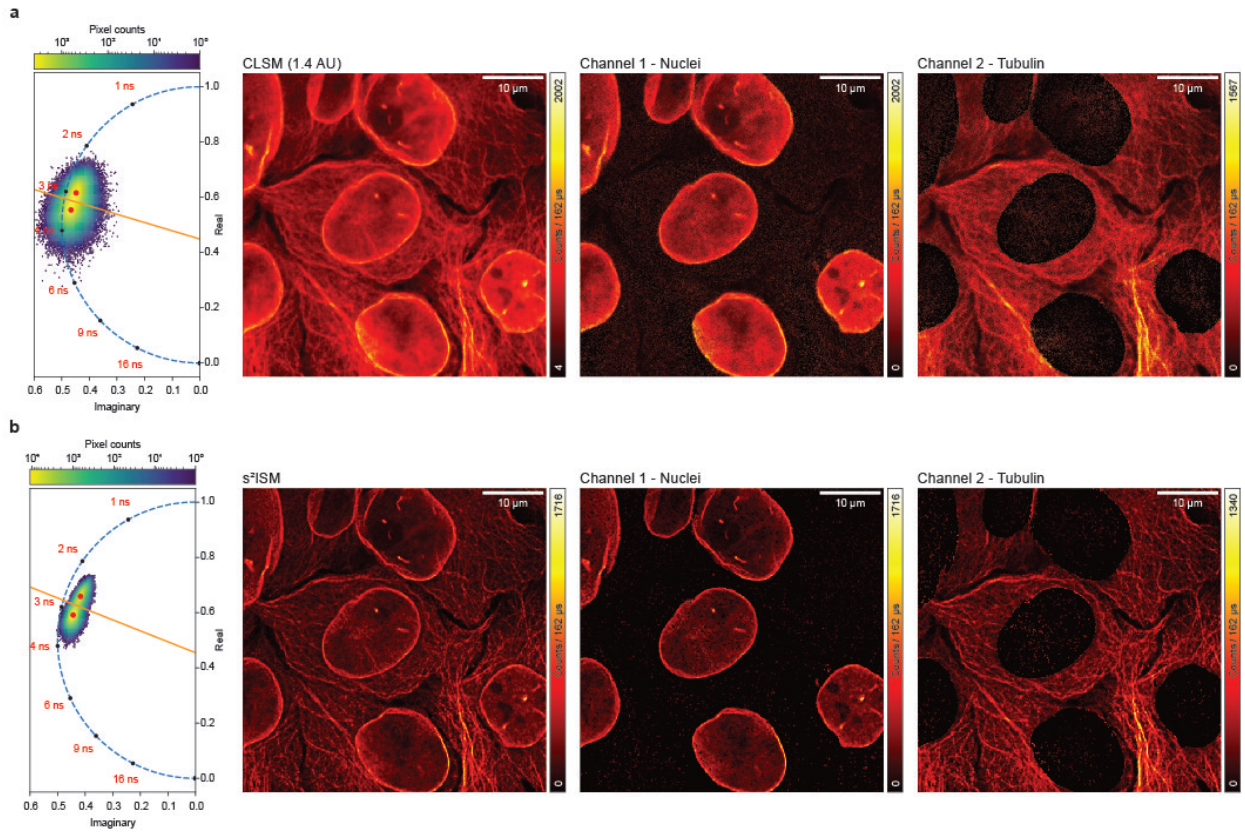

Figure 4.5: **Phasor segmentation.** The phasor cloud is fit to a Gaussian mixture with two components. Using Euclidean distance, each phasor is classified to the closest Gaussian component. The decision boundary is shown as an orange line. The results of the segmentation on the confocal and s^2ISM images are shown in **a** and **b**, respectively. The image quality enhancement coming from s^2ISM enables a more reliable and robust segmentation.

Such benefits are of particular interest when different structures are labeled with fluorophores with different lifetimes. In such a scenario, the correct lifetime estimation is of paramount importance for image segmentation or related tasks. To demonstrate the benefits of s^2FLISM in this context, we imaged a sample of HeLa cells with tubulin filaments and nuclei stained with two fluorophores with the same absorption spectrum, but different lifetime values (Fig. 4.4c). Therefore, a time-resolved measurement enables multi-target imaging with a single excitation laser and a single detector, as long as the different lifetime values can be correctly discriminated. As shown in Fig. 4.4c (Fig. C.13), in the conventional FLISM image, it is possible to distinguish two different populations of emitters. However, the phasor and lifetime histograms are broad, and individual distributions cannot be resolved. With the s^2FLISM reconstruction, the two populations

are clearly identified both in the phasor plot and in the lifetime histogram. Therefore, segmenting the image becomes an easier task thanks to s²FLISM (Fig. 4.5).

4.2.2 s²ISM for Pulse-Interleaved Fluorescence Lifetime Unmixing (PIE-flus²ISM)

At this stage, a careful examination of the mathematical derivation underlying the s²ISM algorithm reveals that optical sectioning was never explicitly assumed as an inherent advantage of incorporating detector-coordinate information. Rather, the framework is fundamentally general: in principle, any information encoded along the detector coordinates of the raw dataset can be recovered. Consequently, if the detector coordinates encode the temporal decay characteristics of distinct fluorescent species within the specimen, this framework admits a natural and novel extension of s²ISM —namely, s²ISM for fluorescence lifetime unmixing. This variant operates on the same raw spatio-temporal data but incorporates prior knowledge of temporal decay profiles to unmix the contributing species effectively. Moreover, flus²ISM framework can be further extended to develop an even more powerful version of spatio-temporal s²ISM, capable of simultaneously demixing species based on their lifetimes and rejecting out-of-focus contributions from the spatio-temporal data. In the Pulse-Interleaved Excitation [9] regime, multiple excitation sources are alternated so that the fluorescence emission from one pulse finishes before the next excitation pulse occurs. This timing ensures that the excitation source responsible for each detected photon is precisely known. The technique has broad applications across various fluorescence methods.

We further generalized the flus²ISM algorithm to the PIE regime acquisition. Namely, in this scenario, we would like to reconstruct the structures with separated emission spectra, where we indexed in the following forward model with P the number of pulses, and $[0, T]$ the timespan of acquisition.

$$i(\boldsymbol{x}_s, t|\boldsymbol{x}_d) = \sum_{k=0}^{K} \sum_{p=0}^{P-1} \sum_{m=1}^{M} o_{m,k}(\boldsymbol{x}_s) * h_{m,k,p}(\boldsymbol{x}_s, t|\boldsymbol{x}_d) \tag{4.6}$$

where

$$h_{m,k,p}(\boldsymbol{x}_s, t|\boldsymbol{x}_d) := h_k(\boldsymbol{x}_s|\boldsymbol{x}_d) \cdot h_{m,p}(t|\boldsymbol{x}_d) \tag{4.7}$$

and

$$h_{m,p}(t|\boldsymbol{x}_d) := IRF\left(t - p\frac{T}{P}|\boldsymbol{x}_d\right) * d_m(t) \tag{4.8}$$

and $d_m(t)$ the decay associated with the m-th fluorophore dyeing the sample, namely encoding the m-th lifetime decay τ_m as in Eq. 4.12. As we can notice, the temporal IRFs are shifted according to the moment in which the p-th excitation pulse occurs on the sample.

Synthetic validation

We simulated a phantom scenario in which we have a train of two pulses in excitation, one at $\lambda^1_{exc} = 488\,\mathrm{nm}$ and the other at $\lambda^2_{exc} = 560\,\mathrm{nm}$, emitting then in the green $\lambda^1_{em} = 510\,\mathrm{nm}$, and in the orange $\lambda^2_{exc} = 600\,\mathrm{nm}$. The first pulse is occurring at $\sim \frac{T}{3}$, the second one at $\sim \frac{2T}{3}$, with T the total acquisition time. As we can see from the temporal profile in the figure 4.6, we are effectively inducing two spikes and the relative temporal decays on the synthetic sample. In the latter, we impose two tubulin populations, both in focus and out of focus, one with a characteristic fluorescence decay of $6\,\mathrm{ns}$ and the other with $3\,\mathrm{ns}$. As we can see, also in the PIE, we are able to eliminate the cross-talk between the two species properly while enhancing both the optical sectioning capabilities and lateral resolution.

Experimental reconstruction

We tested this novel method on an experimental acquisition in the PIE regime of a sample composed of two structures with separated emission spectra, as we can see in Fig.4.7b. The first structure is represented by HeLa cell tubulin dyed with Atto647N, while the second is represented by TOMM20 dyed with Alexa Fluor 488 (Fig.4.7a). The two different fluorophores are excited separately. The total timespan of acquisition is said to be T; we have the first excitation pulse at the time $\sim \frac{T}{3}$ with the blue laser ($\lambda_{exc} = 488\,\mathrm{nm}$), and the second excitation at the time $\sim \frac{2T}{3}$ with the red laser ($\lambda_{exc} = 640\,\mathrm{nm}$) but the non-negligible cross-talk causes a small spike in the temporal kernel of the second species when the first one gets excited and vice versa (Fig.4.7b).

4.2.3 s²ISM for Fluorescence Lifetime Unmixing (flus²ISM)

By setting the index $P = 1$, corresponding to an acquisition with a single pulse, the problem simplifies to a pure unmixing algorithm. In this limiting case, the associated forward model is obtained as a natural specialization of the general PIE formulation.

The flus²ISM forward model, which encodes both types of information, is depicted here:

$$i(\boldsymbol{x}_s, t|\boldsymbol{x}_d) = \sum_{k=1}^{K}\sum_{m=1}^{M} o_{k,m}(\boldsymbol{x}_s) * h_{m,k}(\boldsymbol{x}_s, t|\boldsymbol{x}_d) \tag{4.9}$$

where we indexed with k the axial plane and with m the lifetime species. By inverting this model, we derive what we call flus²ISM, that is, giving as output an intensity-based estimation of the unknown object for each m and k. We have to highlight that $h_{m,k}(\boldsymbol{x}_s, t|\boldsymbol{x}_d)$ is no longer intended as in Eq. 4.2, namely the concatenation of spatial PSFs and temporal IRFs. As hinted before, in this new unmixing framework, we need to encode in the forward model the *a priori* information of the lifetime values of the fluorophore dyeing the specimen. So, we can write:

$$h_{m,k}(\boldsymbol{x}_s, t|\boldsymbol{x}_d) := h_k(\boldsymbol{x}_s|\boldsymbol{x}_d) \cdot g_m(t|\boldsymbol{x}_d) \tag{4.10}$$

with $h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)$ the classical axial-dependent spatial PSFs of the ISM setup, and with

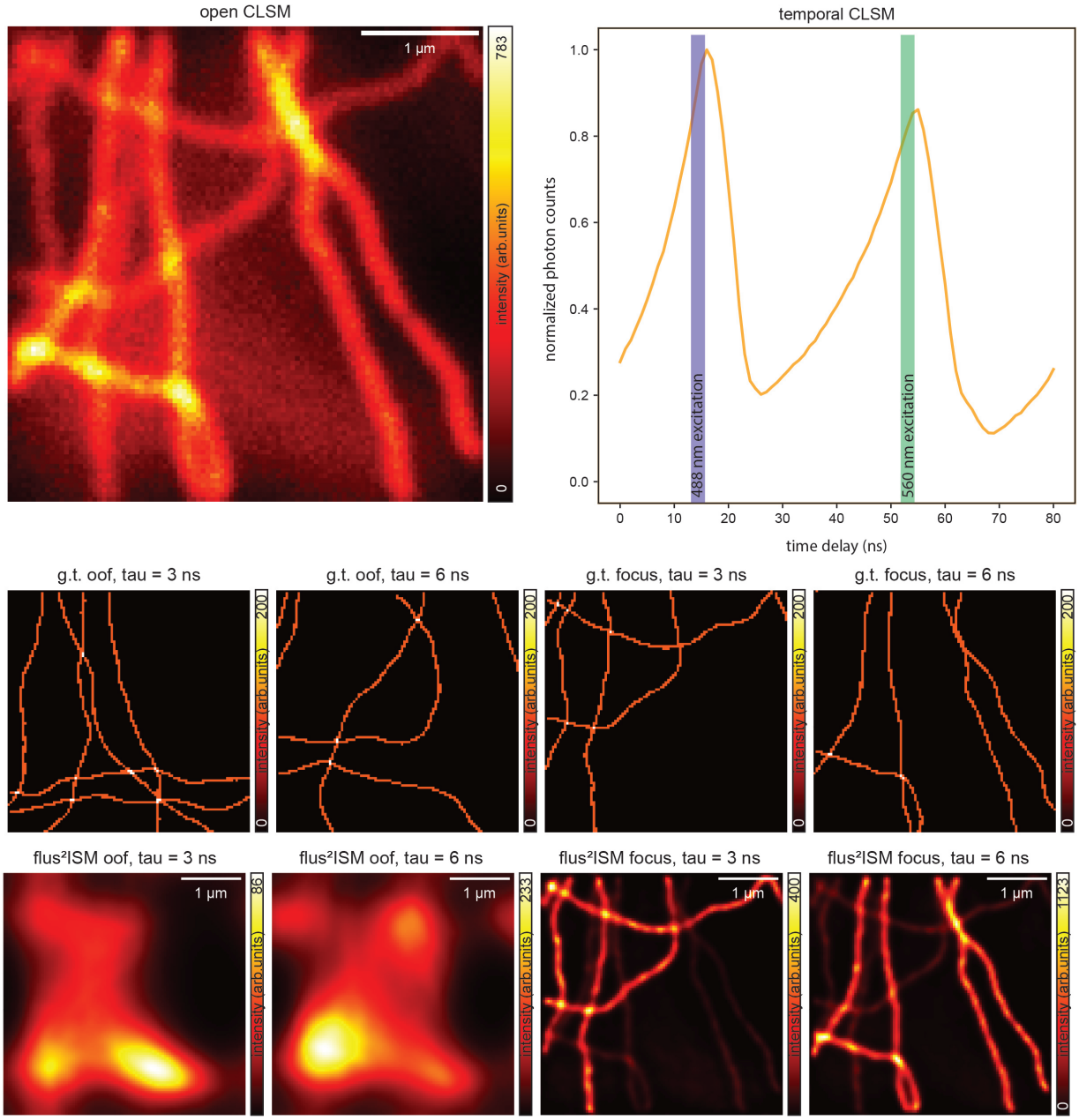

Figure 4.6: **PIE-flus^{2}ISM on synthetic data.**First line, we have the raw spatial and the raw temporal open CLSM data. We highlighted - in the temporal raw data - when the different excitation pulses are occurring in the lifespan of the experiment and the relative excitation wavelength. Second line, we have the list of in-focus and out-of-focus GT structures, with the corresponding different lifetimes. Third line, flus^{2}ISM reconstruction with lifetime-based structural unmixing, for both in-focus and out-of-focus.

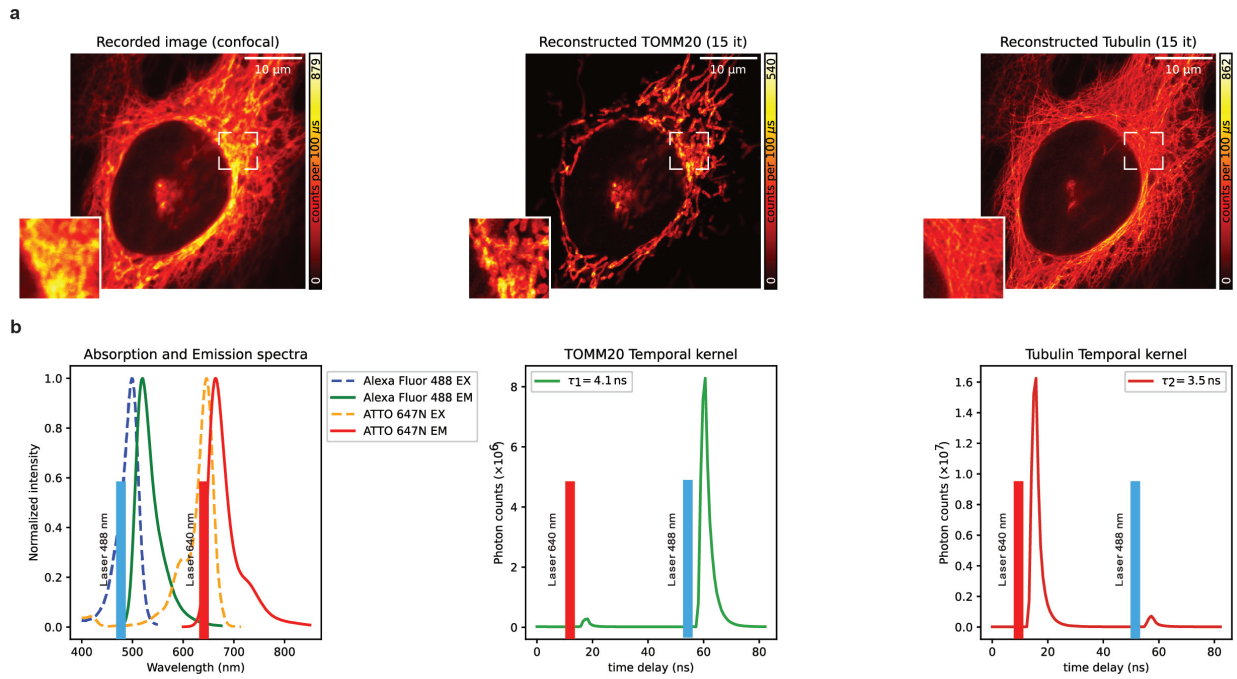

Figure 4.7: **Fluorescence lifetime unmixing with s²ISM in experimental PIE aquisition.** Unmixing two structures with separated emission spectra represented in **a**. The first one is represented by HeLa cell tubulin dyed with Atto647N, while the second on by TOMM20 dyed with Alexa Fluor 488. The two pulses of excitation occur separately in time, as we can appreciate from the time trace in **b**. The first excitation is performed at $488\,\mathrm{nm}$,while the second one at $640\,\mathrm{nm}$.

$$g_m(t|\boldsymbol{x}_d) = IRF(t|\boldsymbol{x}_d) * d_m(t) \tag{4.11}$$

namely the convolution between the temporal kernel and the *a priori* information characterizing the labeling fluorophores, thus

$$d_m(t) = e^{-\frac{t}{\tau_m}} \tag{4.12}$$

where τ_m for $m = 1, \ldots, M$ the fluorophores's lifetime.

Synthetic validation

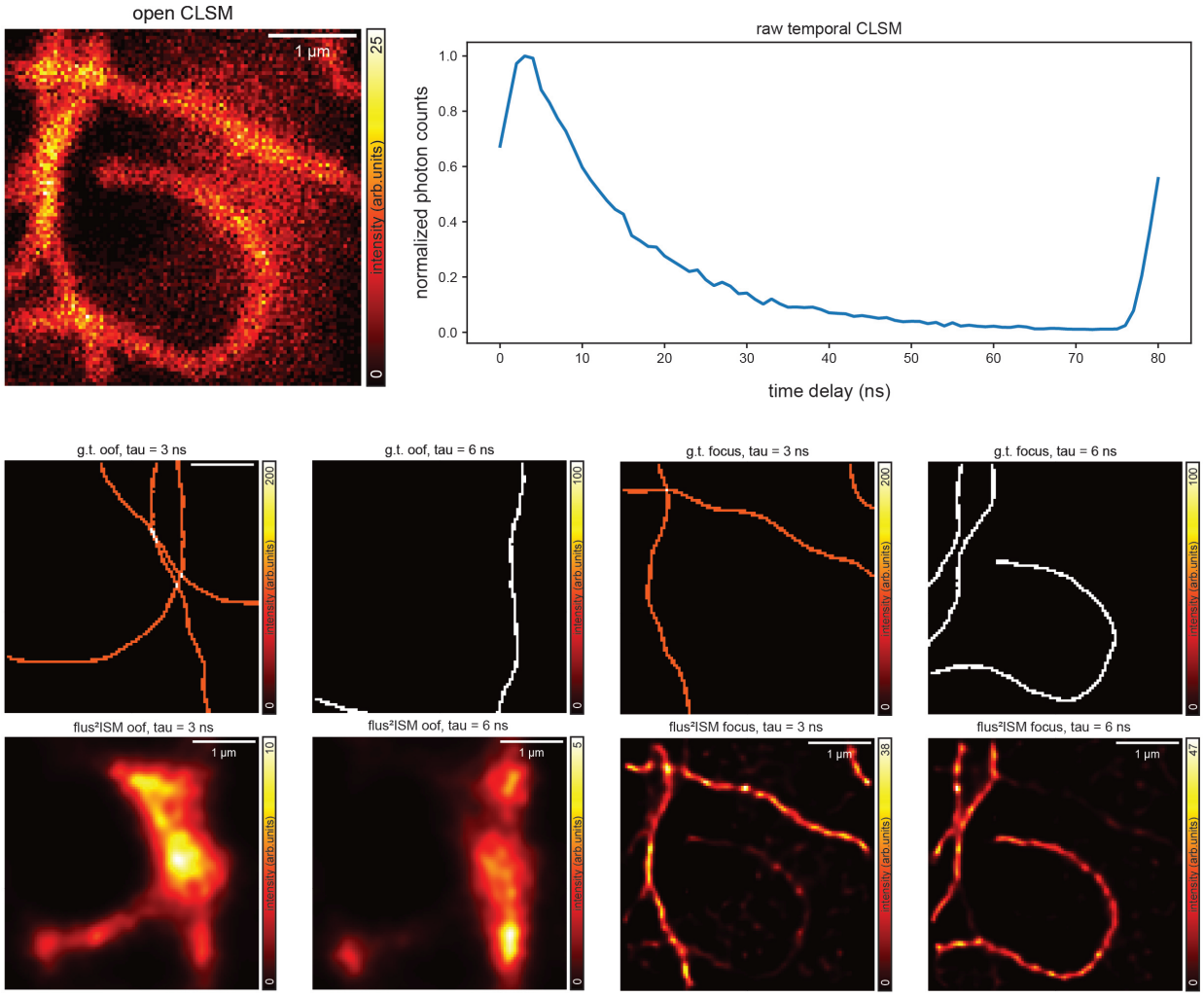

Figure 4.8: **Fluorescence lifetime unmixing with s²ISM.** First line, we have the raw spatial and the raw temporal open CLSM data. Second line, we have the list of in-focus and out-of-focus structures, depending on the different lifetimes. Third line, flus²ISM reconstruction with lifetime-based structural unmixing, for both in-focus and out-of-focus.

We created a thick phantom sample containing both in-focus and out-of-focus emitters. These emitters are divided into two populations, characterized by distinct temporal decay times of 3 ns and 6 ns, respectively. In Fig. 4.8 top-image, we can see the phantom raw spatio-temporal dataset. Then we have the comparison between the ground truth structures composing the ISM data and the respective flus²ISM reconstructions. As for classical s²ISM, background reconstructions have to be discarded due to their very low resolution. In-focus ones reliably estimate the imposed ground-truths, showing the algorithm's amazing capabilities in unmixing structures based on their lifetimes. Still, in this scenario, we have to properly stop the iterative routine to avoid spatial noise amplification.

Experimental reconstruction

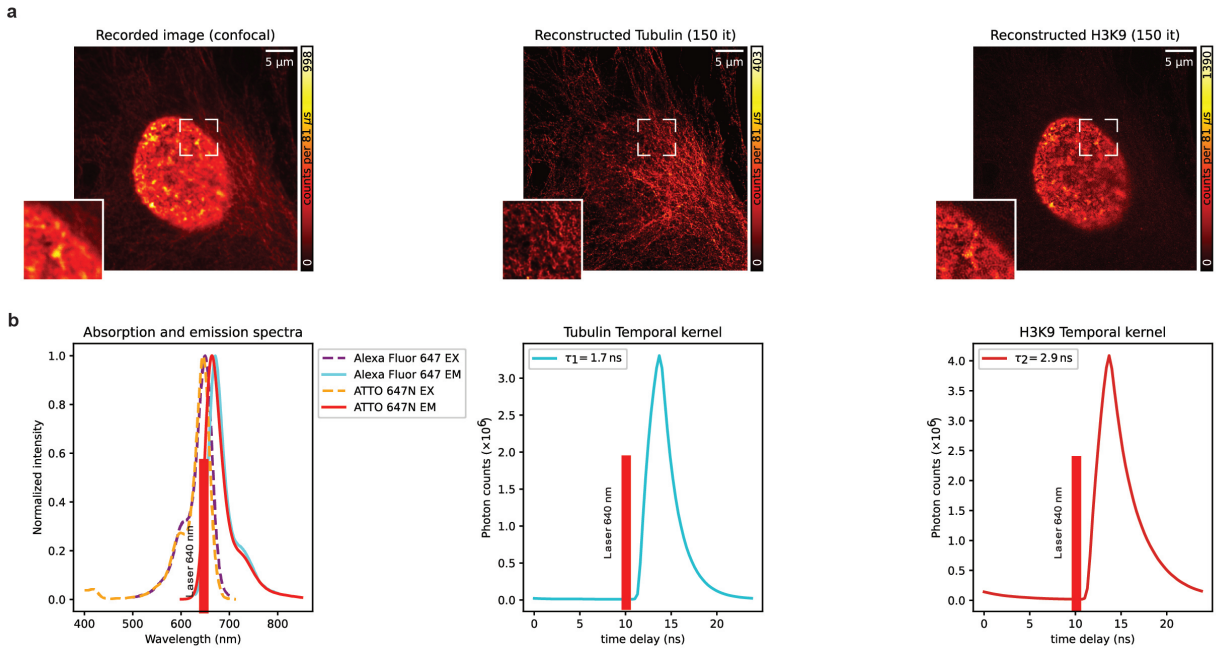

Figure 4.9: **s²FLISM separating two thin structures with overlapping emission spectra.** In **a**, the sample structures. The first structure is represented by HeLa cells' Tubulin dyed with Alexa Fluor 647. The second one by HeLa cells' H3K9 dyed through ATTO647N. Excitation is single, performed at $640\,\mathrm{nm}$. We can appreciate the spatial separation of the two, based on the two different lifetimes inherited by the different labeling fluorophores. Both structures are supposed to be thin. In **b**, we depict the overlapped absorption and emission spectra of the two fluorophores. Additionally, the lifetime values of the individual fluorophores are shown, together with the moment in the time-domain in which the excitation laser impinges on the sample.

We applied flus²ISM to an experimental dataset. Specifically, we analyzed a sample composed of two structures: HeLa cells' tubulin labeled with Alexa Fluor 647 and HeLa cells' H3K9 labeled with ATTO647N. In this scenario, we suppose the experimental sample to be thick. Thus, looking at the forward model in eq. 4.9, we dropped the axial dimension k, asking the algorithm just to perform lifetimes unmixing on structures, with no optical sectioning enhancement. We can appreciate from Fig. 4.9 how our algorithm effectively unmixes the cytoskeleton structure from the histone one, even when they are overlapping with each other, finding again the great benefits that we claimed in the synthetic validation scenario.

Conclusion of Chapter 5

This chapter extended the s²ISM framework to time-resolved imaging by incorporating photon arrival times and the instrument response function into the forward model. The resulting s²ISM approach enables super-resolved, optically sectioned fluorescence lifetime imaging directly from TCSPC data. Furthermore, lifetime-based unmixing strategies were introduced, allowing the sep-

aration of multiple fluorophores based on temporal signatures rather than spectral properties. Synthetic and experimental results demonstrated improved lifetime estimation accuracy and spatial resolution, highlighting the advantages of a joint spatio-temporal reconstruction framework. Furthermore, we demonstrated that the s²ISM framework is not inherently tied to optical sectioning, but is instead fundamentally general, enabling the recovery of any information encoded along the detector coordinates. The s²ISM framework naturally extends to flus²ISM, namely fluorescence lifetime unmixing.

References

[1] Maha Rahim, Jinghui Zhao, Hinesh Patel, Hauna Lagouros, Rajesh Kota, Irma Fernandez, Enrico Gratton, and Jered Haun. Phasor analysis of fluorescence lifetime enables quantitative multiplexed molecular imaging of three probes. *Analytical Chemistry*, 94, 10 2022.

[2] Marco Castello, Giorgio Tortarolo, Mauro Buttafava, Takahiro Deguchi, Federica A. Villa, Sami V. Koho, Luca Pesce, Michele Oneto, Simone Pelicci, Luca Lanzanó, Paolo Bianchini, Colin J. R. Sheppard, Alberto Diaspro, Alberto Tosi, and Giuseppe Vicidomini. A robust and versatile platform for image scanning microscopy enabling super-resolution flim. *Nature Methods*, 16:175 – 178, 2019.

[3] Alessandro Zunino, Eli Slenders, Francesco Fersini, Andrea Bucci, Mattia Donato, and Giuseppe Vicidomini. Open-source tools enable accessible and advanced image scanning microscopy data analysis. *Nature Photonics*, 17:457–458, 6 2023.

[4] Fernando Caprile, Luciano A. Masullo, and Fernando D. Stefani. Pyfocus – a python package for vectorial calculations of focused optical fields under realistic conditions. application to toroidal foci. *Computer Physics Communications*, 275, 6 2022.

[5] Giorgio Tortarolo, Alessandro Zunino, Simonluca Piazza, Mattia Donato, Sabrina Zappone, Agnieszka Pierzyńska-Mach, Marco Castello, and Giuseppe Vicidomini. Compact and effective photon-resolved image scanning microscope. *Advanced Photonics*, 6, 1 2024.

[6] Michelle A. Digman, Valeria R. Caiolfa, Moreno Zamai, and Enrico Gratton. The phasor approach to fluorescence lifetime imaging analysis. *Biophysical Journal*, 94:L14–L16, 1 2008.

[7] Luca Lanzanò, Iván Coto Hernández, Marco Castello, Enrico Gratton, Alberto Diaspro, and Giuseppe Vicidomini. Encoding and decoding spatio-temporal information for super-resolution microscopy. *Nature Communications*, 6:6701, 4 2015.

[8] Giorgio Tortarolo, Yuansheng Sun, Kai Wen Teng, Yuji Ishitsuka, Luca Lanzanó, Paul R. Selvin, Beniamino Barbieri, Alberto Diaspro, and Giuseppe Vicidomini. Photon-separation to enhance the spatial resolution of pulsed sted microscopy. *Nanoscale*, 11:1754–1761, 2019.

[9] Barbara K Müller, Evgeny Zaychikov, Christoph Bräuchle, and Don C Lamb. Pulsed interleaved excitation. *Biophysical journal*, 89(5):3508—3522, November 2005.

5

Regularization of s^{2}ISM

Inverse problems involve determining unknown causes from observed effects. Typically related to direct problems—where future states are predicted from known conditions and governing laws—inverse problems aim to recover past states or system parameters from observations. Applications often seek either to infer past conditions or to guide systems toward desired outcomes.

However, inverse problems frequently lack well-posedness: they may not have unique or stable solutions, and small variations in the data can lead to large deviations in the results [1]. These challenges classify them as ill-posed problems. Overcoming such difficulties often requires the use of regularization methods, which introduce additional constraints or prior information to restore stability. In many cases, solving an ill-posed problem results in an iterative algorithm, where one of the most powerful regularization strategies is determining the optimal number of iterations before the algorithm begins to overfit the noise. The absence of ground truth (GT) data, however, makes it difficult to identify the appropriate stopping point for the iterative process. Mathematical tools capable of predicting a suitable range of iterations to reach the optimal estimator without GT information [2] have been proposed, but they are often challenging to implement in practice. Explicit regularization enables robust recovery by incorporating prior information about the unknown object directly into the optimization framework. Plug-and-Play denoisers [3], instead, are embedded directly in the minimization problem, requiring very little prior information on the noise statistics affecting the raw data, leading to overfitting mitigation as noise explosion control. They do not require any prior structural knowledge. Also, Regularization by Denoising (RED) has been proposed [4]. Unlike PnP, RED incorporates the denoiser directly into the explicit regularization term of the inverse problem, using an explicit, image-adaptive Laplacian-based functional. This formulation clarifies the objective and allows complete flexibility in choosing the optimization procedure. RED can employ any denoising algorithm, handle a wide range of inverse problems effectively, and guarantees convergence to a globally optimal solution. Consequently, the formulation and solution of inverse problems often reduce to determining the most appropri-

ate regularization strategy, depending on the specific application and the characteristics of the underlying signals.

5.1 Model-based regularizers

The most general updating step of the projected gradient descent algorithm is built in its additive form as:

$$o^{k+1}(\boldsymbol{x}_s) = \Pi_{o(\boldsymbol{x}_s)\geq 0}\left[o^k(\boldsymbol{x}_s) - \tau_k \sum_{\boldsymbol{x}_d} \nabla KL(i(\boldsymbol{x}_s|\boldsymbol{x}_d); o^k(\boldsymbol{x}_s) * h(\boldsymbol{x}_s|\boldsymbol{x}_d))\right] \tag{5.1}$$

where the $\Pi_{y\geq 0}$ is the projection operator on the non-negative axis of the quantity y. Our reconstruction framework is built on physical assumptions; in our scenario, we can drop the projection operator, given that we are working with intensity, or photon counts, always semi-positive defined quantities.

We can add to the loss function to be minimized some explicit regularisation $R_\lambda(\boldsymbol{x}_s)$ [5]. In this case, the minimizing problem associated with 2.26, will become:

$$o^*(\boldsymbol{x}_s) = \underset{o(\boldsymbol{x}_s)\geq 0}{\operatorname{argmin}} \sum_{x_d} KL(i(\boldsymbol{x}_s|\boldsymbol{x}_d); o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s|\boldsymbol{x}_d)) + \lambda R_\lambda[o(\boldsymbol{x}_s)] \tag{5.2}$$

and the iteration:

$$o^{k+1}(\boldsymbol{x}_s) = o^k(\boldsymbol{x}_s) - \tau_k \left(\sum_{\boldsymbol{x}_d} \nabla KL(i(\boldsymbol{x}_s|\boldsymbol{x}_d); o^k(\boldsymbol{x}_s) * h(\boldsymbol{x}_s|\boldsymbol{x}_d)) + \lambda \nabla R_\lambda[o(\boldsymbol{x}_s)]\right) \tag{5.3}$$

Sparsity-promoting regularization techniques are widely used in imaging and inverse problems to recover solutions with few non-zero components. Among these, ℓ_0 regularization represents the most direct approach to enforcing sparsity.

The ℓ_0-norm counts the number of non-zero elements in a vector. Although it is commonly referred to as a norm, it is technically a pseudo-norm, as it does not satisfy the homogeneity property nor the triangle inequality.

ℓ_0 regularization explicitly enforces sparsity by assuming that the true solution contains only a small number of non-zero coefficients when represented in a suitable basis or domain (e.g., the spatial domain, wavelet domain, or gradient domain). In imaging applications, this assumption corresponds to scenes composed of a few isolated structures or features against a zero background.

ℓ_0 regularization leads to a non-convex, discontinuous, and combinatorial optimization problem. Minimizing an ℓ_0-penalized functional requires searching over all possible subsets of non-zero coefficients, which rapidly becomes computationally intractable even for moderate problem sizes. Such problems are NP-hard and exhibit strong sensitivity to noise and initialization.

As a consequence, ℓ_0 regularization is rarely employed directly in practical imaging problems. For a vector $x = [x_1, x_2, \ldots, x_n]^T$, the ℓ_0 norm and consequently the regularizer, is defined as

$$R_\lambda[x] = \|x\|_0 = \#\{i \mid x_i \neq 0\},$$

which represents the number of non-zero elements in x. Equivalently, it can be written using the indicator function:

$$R_\lambda[x] = \|x\|_0 = \sum_{i=1}^{n} \mathbf{1}_{\{x_i \neq 0\}},$$

where $\mathbf{1}_{\{.\}}$ is equal to 1 if the condition is true and 0 otherwise.

The ℓ_1 norm of a vector promotes sparsity in a relaxed, convex form and is widely used in imaging and signal processing. Unlike the ℓ_0 pseudo-norm, the ℓ_1 norm is convex and computationally tractable.

For a vector x, the ℓ_1 norm is defined as

$$R_\lambda[x] = \|x\|_1 = \sum_{i=1}^{n} |x_i|.$$

This norm penalizes the absolute values of the coefficients, encouraging many components to be exactly zero while allowing non-zero components to remain. In imaging, ℓ_1 regularization is often applied to wavelet coefficients, gradients (leading to Total Variation regularization), or sparse representations of the image.

The ℓ_2 norm corresponds to the Euclidean norm and penalizes the energy of the vector. It is commonly used in Tikhonov regularization and other smoothing techniques.

For a vector xT, the ℓ_2 norm is defined as

$$R_\lambda[x] = \|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Unlike the ℓ_1 norm, the ℓ_2 norm does not promote sparsity. Instead, it encourages smooth, globally distributed solutions by penalizing large deviations across all components. This makes it highly effective for noise suppression, although it tends to blur sharp edges in images.

Some examples of differentiable regularisation terms based on the ℓ_2-norm are:

- Tikhonov of first order

  $$R_\lambda[o(\boldsymbol{x}_s)] = \frac{\lambda}{2}\|o(\boldsymbol{x}_s)\|_2^2$$

  This form of regularization promotes solutions that vary smoothly while still allowing gradual spatial changes. Compared to zeroth-order Tikhonov regularization, which penalizes the amplitude of the solution itself, first-order regularization better preserves overall intensity levels and reduces bias toward zero. However, it may still smooth out sharp edges or discontinuities, especially for large values of λ.

- Tikhonov of second order

  $$R_\lambda[o(\boldsymbol{x}_s)] = \frac{\lambda}{2}\sum \|\nabla o(\boldsymbol{x}_s)\|_2^2$$

Compared to first-order Tikhonov regularization, the second-order variant generally produces smoother reconstructions and reduces staircase artifacts. However, it may oversmooth fine details and edges if the regularization parameter λ is chosen too large. Consequently, careful parameter selection is crucial to balance noise reduction and resolution preservation.

- Total variation smoothed

$$R_\lambda[o(\boldsymbol{x}_s)] = \frac{\lambda}{2}\sum \|\nabla o(\boldsymbol{x}_s)\|_2 = \frac{\lambda}{2}\sum \sqrt{\nabla_x^2 o(\boldsymbol{x}_s) + \nabla_y^2 (o(\boldsymbol{x}_s)) + \epsilon^2}$$

Unlike first- or second-order Tikhonov regularization, which penalizes squared derivatives and therefore favors globally smooth solutions, TV regularization penalizes the ℓ_1-norm of the gradient. This key property enables TV regularization to strongly suppress small oscillations induced by noise while preserving large gradients associated with edges or discontinuities. As a consequence, TV regularization tends to produce piecewise constant or piecewise smooth solutions. While edges are well preserved, smoothly varying regions may exhibit the so-called *staircasing effect*, in which gradual intensity variations are approximated by flat regions separated by abrupt jumps. The smoothed version also mitigates this undesired effect.

Even if strongly used in the past, the explicit regularizers approach lacks generality: changing the sample or experimental conditions may require redefining the regularizer, which in turn can significantly alter the structure of the reconstruction algorithm.

5.2 Model-free regularizers

SPAD array detector is a high-speed detector that allows sampling the signal at a high temporal scale ($< 1\,\mu\mathrm{s}$). This permits us to collect a time-resolved raw dataset where noise realizations across different time bins are independent and identically distributed (i.i.d.). Furthermore, by integrating multiple time bins, we can adjust the SNR of the final raw ISM data according to our needs. We can exploit the noise realization dimension to reduce the noise level of the raw data immediately or to extract some new information from the specimen. In this study, we exploit this feature to mitigate the NO that can occur during unsupervised s²ISM reconstruction.

5.2.1 Gradient Consensus (GC)

We can leverage the SPAD's multirealization feature to implement custom early stopping criteria. Specifically, we integrate the raw data along the time dimension, ending up with a set of data sharing the same SNR. Then, we analyze each with s²ISM. By computing mean and variance along the realization axis, we estimate the SNR of the reconstruction at each iteration and pick the optimal estimator. Obviously, the SNR estimation consists of a map extended over $\boldsymbol{x}_s$ coordinate. In our framework, to have a single number showing the level of the image SNR, we

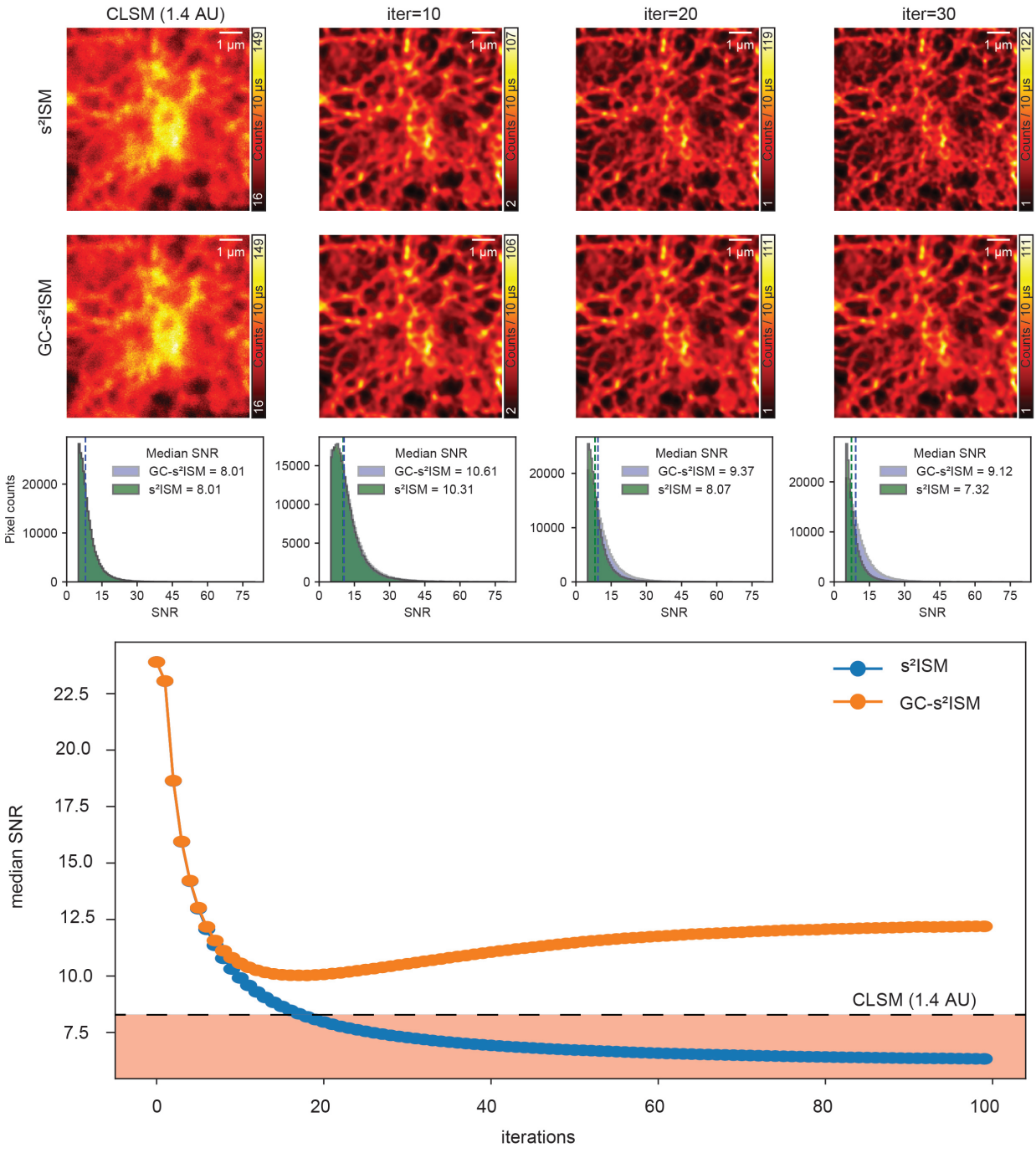

Figure 5.1: **Gradient Consensus mitigating noise overfitting in experimental reconstruction** First two lines depicting qualitative comparison between s^{2}ISM and its regularized version for a bunch of iterations. Bottom line plot showing the median SNR behaviour of the two versions, along the iterative routine. We can see that after $\sim$ 16 iterations, classical s^{2}ISM loses all the gained benefits, while the GC-s^{2}ISM version mitigates overfit, maintaining benefits.

shrink the $\boldsymbol{x}_s$ dimensions to extract the median SNR value for each iteration 5.1. This method is straightforward but has two disadvantages: being non-local and being low-biased.

Indeed, it requires splitting the raw data photon flux on various realizations, leading to an underestimation of the optimal iteration. This noise-estimation-based early-stop criterion has also been implemented within the GC framework. Namely, we launch s²ISM reconstruction on at least two data with i.i.d. noise realizations. At each iteration, we calculate the gradient on both updates. The angle amplitude between the latter provides an estimate of the noise degradation level in that neighborhood of the image, guiding the decision of whether or not to update it. GC is inherently a local stop criterion, bypassing the strong hypothesis of constant SNR across the entire dataset.

Assume to measure several realisations of the image $i_r(\boldsymbol{x}_s)$ with $r \in \{1, \dots, R\}$, each being an independent Poisson process. We define the sum of all the realizations as

$$i(\boldsymbol{x}_s) = \sum_{r=1}^{R} i_r(\boldsymbol{x}_s) \tag{5.4}$$

For each r, the forward model is the following

$$i_r(\boldsymbol{x}_s) \sim \mathcal{P}\{o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s)\} \tag{5.5}$$

where $\mathcal{P}$ indicates the Poisson distribution. The likelihood functional is

$$\mathbb{P}[i(\boldsymbol{x}_s)|\mathbf{o}(\boldsymbol{x}_s)] = \prod_r \prod_{x_s} \frac{[o(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s)]^{i_r(\boldsymbol{x}_s)} e^{-[o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s)]}}{i_r(\boldsymbol{x}_s)!} \tag{5.6}$$

The corresponding negative log-likelihood is

$$\mathcal{L}[\mathbf{o}(\boldsymbol{x}_s)] = \sum_r \sum_{\boldsymbol{x}_s} o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s) - i_r(\boldsymbol{x}_s) \cdot \log\{o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s)\} \tag{5.7}$$

where we neglected the constant terms. The solution of the inverse problem is given by

$$\hat{\boldsymbol{o}}(\boldsymbol{x}_s) = \underset{\boldsymbol{o}(\boldsymbol{x}_s)}{\arg\min}\, \mathcal{L}[\mathbf{o}(\boldsymbol{x}_s)] \tag{5.8}$$

The minimum is found by setting to zero the functional derivative of the log-likelihood. Using the fact that the adjoint operator of the convolution is the convolution with the mirror-reflected kernel, we find

$$\frac{\delta \mathcal{L}}{\delta o} = \sum_r h(-\boldsymbol{x}_s) * \left[1 - \frac{i_r(\boldsymbol{x}_s)}{o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s)}\right] \tag{5.9}$$

We assume the PSFs to be normalized as follows

$$\sum_{\boldsymbol{x}_s} \sum_r h(\boldsymbol{x}_s) = 1 \tag{5.10}$$

which implies

$$\sum_{\boldsymbol{x}_s} h(\boldsymbol{x}_s) = 1/R \tag{5.11}$$

Therefore, we rewrite the gradient as

$$\frac{\delta \mathcal{L}}{\delta o} = \sum_r \frac{\delta \ell_r}{\delta o} = \sum_r \left[\frac{1}{R} - h(-\boldsymbol{x}_s) * \frac{i_r(\boldsymbol{x}_s)}{o(\boldsymbol{x}_s) * h(\boldsymbol{x}_s)} \right] \tag{5.12}$$

We now assume $K = 2$ and check the degree of parallelism of the two partial gradients by calculating their inner product

$$\left\langle \frac{\delta \ell_1}{\delta o} \middle| \frac{\delta \ell_2}{\delta o} \right\rangle = \sum_{\boldsymbol{x}_s} \frac{\delta \ell_1}{\delta o}(\boldsymbol{x}_s) \frac{\delta \ell_2}{\delta o}(\boldsymbol{x}_s) = \left\| \frac{\delta \ell_1}{\delta o}(\boldsymbol{x}_s) \right\| \left\| \frac{\delta \ell_2}{\delta o}(\boldsymbol{x}_s) \right\| \cos\theta \tag{5.13}$$

We define the following local inner product to probe the local consensus of the two gradients

$$c(\boldsymbol{x}_s) = \sum_{\boldsymbol{x}} \frac{\delta \ell_1}{\delta o}(\boldsymbol{x}) w(\boldsymbol{x} - \boldsymbol{x}_s) \frac{\delta \ell_2}{\delta o}(\boldsymbol{x}) = \left[\prod_r \frac{\delta \ell_r}{\delta o}(\boldsymbol{x}_s) \right] * w(\boldsymbol{x}_s) \tag{5.14}$$

where w is a function with finite support. We define a mask function such that

$$m(\boldsymbol{x}_s) = \begin{cases} 1 & \text{if } c(\boldsymbol{x}_s) > 0 \\ 0 & \text{if } c(\boldsymbol{x}_s) \leq 0 \end{cases} \tag{5.15}$$

Finally, we reconstruct the image using the gradient descent method

$$o^{(j+1)} = o^{(j)} - \gamma^{(j)} \frac{\delta \mathcal{L}}{\delta o} \tag{5.16}$$

where

$$\gamma^{(j)}(\boldsymbol{x}_s) = m^{(j)}(\boldsymbol{x}_s) \cdot o^{(j)}(\boldsymbol{x}_s) \tag{5.17}$$

Therefore, the iterative reconstruction algorithm can be written in the following multiplicative form

$$o^{(j+1)} = o^{(j)} \cdot u^{(j)} \tag{5.18}$$

where the update function is

$$u^{(j)}(\boldsymbol{x}_s) = \begin{cases} h(-\boldsymbol{x}_s) * \dfrac{i(\boldsymbol{x}_s)}{o^{(j)}(\boldsymbol{x}_s) * h(\boldsymbol{x}_s)} & \text{if } m^{(j)}(\boldsymbol{x}_s) = 1 \\ 1 & \text{if } m^{(j)}(\boldsymbol{x}_s) = 0 \end{cases} \tag{5.19}$$

Generalization to ISM

Now we assume the object to be thick, imposing its dependency on the z-axis, and we consider the ISM scenario, namely, we introduce the detector coordinates $\boldsymbol{x}_d$ for the SPAD array detector.

Then we have that:

$$i(\boldsymbol{x}_s|\boldsymbol{x}_d) = \sum_{r=1}^{R} i_r(\boldsymbol{x}_s|\boldsymbol{x}_d) \tag{5.20}$$

and that for every realization $r \in \{1, \dots, R\}$:

$$i_r(\boldsymbol{x}_s|\boldsymbol{x}_d) = \sum_{k=1}^{K} o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d). \tag{5.21}$$

The corresponding negative log-likelihood is

$$\mathcal{L}[\mathbf{o}(\boldsymbol{x}_s, z)] = \sum_{r}\sum_{k}\sum_{\boldsymbol{x}_s}\sum_{\boldsymbol{x}_d} o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d) - i_r(\boldsymbol{x}_s|\boldsymbol{x}_d) \cdot \log\left\{\sum_k o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)\right\} \tag{5.22}$$

leading to

$$\frac{\delta\mathcal{L}}{\delta o_k} = \sum_r \sum_{\boldsymbol{x}_d} h_k(-\boldsymbol{x}_s|\boldsymbol{x}_d) * \left[1 - \frac{i_r(\boldsymbol{x}_s)}{\sum_k o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)}\right] \tag{5.23}$$

We assume the PSFs to be normalized as follows

$$\sum_{\boldsymbol{x}_s}\sum_{\boldsymbol{x}_d}\sum_r h_k(\boldsymbol{x}_s|\boldsymbol{x}_d) = 1 \quad \forall k \tag{5.24}$$

which implies

$$\sum_{\boldsymbol{x}_s}\sum_{\boldsymbol{x}_d} h_k(\boldsymbol{x}_s|\boldsymbol{x}_d) = 1/R \quad \forall k \tag{5.25}$$

and therefore that $\forall \quad k$

$$\frac{\delta\mathcal{L}}{\delta o_k} = \sum_{\boldsymbol{x}_d}\sum_r \frac{\delta\ell_r}{\delta o} = \sum_{\boldsymbol{x}_d}\sum_r \left[\frac{1}{R} - h_k(-\boldsymbol{x}_s|\boldsymbol{x}_d) * \frac{i_k(\boldsymbol{x}_s)}{\sum_k o_k(\boldsymbol{x}_s) * h_k(\boldsymbol{x}_s|\boldsymbol{x}_d)}\right] \tag{5.26}$$

The remaining calculations follow the same path as in the non-ISM scenario, since the detector coordinates are already saturated during the estimation process, before the gradient parallelism computations.

We embed this regularizer in the s²ISM framework, and looking at the spectral intensity in the intensity variance domain 5.2 we find out that this technique is properly shrinking down the high intensity close to the cutoff frequency in the Fourier domain, namely mitigating noise explosion during the reconstruction iterative process.

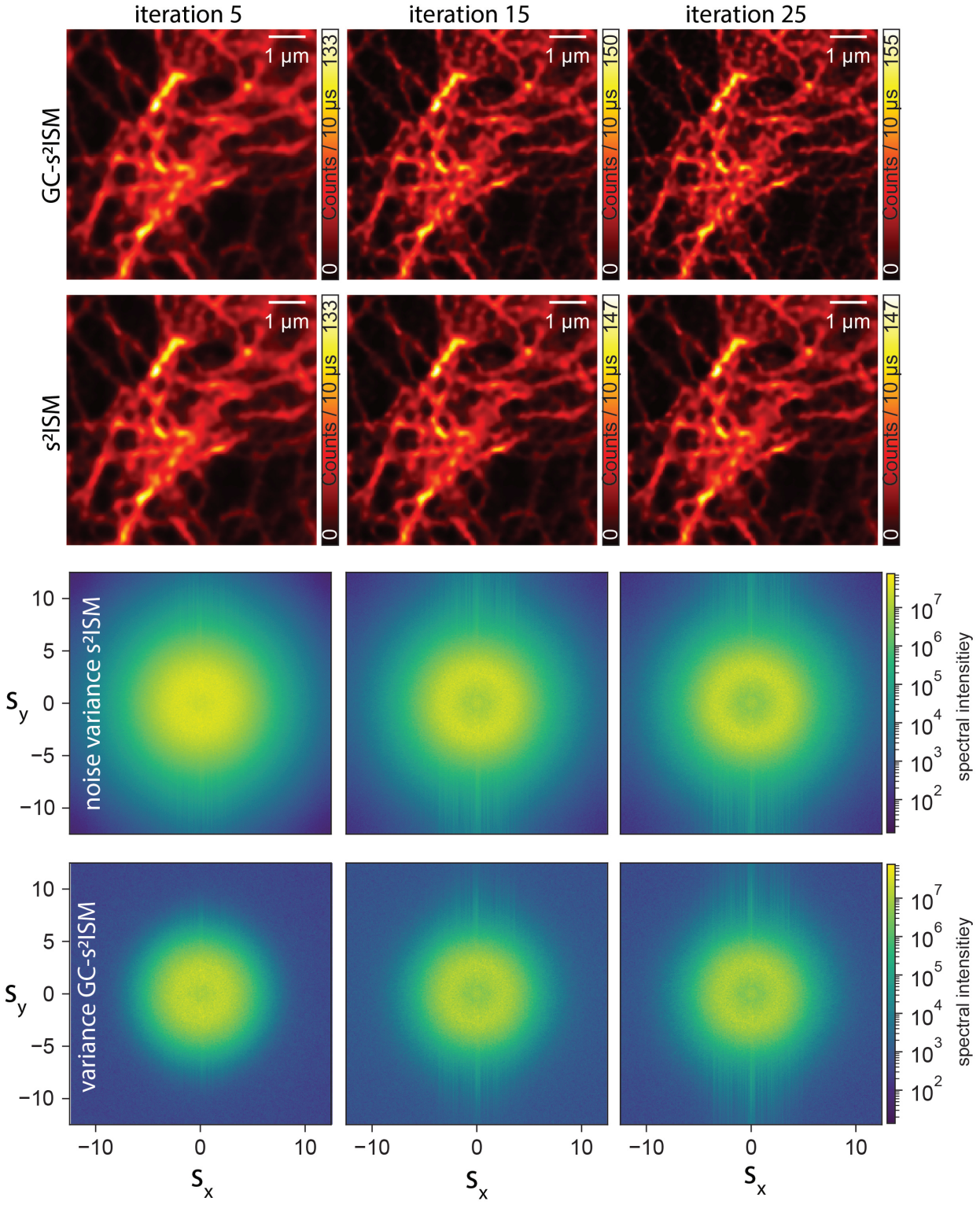

Figure 5.2: **Gradient Consensus mitigating high frequency explosion in the variance domain.** Variance spectral intensities are plotted in log-scale in order to better appreciate the difference along the iterations between the regularized and unregularized s^2ISM reconstruction

Cross-talk quantity characterization

A key step in the Gradient-Consensus framework is the computation of the cross-talk quantity.

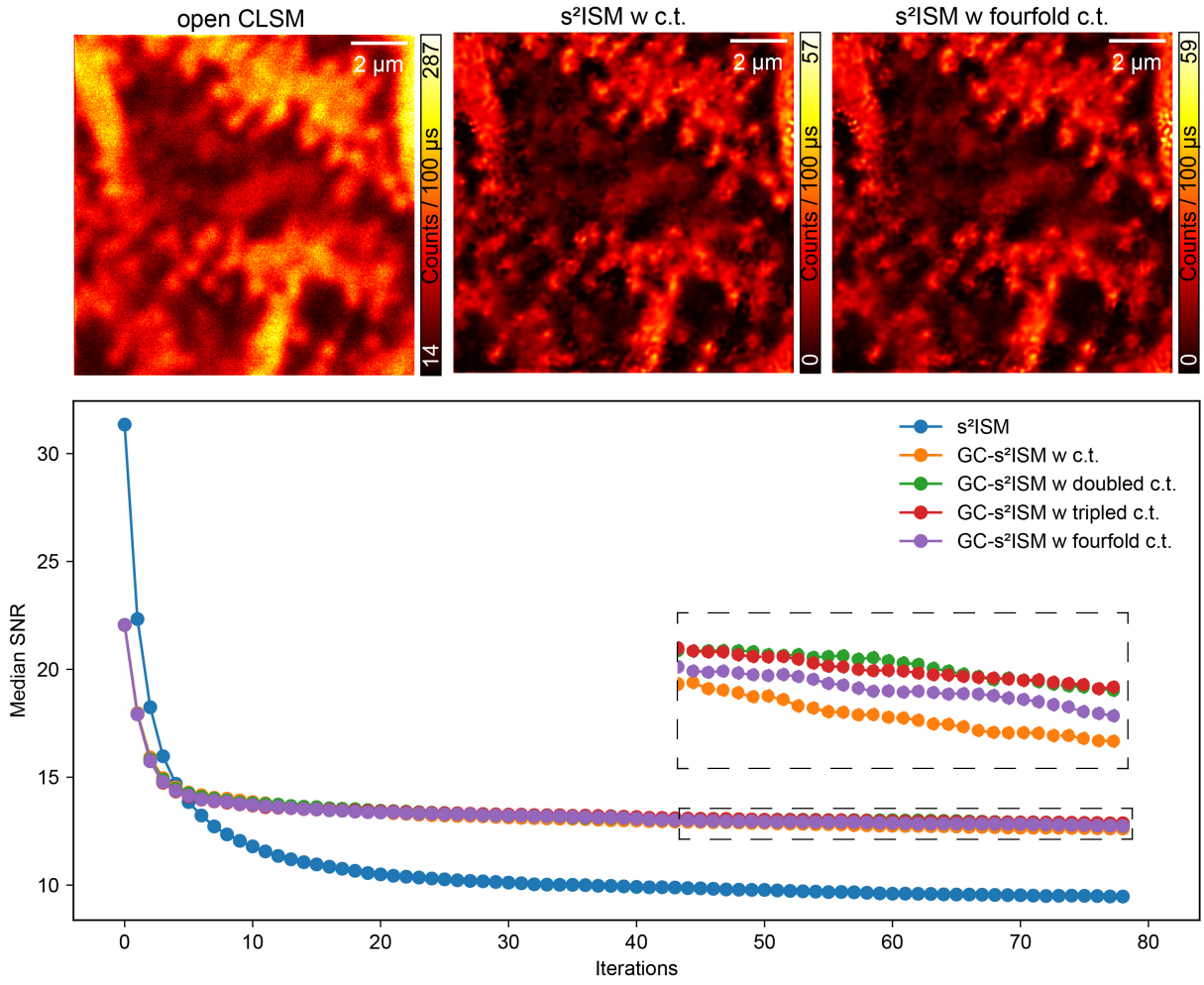

Figure 5.3: **Median SNR estimation for a Purkinje sample along the iterative routine of s²ISM and GC-s²ISM.** Different colors encode the dimension of the cross-talk quantity. With doubled cross-talk, we intend to perform twice the calculation depicted in 5.14, and so on.

After measuring the divergence between the two gradients, a mask function (Eq. 5.3) is constructed to determine where s²ISM iterations should proceed and where they $w(\boldsymbol{x}_s)$, which reflects how the degree of gradient parallelism at a given pixel spreads to its neighbors. For example, setting $w(\boldsymbol{x}_s) = \delta(\boldsymbol{x}_s)$ restricts the updating mask based on the gradient divergence evaluation merely to the specific pixel $\boldsymbol{x}_s$, without considering its neighbors.

Andrew York and James Manton, in their first work about GC [6], simply describes how the well-known RL algorithm can be used to treat blurring inversion, but can also be used to invert the action of any kind of linear operator:

'Note that Richardson-Lucy is often described as "deconvolution", but this is silly; H can be any linear operator. Convolution (blurring) is a linear operator, but so is summing, copying, cropping, etc. RL can be used for deblurring, but it can also calculate volumes from projections, fuse SIM data, unmix spectral crosstalk, process PALM data, etc. In fact, there's nothing about RL that's specific to fluorophores or photons; any linear measurement that's Poisson-noisy (e.g. CT scan, MRI, astronomy) can use RL (and therefore GC) for inference.'

For instance, he showed how modeling the cropping operator, its transpose describes the zero-padding operator. From this idea arises the cross-talk quantity idea. Let us define h as the blurring operator related to the PSF of the optical system. The forward model for our scenario simply lies in the convolution of the unknown object with the latter. The transpose of h, said h^t, then describes which region of the object could have illuminated these detectors' pixels. Defining a mask as the convolution of h with h^t should be the best representer of the forward-backward path for every emitter of the sample. Defining then the cross-talk as the convolution of the gradient product at each iteration with the mask $w(\boldsymbol{x}_s) = h^t(\boldsymbol{x}_s) * h(\boldsymbol{x}_s)$, we define with doubled cross-talk a double convolution of the gradient product with $w(\boldsymbol{x}_s)$ and so on for the others.

Here, Fig. 5.3, we tried to analyze how the dimension of that quantity affects the GC-s²ISM performances on a Purkinje sample, namely a cerebellum specimen, mostly featured by 'big' and thick structures. From the following plot, we can appreciate how effectively the cross-talk dimension affects the overfitting during the iterative process. For this particular scenario, tripled cross-talk seems to have slightly better attitudes with respect to the other tested dimensions. This analysis is merely qualitative, but for future works would be great to find a way to automatically define a cross-talk dimension based on the structures of the data itself. Maybe, a perspective we can envision a pixel-dependent cross-talk dimension pre-calibration for our GC-s²ISM.

5.3 Plug-and-Play regularizers

Plug-and-Play (PnP) denoisers are a class of algorithms that integrate powerful image denoising priors into iterative optimization frameworks for solving inverse problems. Instead of explicitly adding a regularization term into the loss function of the problem inversion, PnP methods replace the proximal operator in classical optimization algorithms with an external denoiser, effectively leveraging advanced denoising models (e.g., BM3D, DnCNN) as implicit priors. This approach enables high-quality reconstructions while maintaining flexibility across diverse imaging applications. However, ensuring convergence and providing theoretical guarantees remain challenging, as the incorporated denoiser may not correspond to a well-defined regularization. The concept of exploiting denoisers as priors for regularization was first proposed by Venkatakrishnan et al. [7], who developed the pioneering PnP Prior framework. Over the past decade, extensive research in this area has produced a wide range of highly effective algorithms [8], many of which achieve near-optimal performance [9].

5.3.1 Block-Matching 3D (BM3D)

For this work, we implemented the Block-Matching 3D (BM3D) PnP denoiser [10] within the iterative routine of s²ISM. BM3D was originally designed for pure Gaussian noise; our measurements are dominated by Poisson noise. To address this mismatch, we included a preprocessing step that transforms Poisson noise into an approximately Gaussian distribution using the Anscombe transform [11], followed by an inverse transform after the convergence of s²ISM. Although BM3D operates within the PnP framework, it still requires careful tuning of certain parameters, with the power spectral density (PSD) size (sigma) being particularly important for performance. Since sigma depends on numerous parameters determined by the experimental acquisition, it must be properly calibrated to achieve optimal results.

s²ISM-BM3D calibrated through N2S

Noise2Self (N2S) is a framework introduced in [12] for denoising high-dimensional data without requiring any prior knowledge about the underlying signal. The method relies solely on the assumption that noise is independent across different dimensions, while the true signal exhibits correlations. The authors introduced a class of functions called J-invariant functions, which estimate *a priori* the performance of a denoiser directly from noisy observations. This property enables the calibration of J-invariant versions of any parameterized denoiser, ranging from a single hyperparameter in a median filter to the millions of weights in a deep neural network.

We implement N2S in order to calibrate the PnP version of the BM3D denoiser, once it has been properly embedded in s²ISM. Namely, at every iteration, we perform a BM3D step on the in-focus reconstruction, but we would like to make the BM3D parameter estimation automatic during the iterations. Fig. 5.4 confirms the observation reported in the original work: in the ground-truth scenario, the minimum of the MSE curve coincides exactly with the optimal value of sigma for the J-invariant implementation of BM3D, as inferred from the self-supervised loss. We further evaluated the performance of s²ISM-BM3D across a range of sigma values by reconstructing the phantom data in Fig. 5.4a for 20 iterations at each simulated PSD. The results show that the retrieved parameter remains optimal when assessed using both the MSE and the Kullback–Leibler divergence (Fig. 5.4f). N2S enables reliable self-supervised calibration of BM3D denoising parameters within the s²ISM framework, eliminating the need for GT or manual parameter tuning and bringing the s²ISM reconstruction routine close to a fully data-driven approach, even in the experimental pipeline.

Conclusion of Chapter 6

This chapter addressed the ill-posed nature of the s²ISM inverse problem and the associated risk of noise amplification during iterative reconstruction. A comprehensive analysis of regularization strategies was presented, including explicit regularizers, Plug-and-Play methods, deep learning approaches, and inherent regularization mechanisms. The Gradient Consensus method was selected as the regularization strategy in our pipeline, as it exploits the intrinsic availability of statistically

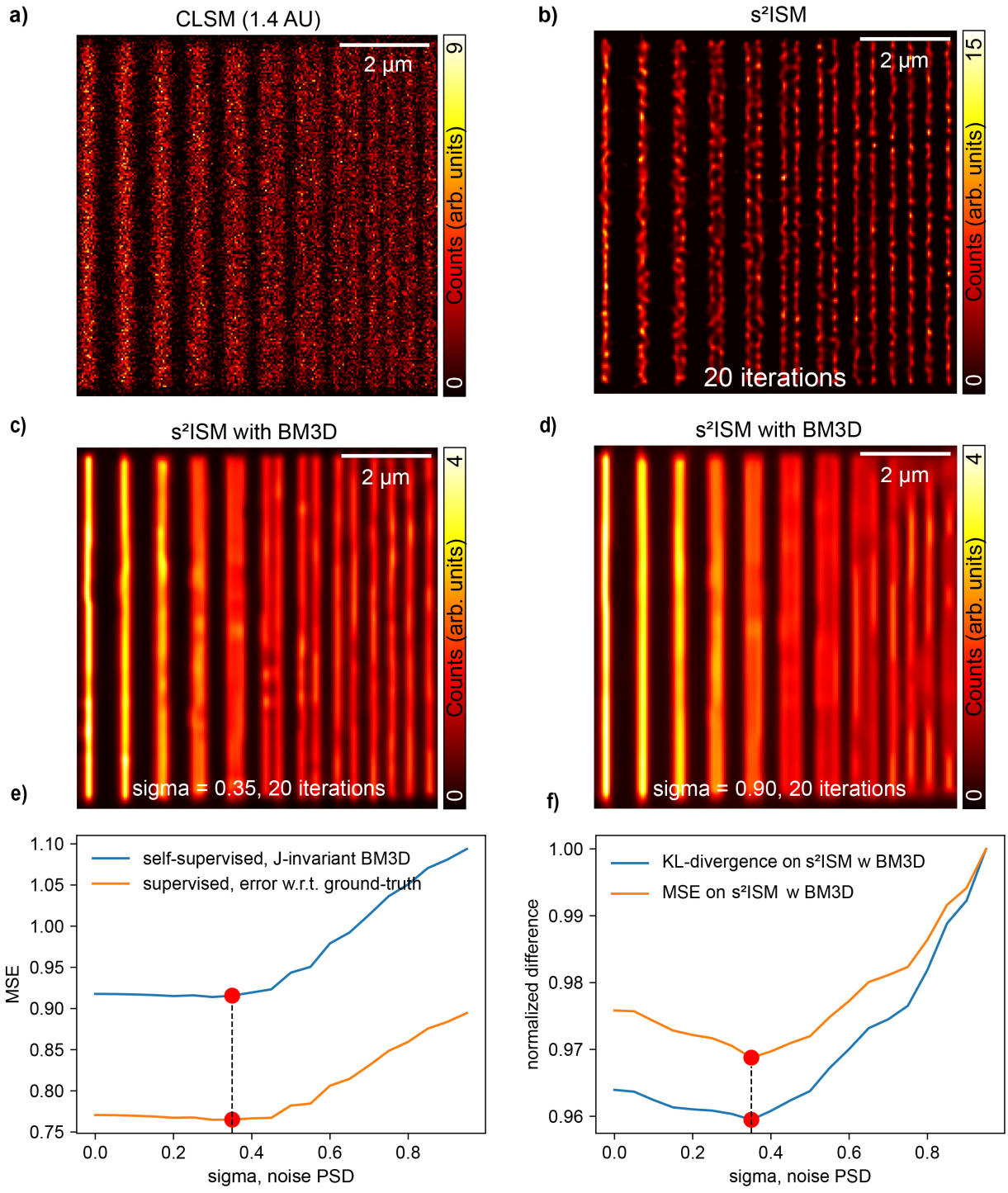

Figure 5.4: **N2S predicting optimal BM3D's denoising parameters in a fully self-supervised manner.** In **a**, we show the synthetic version of the Argolight sample, composed of a pair of lines with increasing separation from left to right. In **b**, we report the s^{2}ISM reconstruction of **a**. In **c**, we present the s^{2}ISM-BM3D reconstruction obtained with optimal BM3D denoising parameters, finely tuned through N2S. In **d**, we show the s^{2}ISM-BM3D reconstruction corresponding to the worst choice of the BM3D denoising parameter. In **e**, the blue curve represents the error as a function of the PSD value for self-supervised *j-invariant* BM3D, while the orange curve shows the MSE between the denoised and ground-truth images. The minima of the two curves coincide, indicating that the self-supervised *j-invariant* approach can blindly estimate the optimal PSD value from noisy data. Finally, in **f**, we compute, for each PSD value, the distance between the BM3D-s^{2}ISM reconstruction and the GT. Both the KL-divergence and the MSE identify the same optimal PSD value.

independent Poisson-distributed measurements in photon-resolved ISM. This approach effectively suppresses noise amplification while preserving spatial resolution and optical sectioning enhancement. A natural and promising next step is the extension of this regularization strategy to the algorithms within the s²ISM family dealing with the temporal dimension.

References

[1] S. Kabanikhin, N Tikhonov, V Ivanov, and M Lavrentiev. Definitions and examples of inverse and ill-posed problems. *Journal of Inverse and Ill-posed Problems - J INVERSE ILL-POSED PROBL*, 16:317–357, 01 2008.

[2] Paolo Massa and Federico Benvenuto. Predictive risk estimation for the expectation maximization algorithm with poisson data. *Inverse Problems*, 37, 03 2021.

[3] Ruturaj G. Gavaskar, Chirayu D. Athalye, and Kunal N. Chaudhury. On plug-and-play regularization using linear denoisers. *IEEE Transactions on Image Processing*, 30, 2021.

[4] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.

[5] A. N. Tikhonov and Vasiliy Yakovlevich Arsenin. Solutions of ill-posed problems. 1977.

[6] Jan Becker Craig Russell Andrew York, James Manton. Automatic data-driven parameter-free optimal deconvolution.

[7] Singanallur Venkatakrishnan, Charles Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. pages 945–948, 12 2013.

[8] Tolga Tasdizen. Principal neighborhood dictionaries for nonlocal means image denoising. *Image Processing, IEEE Transactions on*, 18:2649 – 2660, 01 2010.

[9] A. Levin and B. Nadler. Natural image denoising: Optimality and inherent bounds. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, page 2833–2840, USA, 2011. IEEE Computer Society.

[10] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

[11] Markku Mäkitalo and Alessandro Foi. Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 22, 06 2012.

[12] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision, 2019.

6 Conclusions

6.1 Discussion and perspectives

The ISM dataset contains rich and multidimensional information content that can be exploited both to enhance image quality and to extract additional physical and biological information about the specimen. This richness is further amplified when ISM is implemented using asynchronous readout SPAD array detectors, whose single-photon sensitivity and sub-nanosecond timing resolution enable a truly photon-resolved microscopy paradigm [1]. In this detection regime, fluorescence photons are recorded individually and are associated with detailed spatial and temporal signatures that are typically discarded in conventional microscopy.

Decoding the full informational content encoded in such datasets, however, constitutes a non-trivial inverse problem and requires dedicated reconstruction strategies. Existing ISM reconstruction algorithms have demonstrated effectiveness in achieving specific objectives, most notably lateral resolution enhancement, but they fail to exploit the full information theoretically available in the data. As a consequence, previous approaches have relied on the sequential application of multiple reconstruction or post-processing algorithms, an intrinsically sub-optimal strategy that is not always feasible and may lead to information loss.

In contrast, s^2ISM represents, to the best of our knowledge, the first reconstruction framework capable of comprehensively exploiting the information content of array-detected ISM data within a single, unified computational approach, without introducing intrinsic compromises or architectural drawbacks. Specifically, s^2ISM leverages the ISM dataset to reconstruct a single super-resolution image, Fig. 6.3c,d, that surpasses both the Abbe diffraction limit and the Nyquist sampling criterion, while simultaneously providing enhanced optical sectioning Fig. 6.3e,f.

Although s^2ISM is particularly well suited for sub-cellular super-resolution microscopy employing high numerical aperture objective lenses, its working principles are general and scalable. As

demonstrated experimentally in Fig. 6.1, the framework can be adapted to imaging across a wide range of spatial scales using arbitrary combinations of NA and magnification. More generally, s^{2}ISM can be applied to any laser-scanning microscope equipped with a detector array Fig. 6.2, independently of the specific optical configuration.

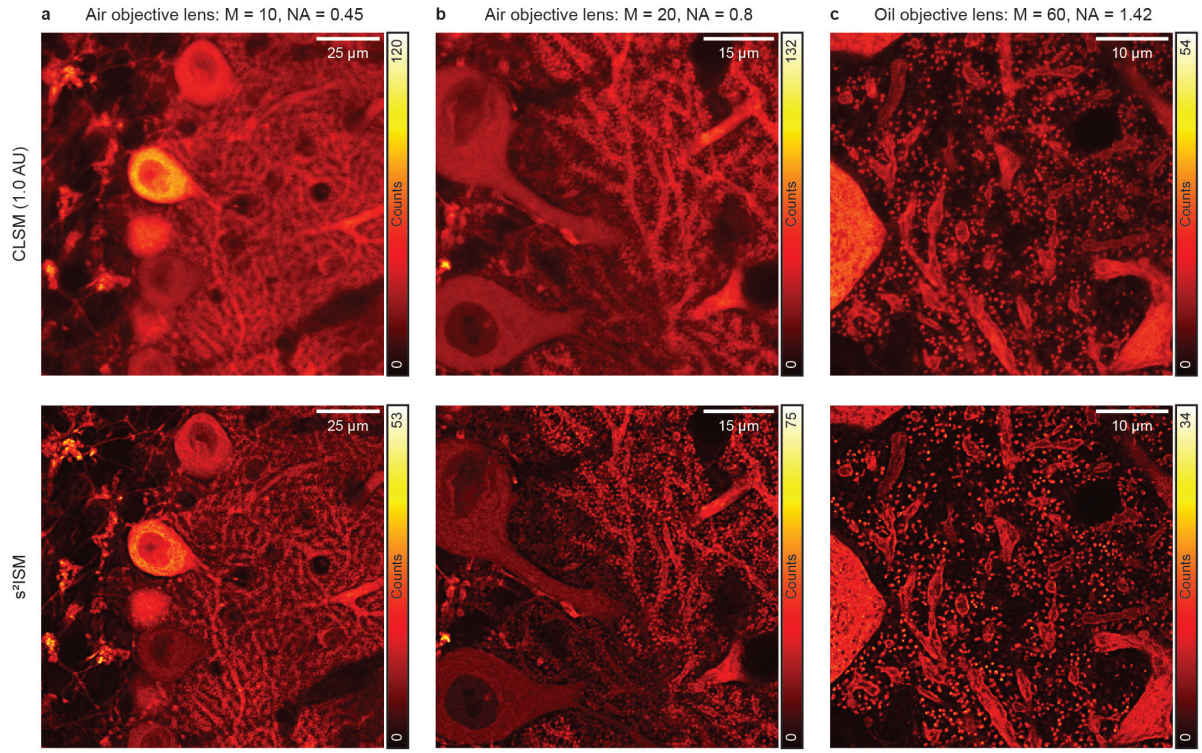

Figure 6.1: **Imaging at various magnifications and numerical apertures.** Slice of mouse cerebellum excited at $\lambda_{\text{exc}} = 488\,\text{nm}$ and imaged with a commercial NSPARC microscope. Top row: confocal image. Bottom row: corresponding s^{2}ISM reconstruction (5 iterations). **a**, data acquired with an air objective lens $10 \times /0.45$, total magnification $M = 172$. **b**, data acquired with an air objective lens $20 \times /0.8$, total magnification $M = 305$. **c**, data acquired with an oil-immersion objective lens $60 \times /1.42$, total magnification $M = 543$.

The feasibility and versatility of the proposed framework were demonstrated in the context of 2PE fluorescence microscopy for tissue imaging. Beyond this implementation, s^{2}ISM is inherently compatible with a broad class of nonlinear excitation schemes. For instance, fluorescence saturation effects exploited in saturated excitation ISM [2] can be combined with s^{2}ISM to further enhance spatial resolution. Similarly, single-molecule ISM [3] can benefit from the combined resolution enhancement and background suppression provided by s^{2}ISM, leading to improved localization precision. These improvements propagate synergistically to any additional dimension encoded in the dataset.

A particularly important extension arises from the temporal dimension naturally provided by SPAD array detectors. Their sub-nanosecond timing resolution enables the direct incorporation of time-resolved information into the reconstruction process. This capability was exploited to demonstrate that s^{2}ISM significantly improves the robustness and precision of fluorescence life-

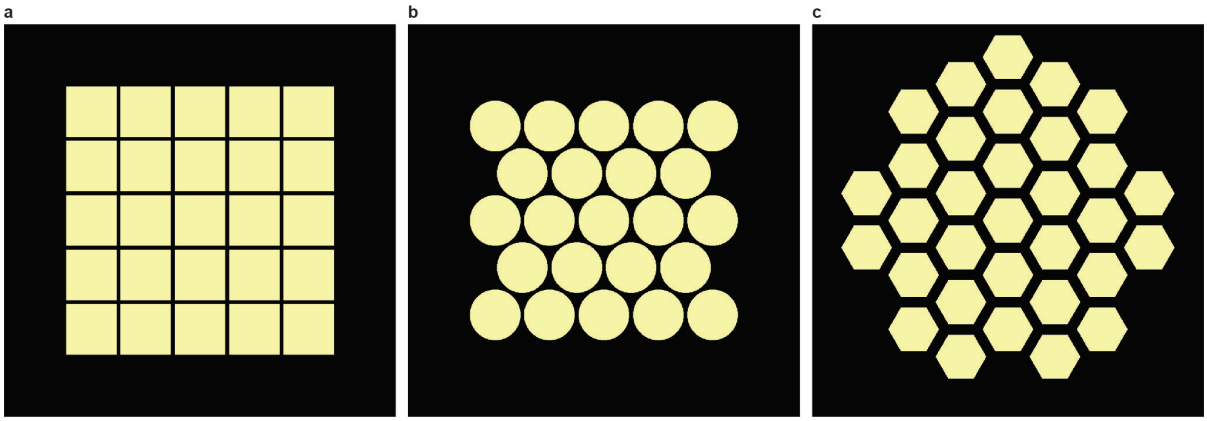

Figure 6.2: **Array detectors.** Examples of array detectors that can be simulated using our Python package. Any combination of grid and detector geometry is possible. **a**, rectangular array of squared detectors. **b**, hexagonal array of circular detectors. **c**, AiryScan array with hexagonal detectors. For the data presented in this work, we used the array detector as represented in **a**.

time imaging microscopy (FLIM) measurements. Comparable performance gains are expected for hyperspectral imaging applications [4]. Furthermore, s^{2}ISM was extended to lifetime-based structural unmixing by explicitly encoding fluorescence decay dynamics into the forward model, enabling the simultaneous reconstruction of optically sectioned images and the separation of structures characterized by distinct lifetimes.

Importantly, the current formulation of s^{2}ISM does not rely on a specific fluorescence dynamics model and is therefore applicable to a wide range of temporal phenomena. These include intensity fluctuations associated with transitions to dark states [5, 6], as well as photon antibunching effects from single-photon emitters [7]. Additional promising extensions involve stimulated emission depletion (STED) microscopy [8], particularly in time-resolved STED implementations [9, 10], where fluorescence lifetime information could be exploited to further enhance spatial resolution through separation-by-lifetime tuning concepts [11, 12].

By appropriately modeling multi-spot excitation, the s^{2}ISM framework could also be generalized to camera-based ISM architectures, enabling computational optical sectioning without the need for physical or virtual pinholes [13]. With suitable modifications, the conceptual framework may further be extended to coherent imaging modalities. One possible approach involves adapting the forward model to account for coherent image formation processes, although it remains unclear whether such implementations would provide benefits comparable to those observed in incoherent modalities such as fluorescence, spontaneous Raman and Brillouin scattering, or photothermal imaging. Alternatively, interferometric detection schemes could be employed [14, 15], in which case s^{2}ISM could be formulated in terms of optical fields rather than intensities.

From a numerical perspective, the reconstruction problem underlying s^{2}ISM is intrinsically ill-posed and, as with all maximum-likelihood estimation frameworks, is susceptible to noise amplification with increasing iteration number. Rather than relying on heuristic stopping criteria, this work adopts a regularization strategy based on the Gradient Consensus framework [16], which exploits the intrinsic statistical independence of photon-resolved measurements enabled by SPAD

detectors. The dense temporal sampling intrinsic to SPAD-based acquisition naturally enables the generation of multiple statistically independent realizations within a single experiment. The GC approach effectively suppresses noise overfitting during the reconstruction process while preserving spatial resolution and physical interpretability, and can be straightforwardly embedded within the s^{2}ISM framework.

Additional regularization strategies — including explicit regularizers [17], P-n-P methods, and deep-learning-based denoisers [18], such as Noise2Noise [19, 20], and related approaches [21] — have already begun to be investigated and represent promising complementary directions

Finally, while s^{2}ISM was developed under the assumption of negligible optical aberrations, the framework can be extended to aberrated systems by operating on local isoplanatic patches, provided that the corresponding wavefront distortions are measured and incorporated into the PSFs models. In this context, the integration of wavefront sensing techniques [22] with s^{2}ISM offers a promising route toward extending imaging depth and robustness in heterogeneous and scattering samples.

In conclusion, the s^{2}ISM framework enhances the capabilities of scanning microscopes equipped with detector arrays by enabling the extraction of the full informational content encoded in ISM data. Super-resolution, enhanced optical sectioning, SNR, and improved sampling efficiency are achieved simultaneously, without modifications to the microscope architecture and without compromising existing ISM functionalities. These advantages are further amplified when SPAD array detectors are employed, fully realizing the potential of the photon-resolved microscopy paradigm. We therefore expect s^{2}ISM to be widely adopted both as a practical imaging tool by the microscopy community and as a foundational platform for the development of future advanced reconstruction methodologies.

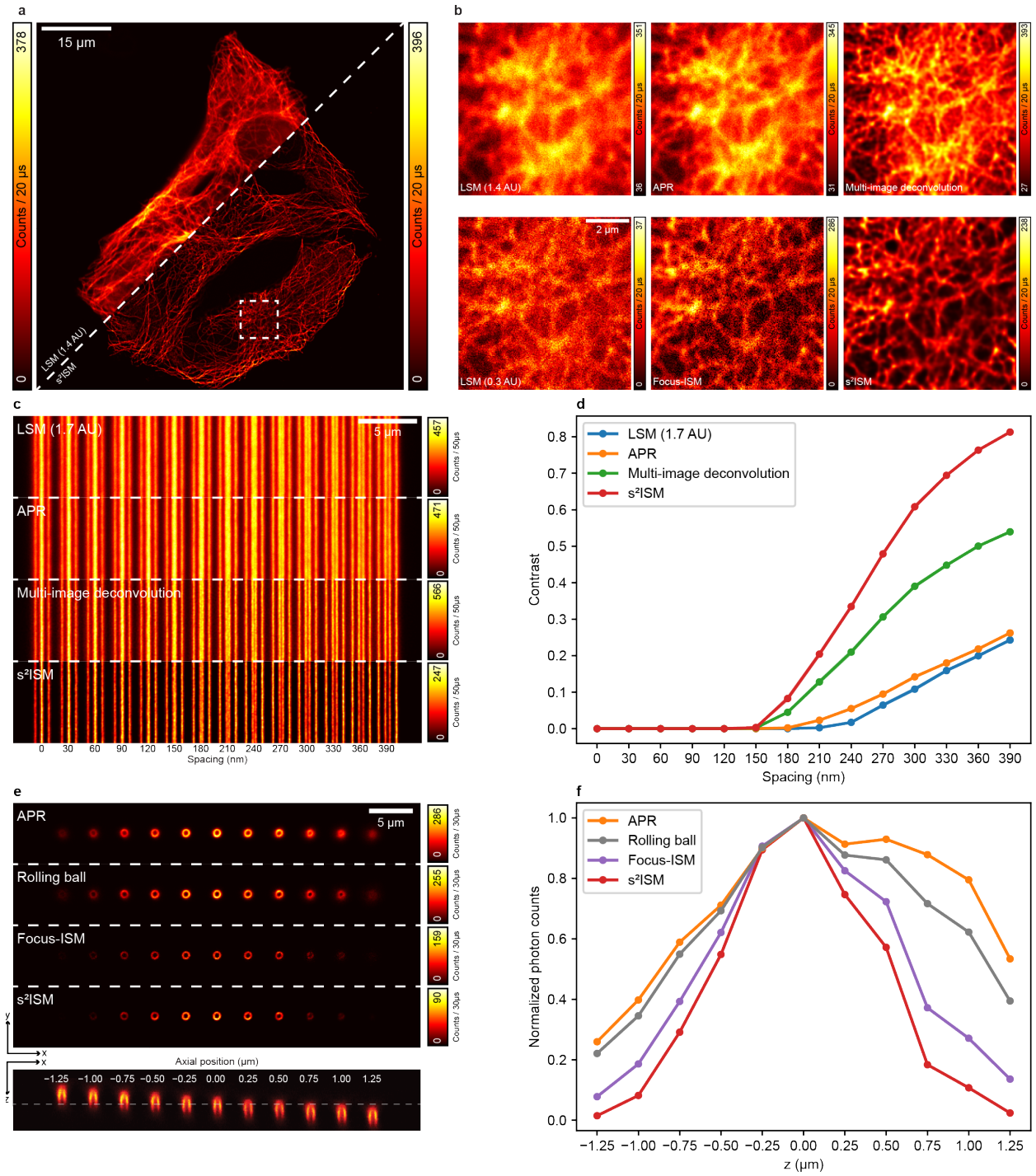

Figure 6.3: **Lateral resolution and optical sectioning.** **a**, Confocal image (left) of the tubulin network of a HeLa cell (mouse anti-α-tubulin combined with anti-mouse Abberior STAR RED) compared to the s^{2}ISM reconstruction (right). **b**, Detail of the image in **a** (white dashed box) reconstructed using different algorithms. From left to right, lateral resolution and SNR are improved. From top to bottom, optical sectioning is improved. Both multi-image deconvolution and s^{2}ISM algorithms are stopped at 20 iterations. We excited the specimen using the laser wavelength $\lambda = 640\,\mathrm{nm}$. **c**, Compared images of a resolution target composed of gradually spaced lines. **d**, Corresponding modulation transfer function, experimentally measured by calculating the contrast of the dip relative to the two adjacent lines. When no dip is discernible, the contrast is set to zero. **e**, Compared single-plane images of a three-dimensional stair of rings evenly spaced on the axial direction ($\Delta z = 250\,\mathrm{nm}$). On the bottom, we show the xz slice of a confocal 3D stack of the same target. The dashed line indicates the axial plane of the images above. **f**, Corresponding normalized optical sectioning function, calculated by summing the photon counts from each ring. We excited both targets using the laser wavelength $\lambda = 488\,\mathrm{nm}$.

References

[1] Alessandro Rossetta, Eli Slenders, Mattia Donato, Sabrina Zappone, Francesco Fersini, Martina Bruno, Francesco Diotalevi, Luca Lanzanò, Sami Koho, Giorgio Tortarolo, Andrea Barberis, Marco Crepaldi, Eleonora Perego, and Giuseppe Vicidomini. The brighteyes-ttm as an open-source time-tagging module for democratising single-photon microscopy. *Nature Communications*, 13:7406, 12 2022.

[2] Kenta Temma, Ryosuke Oketani, René Lachmann, Toshiki Kubo, Nicholas I. Smith, Rainer Heintzmann, and Katsumasa Fujita. Saturated-excitation image scanning microscopy. *Optics Express*, 30:13825, 4 2022.

[3] Niels Radmacher, Oleksii Nevskyi, José Ignacio Gallea, Jan Christoph Thiele, Ingo Gregor, Silvio O Rizzoli, and Jörg Enderlein. Doubling the resolution of single-molecule localization microscopy with image scanning microscopy. *bioRxiv*, page 2023.08.23.554438, 1 2023.

[4] Franziska Strasser, Martin Offterdinger, Rafael Piestun, and Alexander Jesacher. Spectral image scanning microscopy. *Biomedical Optics Express*, 10:2513, 5 2019.

[5] Aleksandra Sroda, Adrian Makowski, Ron Tenne, Uri Rossman, Gur Lubin, Dan Oron, and Radek Lapkiewicz. Sofism: Super-resolution optical fluctuation image scanning microscopy. *Optica*, 7:1308, 10 2020.

[6] Alexander Krupinski-Ptaszek, Adrian Makowski, Aleksandra Mielnicka, Monika Pawłowska, Ron Tenne, and Radek Lapkiewicz. Super-resolution microscopy based on the inherent fluctuations of dye molecules. *Biomedical Optics Express*, 16:910, 3 2025.

[7] Ron Tenne, Uri Rossman, Batel Rephael, Yonatan Israel, Alexander Krupinski-Ptaszek, Radek Lapkiewicz, Yaron Silberberg, and Dan Oron. Super-resolution enhancement by quantum image scanning microscopy. *Nature Photonics*, 13:116–122, 2 2019.

[8] Giuseppe Vicidomini, Paolo Bianchini, and Alberto Diaspro. Sted super-resolved microscopy. *Nature Methods*, 15:173–182, 3 2018.

[9] Giuseppe Vicidomini, Gael Moneron, Kyu Y Han, Volker Westphal, Haisen Ta, Matthias Reuss, Johann Engelhardt, Christian Eggeling, and Stefan W Hell. Sharper low-power sted nanoscopy by time gating. *Nature Methods*, 8:571–573, 7 2011.

[10] Giuseppe Vicidomini, Andreas Schönle, Haisen Ta, Kyu Young Han, Gael Moneron, Christian Eggeling, and Stefan W. Hell. Sted nanoscopy with time-gated detection: Theoretical and experimental aspects. *PLoS ONE*, 8:e54421, 1 2013.

[11] Luca Lanzanò, Iván Coto Hernández, Marco Castello, Enrico Gratton, Alberto Diaspro, and Giuseppe Vicidomini. Encoding and decoding spatio-temporal information for super-resolution microscopy. *Nature Communications*, 6:6701, 4 2015.

[12] Giorgio Tortarolo, Alessandro Zunino, Francesco Fersini, Marco Castello, Simonluca Piazza, Colin J.R. Sheppard, Paolo Bianchini, Alberto Diaspro, Sami Koho, and Giuseppe Vicidomini. Focus image scanning microscopy for sharp and gentle super-resolved microscopy. *Nature Communications*, 13, 12 2022.

[13] Andrew G York, Sapun H Parekh, Damian Dalle Nogare, Robert S Fischer, Kelsey Temprine, Marina Mione, Ajay B Chitnis, Christian A Combs, and Hari Shroff. Resolution doubling in live, multicellular organisms via multifocal structured illumination microscopy. *Nature Methods*, 9:749–754, 7 2012.

[14] Dekel Raanan, Man Suk Song, William A. Tisdale, and Dan Oron. Super-resolved second harmonic generation imaging by coherent image scanning microscopy. *Applied Physics Letters*, 120, 2 2022.

[15] Anna Zhitnitsky, Elad Benjamin, Ora Bitton, and Dan Oron. Super-resolved cars by coherent image scanning. *ArXiv*, 4 2024.

[16] James Manton. Richardson-Lucy Gradient Consensus. https://github.com/jdmanton/rlgc.

[17] Weisong Zhao, Shiqun Zhao, Liuju Li, Xiaoshuai Huang, Shijia Xing, Yulin Zhang, Guohua Qiu, Zhenqian Han, Yingxu Shang, De en Sun, Chunyan Shan, Runlong Wu, Lusheng Gu, Shuwen Zhang, Riwang Chen, Jian Xiao, Yanquan Mo, Jianyong Wang, Wei Ji, Xing Chen, Baoquan Ding, Yanmei Liu, Heng Mao, Bao-Liang Song, Jiubin Tan, Jian Liu, Haoyu Li, and Liangyi Chen. Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy. *Nature Biotechnology*, 40:606–617, 4 2022.

[18] Christian Daniele, Silvia Villa, Samuel Vaiter, and Luca Calatroni. Deep equilibrium models for poisson imaging inverse problems via mirror descent. 2025.

[19] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *ArXiv*, 3 2018.

[20] Liying Qu, Shiqun Zhao, Yuanyuan Huang, Xianxin Ye, Kunhao Wang, Yuzhen Liu, Xianming Liu, Heng Mao, Guangwei Hu, Wei Chen, Changliang Guo, Jiaye He, Jiubin Tan, Haoyu Li, Liangyi Chen, and Weisong Zhao. Self-inspired learning to denoise for live-cell super-resolution microscopy. *bioRxiv*, 2024.

[21] Chang Qiao, Yunmin Zeng, Quan Meng, Xingye Chen, Haoyu Chen, Tao Jiang, Rongfei Wei, Jiabao Guo, Wenfeng Fu, Huaide Lu, Di Li, Yuwang Wang, Hui Qiao, Jiamin Wu, Dong Li, and Qionghai Dai. Zero-shot learning enables instant denoising and super-resolution in optical fluorescence microscopy. *Nature Communications*, 15:4180, 5 2024.

[22] Daniele Ancora, Tommaso Furieri, Stefano Bonora, and Andrea Bassi. Spinning pupil aberration measurement for anisoplanatic deconvolution. *Optics Letters*, 46:2884, 6 2021.

A

Supplementary notes

A.1 Custom ISM architecture

For this work, we built a custom ISM setup (Fig. A.1). The excitation beams are provided by three triggerable pulsed ($80\,\mathrm{ps}$ pulse-width) diode lasers emitting at $640\,\mathrm{nm}$, $561\,\mathrm{nm}$, and $488\,\mathrm{nm}$ (LDH-D-C-640, LDH-D-C-560, and LDH-D-C-488 – Picoquant). We control the coarse power of the visible laser using their respective drivers and control software. We performed the fine control of the power using acoustic optical modulators (AOM, MT80-A1-VIS, AAopto-electronic). All laser beams are coupled into a different polarising-maintain fibre (PMF) to transport the beams to the microscope. In all cases, we used a half-wave plate (HWP) to adjust the beam polarization parallel to the fast axis of the PMF.

The beam for two-photon excitation is provided by a tunable ultrafast laser (Chameleon Vision – Coherent), emitting at $900\,\mathrm{nm}$ ($140\,\mathrm{fs}$ pulse-width). The power is controlled with HWP and a polarizing beam-splitter (PBS), which redirects a fraction of the light onto a beam dump, depending on the rotation angle of the HWP. The beam is magnified by a factor of 3 using a telescope. A set of dichroic mirrors (491 short-pass, 590 short-pass, 750 short-pass) allows the combination of all laser beams. The excitation and fluorescence light are separated by a different dichroic mirror (multi-reflection band 488-560-640-775 or 720 short-pass), depending on the excitation modality (one or two photons, respectively). Two galvanometer scanning mirrors (6215HM40B, CT Cambridge Technology), a scan lens and a tube lens – of a commercial confocal microscope (C2, Nikon) – deflect and direct all the beam towards the objective lens (CFI Plan Apo VC 60×, 1.4 NA, Oil, Nikon) to perform the raster scan on the specimen. The objective lens is mounted over a nanopositioner (FOC.100, Piezoconcept), enabling z-scanning. The fluorescence light is collected by the same objective lens, de-scanned, and sent towards the detection path. This latter consists of a set of lenses to form a telescopic system that

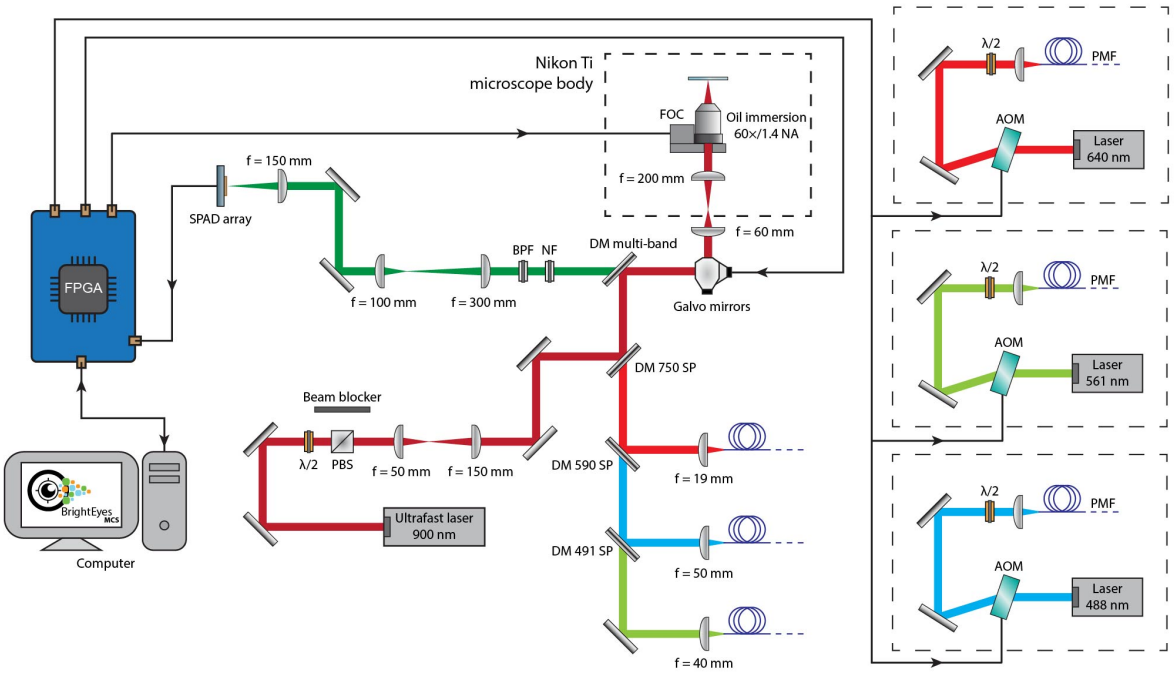

Figure A.1: **Scheme of the image scanning microscope.** AOM: acousto-optic modulator. PMF: single-mode polarization maintaining fibre. DM: dichroic mirror. BPF: band-pass filter. NF: notch filter. PBS: polarizing beam splitter. $\lambda/2$: half-wave plate.

conjugates the sample plane onto the detector plane with an overall magnification of 450×. Spectral filters are installed in the detection path to discard residual excitation light. Depending on the experiments, fluorescence light is selected by using a dedicated set of filters (red set: ZET633TopNotch and ET685/70M, green set: ZET561NF and ET575LP, blue set: ZET488NF and ET525/50M, two-photon set: 720SP and ET525/50M). The detector is a 7×7 SPAD array (PRISM-light kit, TTL version, Genoa Instruments) with a pixel pitch of $75\,\mu\mathrm{m}$, but only the inner 5×5 array is read due to limitations of the read-out system used in this work. Every photon detected by any element of the SPAD array generates a TTL signal that is delivered through a dedicated channel to a multifunction FPGA-based I/O device (NI USB-7856R from National Instruments), which acts both as a data-acquisition system and a control unit. The BrightEyes-MCS software [1] controls the entire microscope, including the galvanometric mirrors, the FOC, and the AOMs. The software also provides real-time image visualisation during the scan and saves the raw data in a hierarchical data format (HDF5) file. The saved file contains metadata as a dictionary and data as a six-dimensional array (repetition, axial position, vertical position, horizontal position, time, detector channel).

A.2 Time-resolved acquisition system

To implement TCSPC recording for the SPAD array detector, we used the multi-channel digital frequency domain (DFD) method [2]. This scheme allows for measuring the fluorescence decay at each scan point and for each detector array element. The DFD strategy enables the acquisition of periodic signals with a timing precision superior to direct sampling through a heterodyne measurement. Laser pulses are emitted at frequency f_{exc}, and the fluorescence is sampled at frequency f_s. Those frequencies are slightly detuned, $kf_{exc} = (k-1)f_s$. Therefore, the sampling accumulates a delay with every cycle, resulting in a sliding window that spans over $k-1$ excitation periods until the two signals are back in phase. Each period is more finely sampled in n shorter windows of duration T_w by a frequency $f_w = nf_s$. Defining the counters $w \in [0, n)$ and $\varphi \in [0, k)$, we reconstruct the time index γ as follows

$$\gamma = (mw - \varphi) \bmod k \tag{A.1}$$

where and $m = \frac{k-1}{n}$, and $k \in \mathbb{N}$, $n \in \mathbb{N}$ such that $m \in \mathbb{N}$. Finally, the photon arrival time is given by $t = \gamma T_{exc}/k$.

In our implementation on an FPGA board (NI USB-7856R, National Instruments), we used a base clock of $f_0 = 40\,\mathrm{MHz}$. From this latter, we derived the frequencies $f_{\mathrm{exc}} = \frac{28}{27} f_0 = 41.48\,\mathrm{MHz}$, $f_s = \frac{21}{20} f_0 = 42\,\mathrm{MHz}$, $f_w = \frac{21}{2} f_0 = 420\,\mathrm{MHz}$. The DFD parameters are $n = 10$, $k = 81$, and $m = 8$. Therefore, we obtain a timing precision of $\Delta t = T_{\mathrm{exc}}/k = 298\,\mathrm{ps}$ and a temporal resolution of $T_w = 2.38\,\mathrm{ns}$.

The DFD system builds the fluorescence decay histogram using Eq. A.1 for each detector coordinate $\boldsymbol{x}_d$. Additionally, the system returns an extra channel that samples the laser trigger signal. This latter is used as a reference to align different measurements to a common reference frame, whose origin is the instant of emission of the laser pulse.

A.3 Reconstruction with experimental PSFs

If the ISM PSFs are experimentally available, they can be used to replace the simulated PSFs in the s^{2}ISM workflow. In this case, the axial positions of the axial planes required to perform the reconstruction are unknown *a priori*.

Therefore, we need to acquire the full volumetric PSF and estimate the position of the planes in post-processing. We estimate the focal plane to be the one with the sharpest PSF. Equivalently, it is the plane where the normalized 2D modulation transfer function (MTF) has the largest integral. Since the size of the pinhole does not alter the position of the focal plane, we work with the open-pinhole PSF $h(\boldsymbol{x}_s, z) = \sum_{\boldsymbol{x}_d} h(\boldsymbol{x}_s, z|\boldsymbol{x}_d)$. The normalized MTF integral is

$$\gamma(z) = \int_{\mathbb{R}^2} \left| \mathcal{F}\left\{ \frac{h(\boldsymbol{x}_s, z)}{\int h(\boldsymbol{x}_s, z)\,\mathrm{d}\boldsymbol{x}_s} \right\} (\boldsymbol{k}_s) \right| \mathrm{d}\boldsymbol{k}_s \tag{A.2}$$

where the Fourier Transform $\mathcal{F}$ is calculated with respect to the lateral coordinates $\boldsymbol{x}_s$. The focal plane is estimated as

$$z_1 = \arg\max_z \gamma(z)$$

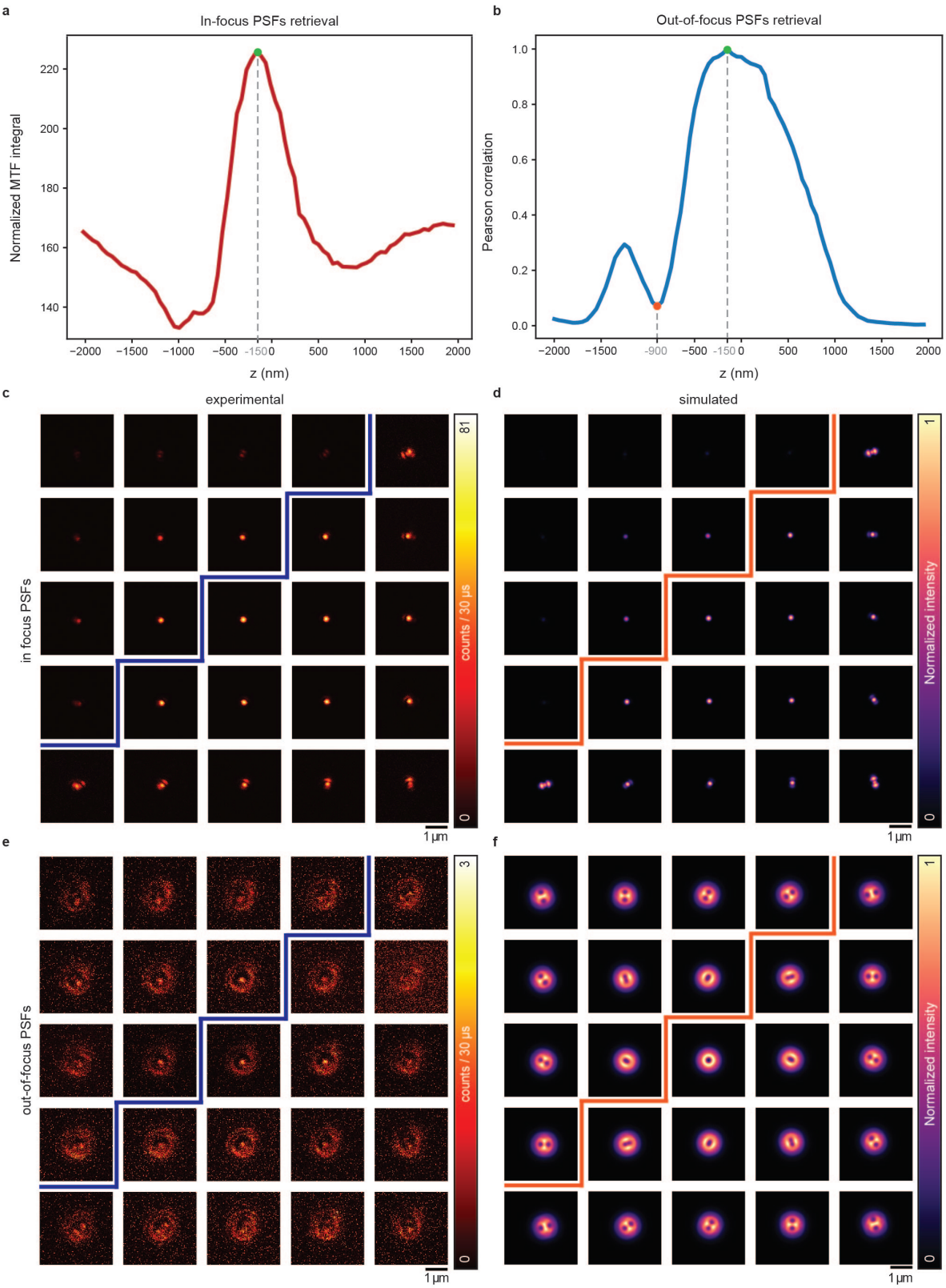

Figure A.2: **Comparison between simulated and experimental PSFs.** We calculate the axial planes required to run the s^{2}ISM algorithm from the 3D ISM dataset of the experimental PSF shown in Supp. Fig. **??**. **a**, we find the focal plane as the maximum position of the normalized MTF integral. **b**, we find the out-of-focus plane by minimizing the correlation of the in-focus PSFs with the defocused PSFs. We compare the measured **c** and simulated **d** in-focus PSFs, calculated using the procedure reported in Supp. Fig. 3.5. Similarly, we show the measured **c** and simulated **d** in-focus PSFs. Note that the defocus position is not the same, but is given by the procedure depicted in **b** and Supp. Fig. 3.3. For each dataset, the top-left corner images are normalized to the full dataset and the bottom-right images are normalized to themselves.

The out-of-focus position z_2 is calculated as in the simulation case

$$z_2 = \arg\max_z D[h(\boldsymbol{x}_s, z_1|\boldsymbol{x}_d) \parallel h(\boldsymbol{x}_s, z|\boldsymbol{x}_d)] \tag{A.3}$$

However, in the experimental scenario, the PSFs might not be axially symmetric due to the presence of some small aberrations. Therefore, we arbitrarily chose one side of the z-axis. Another valid approach would be to perform the reconstruction using two out-of-focus planes, one for each side. We show the result of a reconstruction using experimental PSFs in Supp. Fig. A.2.

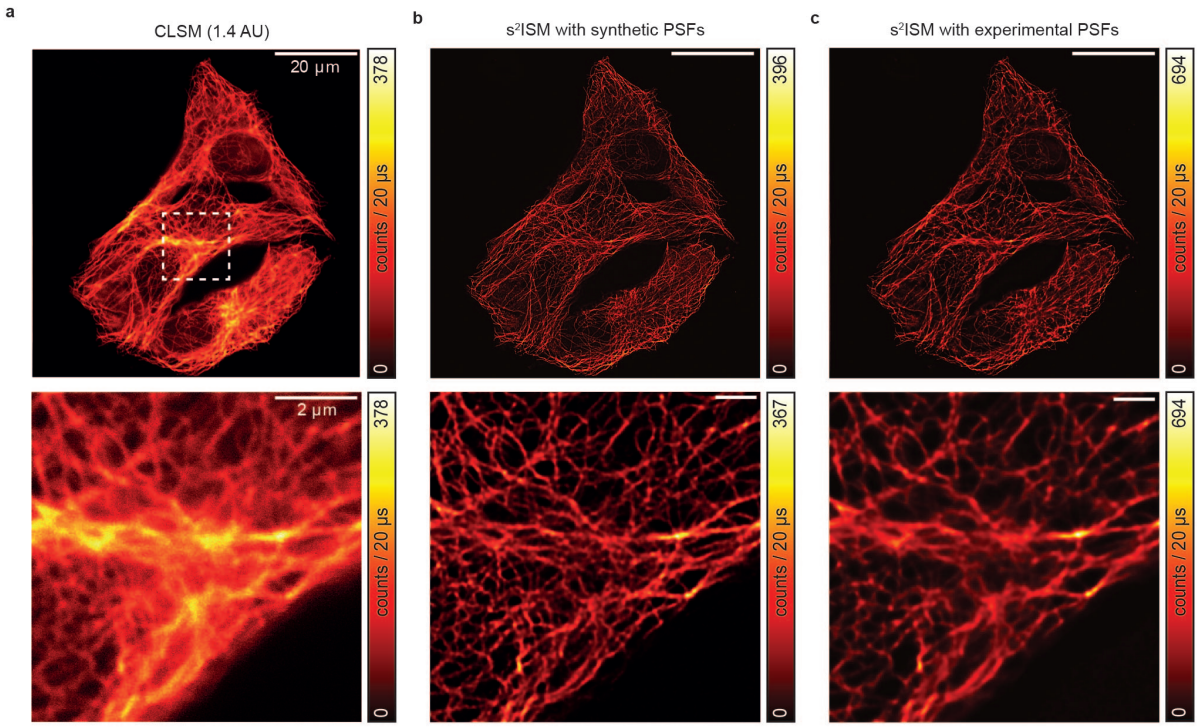

Figure A.3: **s²ISM with experimental and synthetic PSFs.** From the ISM dataset of the tubulin network of a group of HeLa cells, we performed the following reconstructions. **a**, open-pinhole confocal image. **b**, s²ISM image obtained using synthetic PSFs. **c**, s²ISM image obtained using esperimental PSFs. The used PSFs are shown in Supp. Fig. A.2. Both reconstructions are stopped at 10 iterations.

A.4 SNR estimation

Thanks to the high speed and sensitivity of the SPAD array detector, we can finely tune the duration of photon counting windows on each scan point. We exploited the detector's speed of to subdivide each pixel dwell time T into n_T temporal bins. To estimate the SNR of the final images, we summed the temporal bins into l realizations of the same ISM dataset. Then, we independently reconstructed each realization with the s²ISM algorithm and saved the result at each iteration. Finally, we calculated pixel-wise mean and standard deviation among the images,

obtaining an SNR map for each algorithm update. We show the result of this analysis in Supp. Fig. A.4, where we also report the histogram of the SNR values. To exclude pixels with no signal from the histogram, we bin the values in the range $[5, 60]$ and we calculate the median SNR for each iteration.

The results indicate that the best increase in SNR occurs at a small number of iterations. However, note that this analysis cannot be used to choose the best s^{2}ISM iteration since each realization has a lower starting SNR (smaller pixel dwell time) that the dataset integrated over the time bins.

A.5 Analysis of the reference target

A.5.1 Gradually spaced lines

First, we calculated the confocal image by summing the raw images of the ISM dataset. Then, we smoothed the image using a Gaussian kernel with a standard deviation of two pixels. We binarized the image using Otsu's method and filled any residual hole. Later, we calculated the countour of the binary mask, obtaining a contour for each region containing a single pair. We fit the coordinates of these latter to a line, obtaining the rotation angle for the target with respect to the reference frame of the image. We reconstructed the ISM dataset using different algorithms and counter-rotated the resulting images by the angle found previously.

We calculated 40 line profiles obtained from the central region of the rotated images, averaged over 10 rows. We chose to sum multiple lines over a limited range to increase the SNR of each line profile without induce blurring due to imprecise correction of the rotation. We identified the relative maxima and minima corresponding to each line pair and calculated the contrast for each spacing as

$$\text{Contrast} = \frac{\max - \min}{\max} \tag{A.4}$$

where $\max$ is the average peak value of the line pair, and $\min$ is the value of the dip between the two peaks. In case only a single peak can be identified, the contrast is set to zero. We repeated the analysis for each line profile, obtaining an average contrast and the relative standard error, and for the reconstructed image. The results are shown in Fig. 3.17.

A.5.2 3D crossing stairs

We acquired a 3D stack of images of the crossing stair, and summed the raw images of the ISM dataset to obtain a confocal stack. Then, we projected the stack over the axial dimension to obtain a single 2D image which we smoothed with a Gaussian kernel with a standard deviation of five pixels. Later, we binarized the image using Otsu's method, calculated the morphological closing, and cleaned the borders. We labeled each disconnected region, obtaining a segmentation of the 2D image.

Then, we extracted the central dataset of the raw stack and we reconstructed it using the s^{2}ISM algorithm. Finally, we calculated the total photon counts inside each segmented region

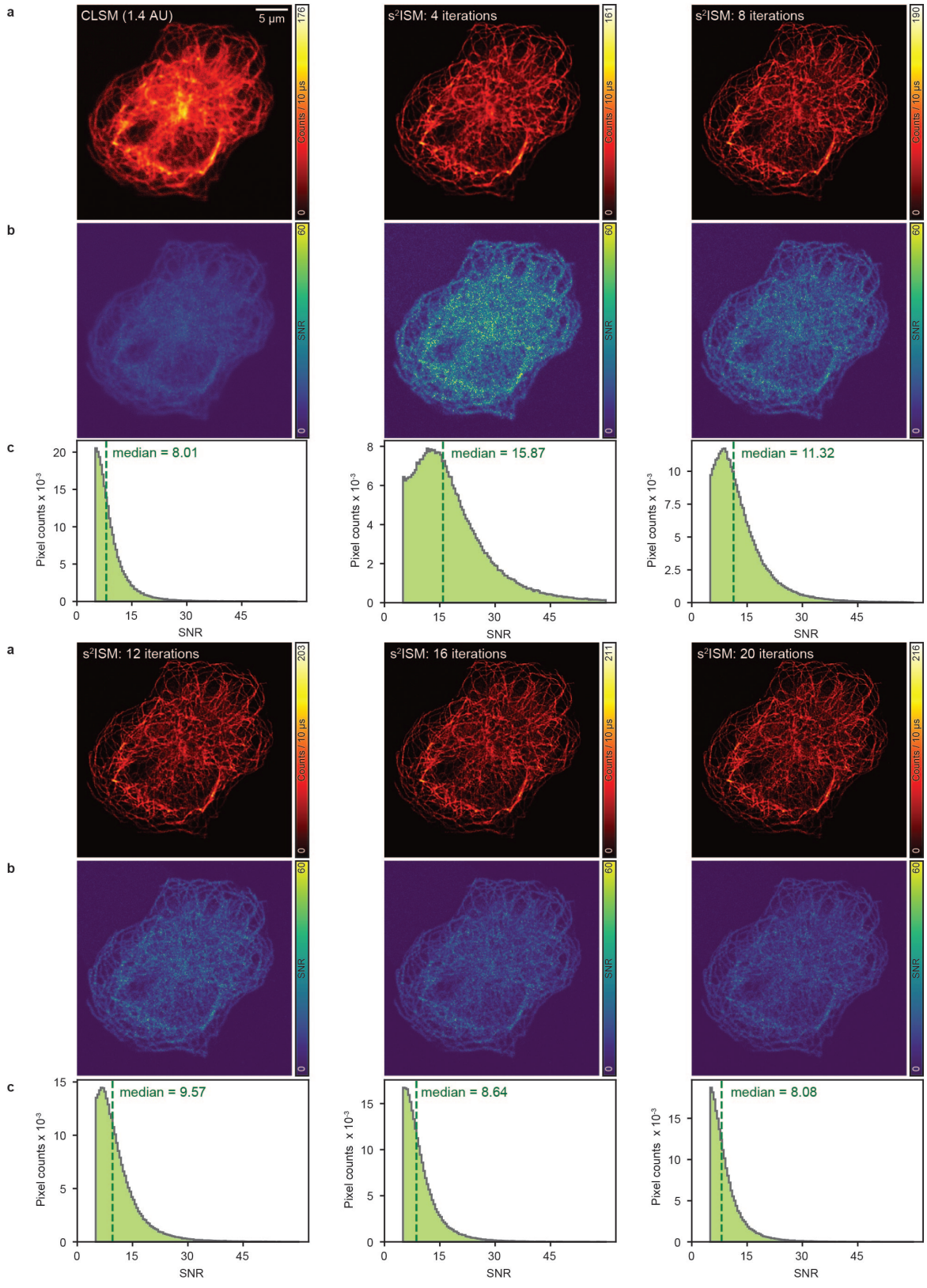
a
CLSM (1.4 AU)
5 µm
s²ISM: 4 iterations
s²ISM: 8 iterations
s²ISM: 12 iterations
s²ISM: 16 iterations
s²ISM: 20 iterations
Counts / 10 µs
SNR
b
c
median = 8.01
median = 15.87
median = 11.32
median = 9.57
median = 8.64
median = 8.08
Pixel counts x 10⁻³
SNR

Figure A.4:

before and after the reconstruction. From the manufacturer's documentation, we can associate each lateral position to an axial position, allowing for the calculation of an axial contrast curve. The results are shown in Supp. Fig. C.4.

A.6 Phasor analysis

The signal acquired by TCSPC system at each spatial coordinate can be modeled as the convolution of the fluorescence decay $d(t|\boldsymbol{x}_s, \boldsymbol{x}_d)$ and the IRF $h(t)$, expressed as

$$f(t|\boldsymbol{x}_s, \boldsymbol{x}_d) = d(t|\boldsymbol{x}_s, \boldsymbol{x}_d) * h(t). \tag{A.5}$$

For a single-exponential decay with lifetime τ,

$$d(t) = \begin{cases} 0 & t < 0 \\ d_0 \exp(-t/\tau) & t \geq 0, \end{cases} \tag{A.6}$$

where the temporal reference origin corresponds to the laser excitation pulse.

In Fourier space, the normalized signal's transform can be written as

$$\frac{\mathcal{F}\{d(t)\}(\omega)}{\int_0^{+\infty} d(t)\, dt} = \frac{1}{1 + i\omega\tau} = \frac{1}{1 + (\omega\tau)^2} - i\frac{\omega\tau}{1 + (\omega\tau)^2} = g(\omega) - is(\omega) \tag{A.7}$$

where we defined $g(\omega)$ and $s(\omega)$ respectively as the real and imaginary part of the Fourier transform of the decay. Notably, these quantities are related by the following equation

$$[g(\omega) - 1]^2 + s^2(\omega) = \frac{1}{4} \tag{A.8}$$

defining a semicircle on the complex plane known as the universal circle. The vector (g, s), called the phasor, lies on this circle for single-exponential decays, while multi-exponential decays correspond to linear combinations of phasors lying within the circle (Fig. 4.1).

From these components, the fluorescence lifetime can be estimated via two equivalent expressions:

$$\tau_\phi = \frac{1}{\omega} \tan\left[\phi(\omega)\right], \quad \text{where } \phi(\omega) = \arctan\frac{s(\omega)}{g(\omega)}, \tag{A.9}$$

and

$$\tau_m = \frac{1}{\omega}\sqrt{\frac{1}{m^2(\omega)} - 1}, \quad \text{where } m^2(\omega) = g^2(\omega) + s^2(\omega). \tag{A.10}$$

For single-exponential decays, these two estimates coincide and are frequency-independent. Because sampled data are discrete, the analysis is generalized by expressing the phasor coordinates

as the real and imaginary parts of the discrete Fourier transform (DFT) of the measured decay:

$$g(h) = \frac{1}{I}\sum_{p=0}^{n_p-1} d(p)\cos\left(\frac{2\pi hp}{n_p}\right), \quad s(h) = \frac{1}{I}\sum_{p=0}^{n_p-1} d(p)\sin\left(\frac{2\pi hp}{n_p}\right), \tag{A.11}$$

where $I = \sum_p F(p)$ and n_p is the number of data points. Identifying $\omega = 2\pi h f_{\text{exc}}$, numerical lifetime estimates follow directly.

Using the convolution theorem, the phasor of the measured signal F can be decomposed as the product of the phasors of the fluorescence decay and the IRF:

$$F = m_F e^{i\phi_F} = (m_D \cdot m_H)\exp[i(\phi_D + \phi_H)], \tag{A.12}$$

where m and ϕ denote magnitude and phase. The IRF phasor $m_H e^{i\phi_H}$ can be estimated either indirectly from a known sample or directly from an almost instantaneous response, such as back-scattered light from a gold bead. We adopted the latter, measuring the signal with spectral filters retained and applying a rectangular time window to remove multiple reflections, thereby ensuring a high signal-to-noise ratio. The laser trigger channel provided a common temporal reference, allowing us to calibrate the phasor phase by subtracting the reference phase φ.

This calibration corresponds to a Wiener deconvolution in frequency space, expressed as

$$\phi_D(\mathbf{x}_s, \mathbf{x}_d) = \phi_F(\mathbf{x}_s, \mathbf{x}_d) - \phi_H(\mathbf{x}_d) + \varphi_H - \varphi_F, \tag{A.13}$$

$$m_D(\mathbf{x}_s, \mathbf{x}_d) = \frac{m_F(\mathbf{x}_s, \mathbf{x}_d)}{m_H(\mathbf{x}_d)}. \tag{A.14}$$

For CLSM and ISM, phasors from different detector positions are temporally aligned (Fig. 4.2a) and summed before analysis, effectively treating the data as single-channel decays (Fig. 4.2d). The fluorescence lifetime maps are then computed as

$$\tau_\phi = \frac{1}{2\pi f_{\text{exc}}}\tan(\phi_D), \quad \tau_m = \frac{1}{2\pi f_{\text{exc}}}\sqrt{\frac{1}{m_D^2} - 1}. \tag{A.15}$$

Between the two, τ_m is generally more robust and less sensitive to calibration errors, and was thus preferred in this work.

Phasor analysis offers a simple and intuitive graphical framework, with the notable feature that the phasor of a mixture lies on the line connecting the phasors of its components, which lies on the universal semi-circle. Its model-independent nature avoids assumptions on decay form, enabling broad applicability and high-throughput automation. However, phasor accuracy degrades significantly in low signal-to-noise regimes and is sensitive to distortions from the instrument response.

Importantly, when analyzing decay data reconstructed via the s^{2}ISM algorithm, the IRF effect is inherently compensated through temporal deconvolution, eliminating the need for IRF correction beyond phase alignment with the reference channel.

References

[1] Mattia Donato, Eli Slenders, Alessandro Zunino, Luca Bega, and Giuseppe Vicidomini. Brighteyes-mcs: a control software for multichannel scanning microscopy. *Journal of Open Source Software*, 9(103):7125, 2024.

[2] Giorgio Tortarolo, Alessandro Zunino, Simonluca Piazza, Mattia Donato, Sabrina Zappone, Agnieszka Pierzyńska-Mach, Marco Castello, and Giuseppe Vicidomini. Compact and effective photon-resolved image scanning microscope. *Advanced Photonics*, 6, 1 2024.

B Measures

B.1 Kullback-Leibler divergence

The generalized KL divergence (or Csiszár I-divergence) is defined as follows

$$D_{\mathrm{KL}}(o \,\|\, \hat{o}) = \sum_{x_s \in \chi} \left(o(x_s) \log \frac{o(x_s)}{\hat{o}(x_s)} - o(x_s) + \hat{o}(x_s) \right) \tag{B.1}$$

where o is the reference image, $\hat{o}$ is an estimate of o, and the sum is performed only on the subset of pixels $\chi = \{x_s : o(x_s) \neq 0\}$. In our analysis, we choose to remove the intensity offset. We noticed that in the deconvolution framework, the quantity $-o(x_s) + \hat{o}(x_s)$ can lead to weird values of the divergence, because of the reassigment of the deconvolution and the locality of the divergence itself. If not specified, we will refer to the KL-divergence while instead using the entropy measure:

$$D_{\mathrm{KL}}(o \,\|\, \hat{o}) = E(o \,\|\, \hat{o}) = \sum_{x_s \in \chi} \left(o(x_s) \log \frac{o(x_s)}{\hat{o}(x_s)} \right) \tag{B.2}$$

B.2 Upsampling error

We quantified the upsampling error ϵ as follows

$$\epsilon(i, \hat{i}) = \sum_{x_s} \left| \frac{i(x_s)}{\sum_{x_s} i(x_s)} - \frac{\hat{i}(x_s)}{\sum_{x_s} \hat{i}(x_s)} \right| \tag{B.3}$$

where i is the conventionally sampled image and $\hat{i}$ is the corresponding upsampled reconstruction. Note that each pixel is normalized to the total counts of the corresponding image to compensate for the fact that the upsampled image contains roughly one-fourth of the counts of the reference image.

B.3 Structural similarity index measure (SSIM)

In 3.3.3, we generalized s^{2}ISM to enable upsampling. To measure the quality of the reconstruction, we acquired a dataset using the condition $2\Delta\boldsymbol{x}_s = \Delta\boldsymbol{x}_d$ and later downsampled the acquisition by a factor of two. Then, we reconstructed the original and downsampled ISM dataset, with and without upsampling. The results are shown in Fig. 3.19b and extended in Supp. Fig. C.8. We analyzed the reliability of the reconstructions by calculating the structural similarity index measure (SSIM) between the two reconstructions. Naming the images x and y, the SSIM index is defined as

$$\mathrm{SSIM} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{B.4}$$

where μ_x, μ_y are the mean values, σ_x, σ_y the respective standard deviations, and σ_{xy} the covariance. The above quantities are calculated in a subregion of the images defined by a Gaussian window with a standard deviation of 10 pixels. We set the regularizers constants to zero, $c_1 = c_2 = 0$. The SSIM metric returns a value in the $[-1, 1]$ range, where 1 indicates optimal similarity. By sliding the window on the full image, we calculated an SSIM map. The result is shown in Supp. Fig. 3.16.

B.4 Radial spectrum

We used the radial spectrum to compare the reconstructions from different algorithms. The spatial spectrum of an image $i(\boldsymbol{x}_s)$ as its Fourier transform

$$\mathcal{F}\{i(\boldsymbol{x}_s)\} = I(\boldsymbol{k}_s) = O(\boldsymbol{k}_s) \cdot H(\boldsymbol{k}_s) \tag{B.5}$$

where the capital letters indicate the Fourier transform of the corresponding quantities and $\boldsymbol{k}_s = (k_s \cos\phi, k_s \sin\phi)$ is the spatial frequency vector written in polar coordinates. The radial spectrum is given by the absolute value of the average on the angular dimension

$$S(k) = \frac{1}{2\pi} \int_0^{2\pi} |I(k, \varphi)| \, \mathrm{d}\varphi \tag{B.6}$$

We report the results of the calculation in Supp. Fig. B.1.

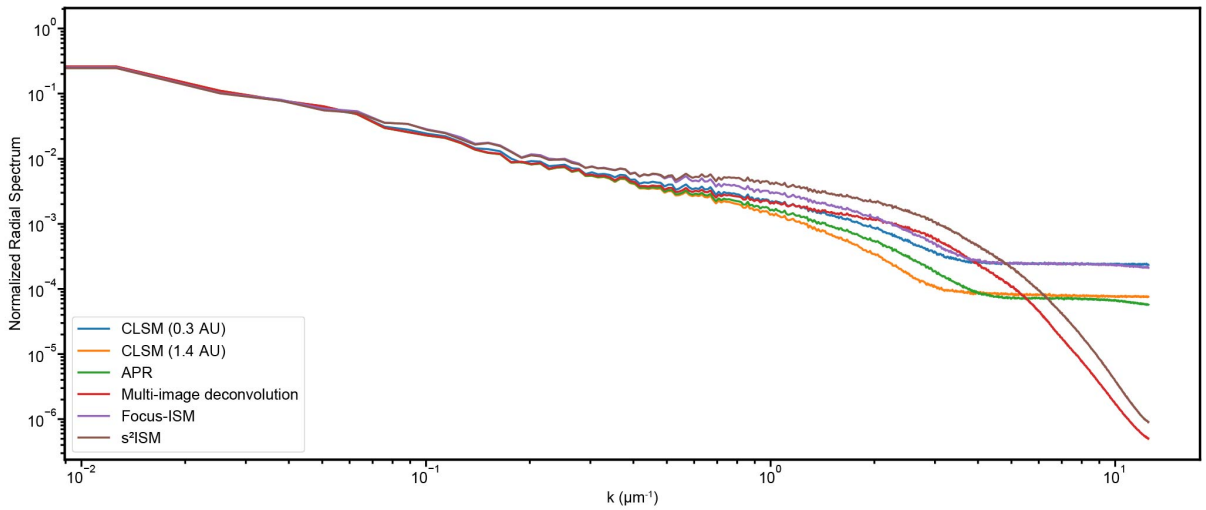

Figure B.1: **Radial spectrum analysis.** Normalized radial spectra calculated on the images of Supp. Fig. C.3. The high-frequency plateau is the noise-level. A larger cut-off correlates with higher lateral resolution, while larger high-frequency values (before the cut-off) correlate with higher optical sectioning. Lower values after the cut-off correlate with a higher signal-to-noise ratio.

B.5 Axial position

We first position the out-of-focus plane as far as possible from the in-focus plane to slightly shift the well-position of the inversion. However, caution is required: PSFs located very far from the in-focus plane, i.e., beyond the DOF of the microscope, will be essentially null, resulting in no out-of-focus rejection. This observation led to our initial idea: to place the background plane

just before the edge of the setup's DOF. Although this approach is largely empirical, it produced impressive results as a first step. As previously mentioned, s^{2}ISM can be run using two axial planes in the idea scenario. The first plane is set at the focal plane ($z_1 = 0$), while the second, out-of-focus plane can be chosen according to different criteria.

Maximum discrepancy

This criterion defines the out-of-focus plane as the one that maximizes the conditioning of the inverse problem. To ensure this latter condition, we maximize the discrepancy between the computationally-generated in-focus and out-of-focus PSFs

$$z_2 = \arg\max_z D[h(\boldsymbol{x}_s, 0|\boldsymbol{x}_d) \parallel h(\boldsymbol{x}_s, z|\boldsymbol{x}_d)] \tag{B.7}$$

where the PSFs are normalized by the flux at the corresponding axial planes

$$\int_{\mathbb{R}^4} h(\boldsymbol{x}_s, z|\boldsymbol{x}_d)\,\mathrm{d}\boldsymbol{x}_d\,\mathrm{d}\boldsymbol{x}_s = 1 \qquad \forall z \in \mathbb{R} \tag{B.8}$$

As a discrepancy functional, we used either the Kullback-Leibler divergence or the negative Pearson correlation. Since the non-aberrated PSFs are axially symmetric, we run the maximization algorithm only for positive z values up to a distance equal to the depth of field of the microscope, calculated using the excitation wavelength. We explore the axial range in 40 steps to be densely partitioned. As shown in Fig. 3.3, both metrics yield the same result.

Cumulative method

Furthermore, we tried to approach this question with a slightly different technique (Fig. B.2). We generate a thick phantom, densely discretized along the z-axis. Then we compute the forward and we give the latter as input for s^{2}ISM, where we decide the DOF to be the depth of the out-of-focus reconstruction. In the final step, we compute the Pearson correlation between the in-focus reconstruction and the cumulative ground-truth at each depth. As we can see from the plot Fig. B.2d, there is a position maximizing the correlation. The good result is that the conclusion of those two analyses concords.

Maximum optical sectioning

To evaluate our approach, we simulated the acquisition of an under-diffraction-limit gold bead by generating the ideal PSFs of our system. We then applied s^{2}ISM, varying the depth of the out-of-focus plane from 0 nm to 980 nm in 40 nm steps.

For each configuration, the algorithm was run for 10 iterations, and we quantified both the optical sectioning performance and the fraction of the total photon budget allocated to the in-focus reconstruction (Fig. 3.7). Thanks to the photon flux conservation property, these metrics were readily accessible. The optimal axial position was selected by optimizing along the inherent Pareto front, which aims to maximize the contrast between in-focus and out-of-focus emitters.

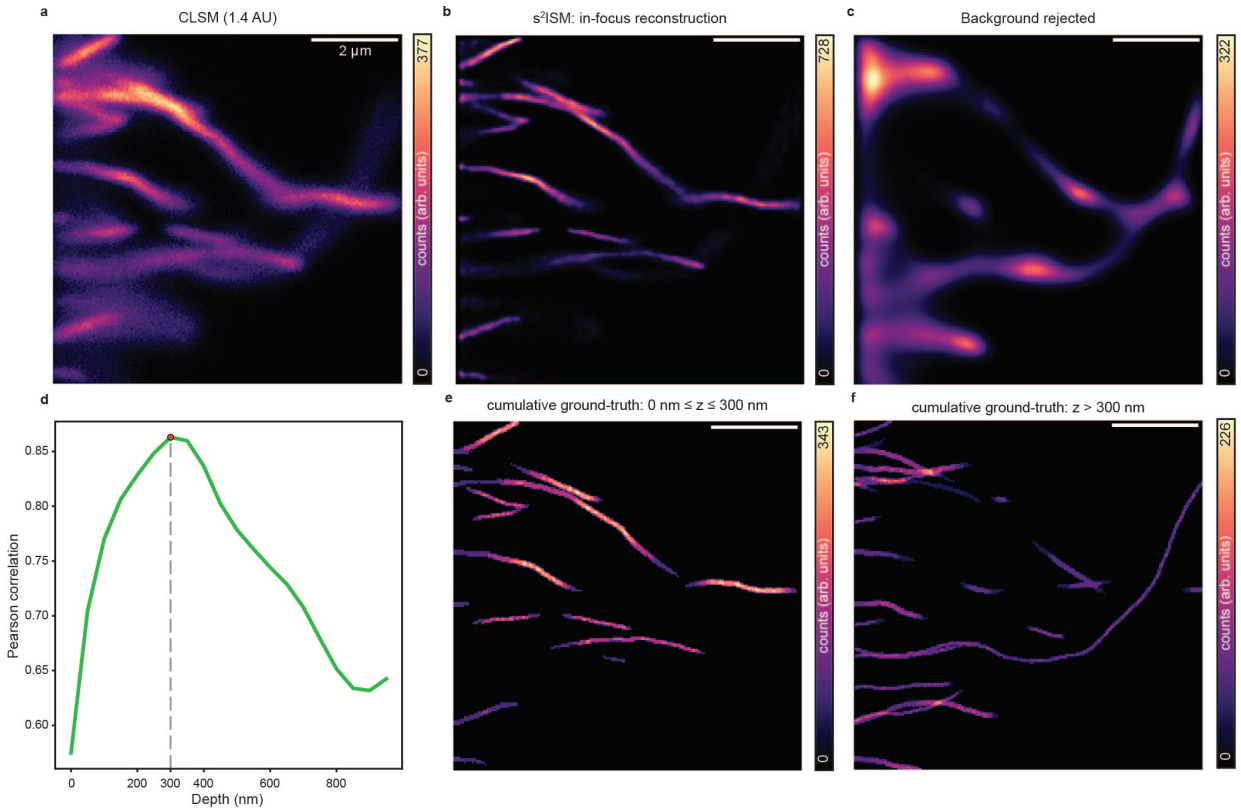

Figure B.2: **s²ISM on a synthetic thick sample.** We generated a volumetric phantom of $20 \times 201 \times 201$ voxels (zxy). We simulate the corresponding PSF dataset of $20 \times 201 \times 201 \times 25$ voxels $(zxyc)$ with a lateral and axial size of $40\,\mathrm{nm}$ and $50\,\mathrm{nm}$, respectively. The raw ISM dataset with size $201 \times 201 \times 25$ voxels (xyc) is obtained by convolving plane-by-plane and channel-by-channel the ground truth with the PSF and then summing over the axial dimension. **a**, confocal image obtained summing over the channel dimension. Images of the in-focus **b** and out-of-focus **c** reconstructed by s²ISM. **d** correlation of the in-focus reconstruction with the partial axial integral of the ground-truth up to z. The position of the correlation maximum indicates the effective size of the optical section. Ground-truth axially integrated up to **e** and above **f** the correlation maximum at $z = 300\,\mathrm{nm}$.

To do this, we performed s²ISM reconstruction on a 3D dataset of a single point emitter at various defocused positions. For each plane, the total photon counts were computed by summing over the xy axes and normalizing the resulting curves to their respective maxima, producing the contrast curves. Narrower contrast curves correspond to smaller integrals, which we then integrated to estimate the algorithm's optical sectioning power . The optimal out-of-focus plane was defined as the position that minimizes the integral of the contrast curve.

Because this procedure is computationally intensive, we propose an alternative estimate for the optimal axial position: the inverse of the ideal confocal cut-off axial frequency, which is defined as twice the wide-field axial cut-off.

Thus, the optimal out-of-focus position can be approximated to

$$z_2 = \frac{1}{2k_z^{\mathrm{max}}} = \frac{\lambda}{4n\sin^2\alpha/2} \tag{B.9}$$

We applied the same analysis in an experimental setting (Fig 3.8) and observed a strong correspondence between the in silico and experimental results. Then, to validate the approach, we applied the same idea, namely, ranging the out-of-focus position along the z-axis and comparing the optical sectioning gain on an Argolight structured sample in the xz dimension B.3.

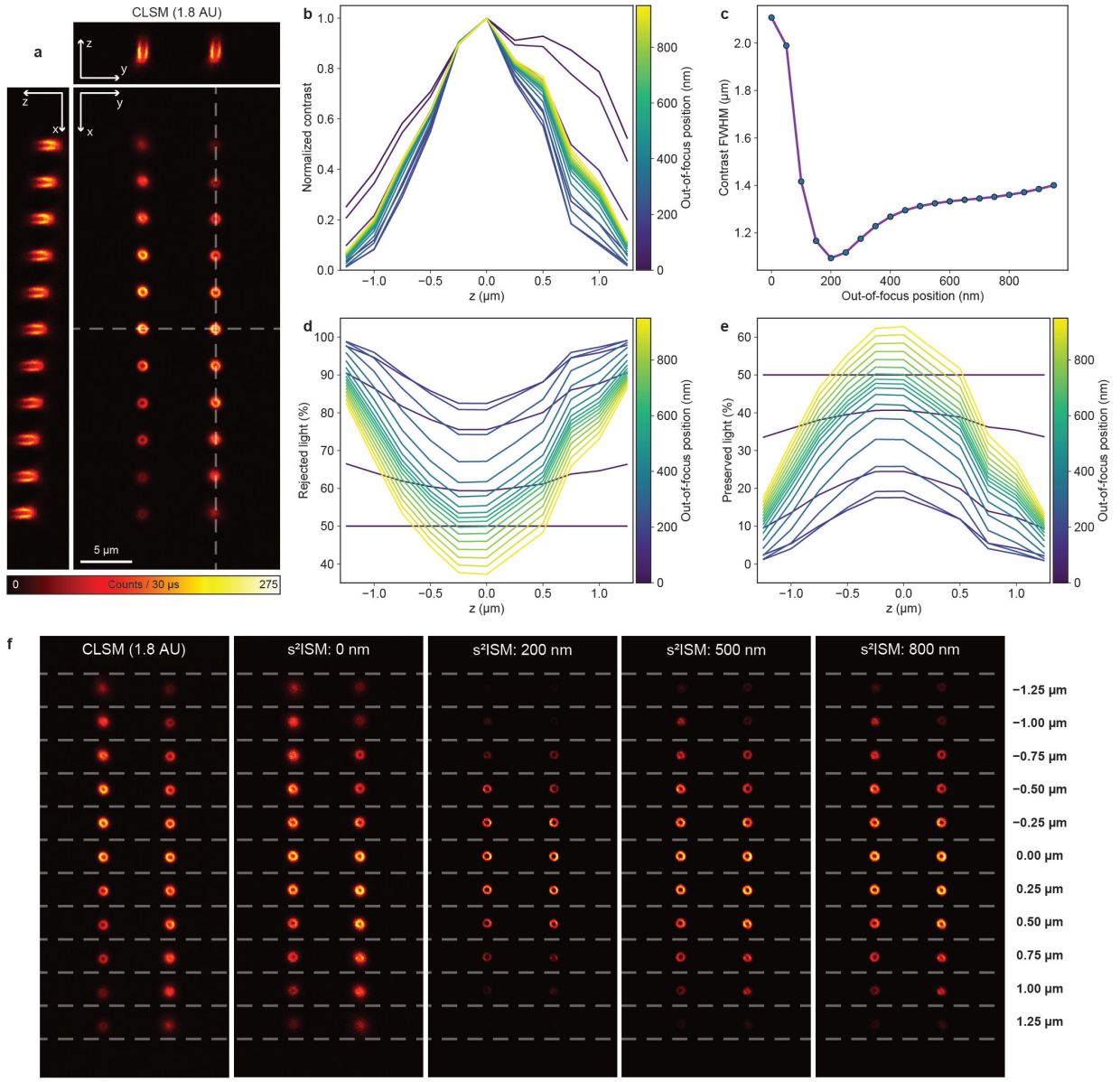

Figure B.3: **Argolight stairs: reconstruction at different out-of-focus positions.** **a**, slices of the stack of confocal images of the 3D crossing stairs target. The dashed lines indicate the slicing planes. **b**, Plot of photon counts of the rings at each axial positions, normalized with respect to the maximum. The colormap represents the curve obtained by reconstructing the image with s^{2}ISM using a a different out-of-focus value. **c**, plot of the full with at half maximum (FWHM) of the curves represented in **b**. **d**, plot of rejected photons relative to the total photon counts at different axial positions. **e**, plot of preserved photons relative to the total photon counts at different axial positions. **f**, Reconstructed images of the central plane of the stack in **a**, obtained running s^{2}ISM with different out-of-focus-values. For each reconstruction we used 20 iterations.

C

Extended figures

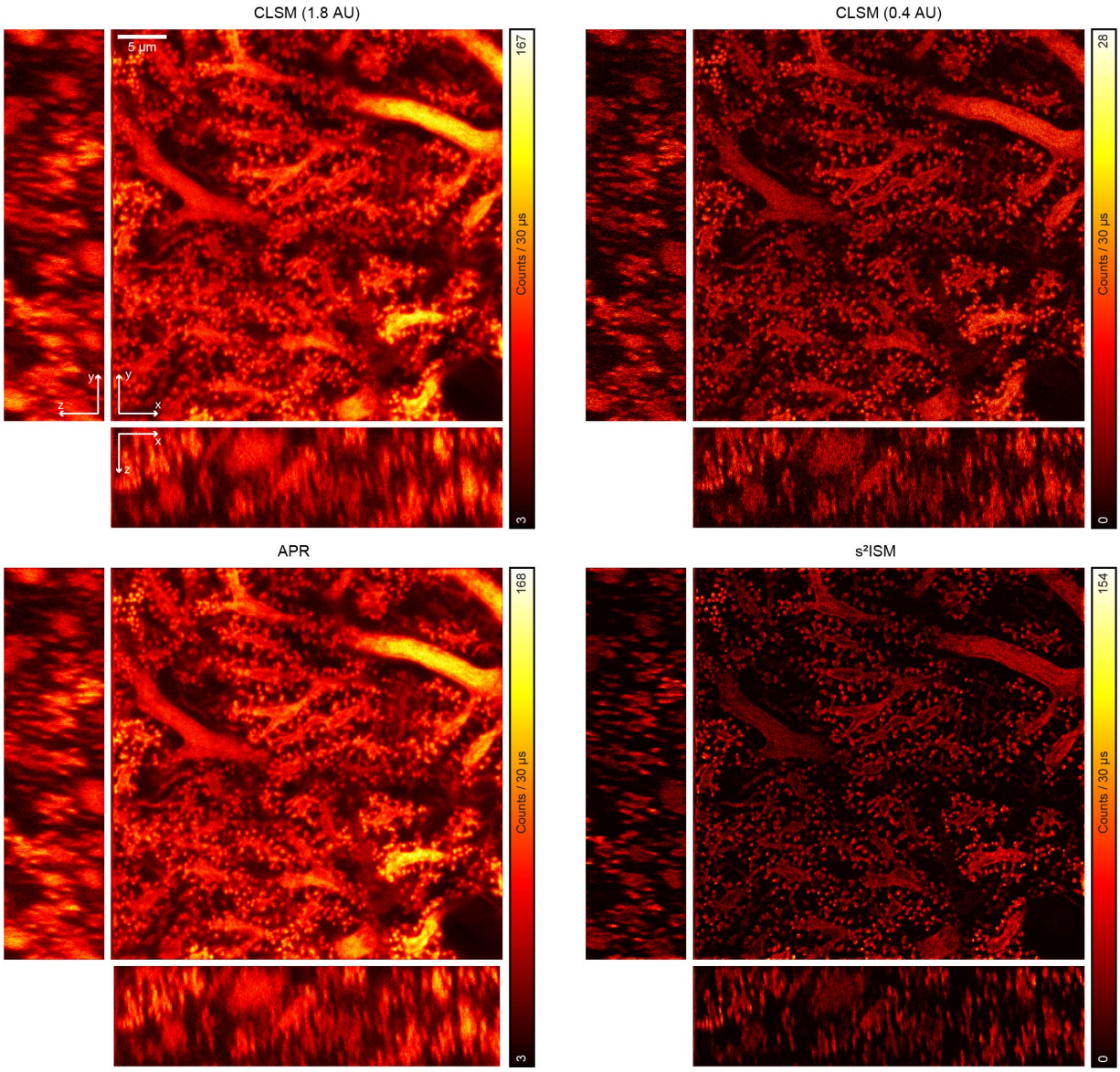

Figure C.1: **Volumetric reconstruction: Purkinje cells in a cerebellum slice.** Stack of images of Purkinje cells in a mouse cerebellum slice, obtained with linear excitation at $\lambda_{\text{exc}} = 488\,\text{nm}$ and reconstructed plane-by-plane. Field-of-view: $10.25\,\mu\text{m} \times 40\,\mu\text{m} \times 40\,\mu\text{m}$ (zxy), image size: $50 \times 1000 \times 1000$ pixels (zxy), s^{2}ISM and Multi-Image deconvolution iterations $= 20$ per plane. The three slices (xy, xz, yz) are taken from the center of the stack.

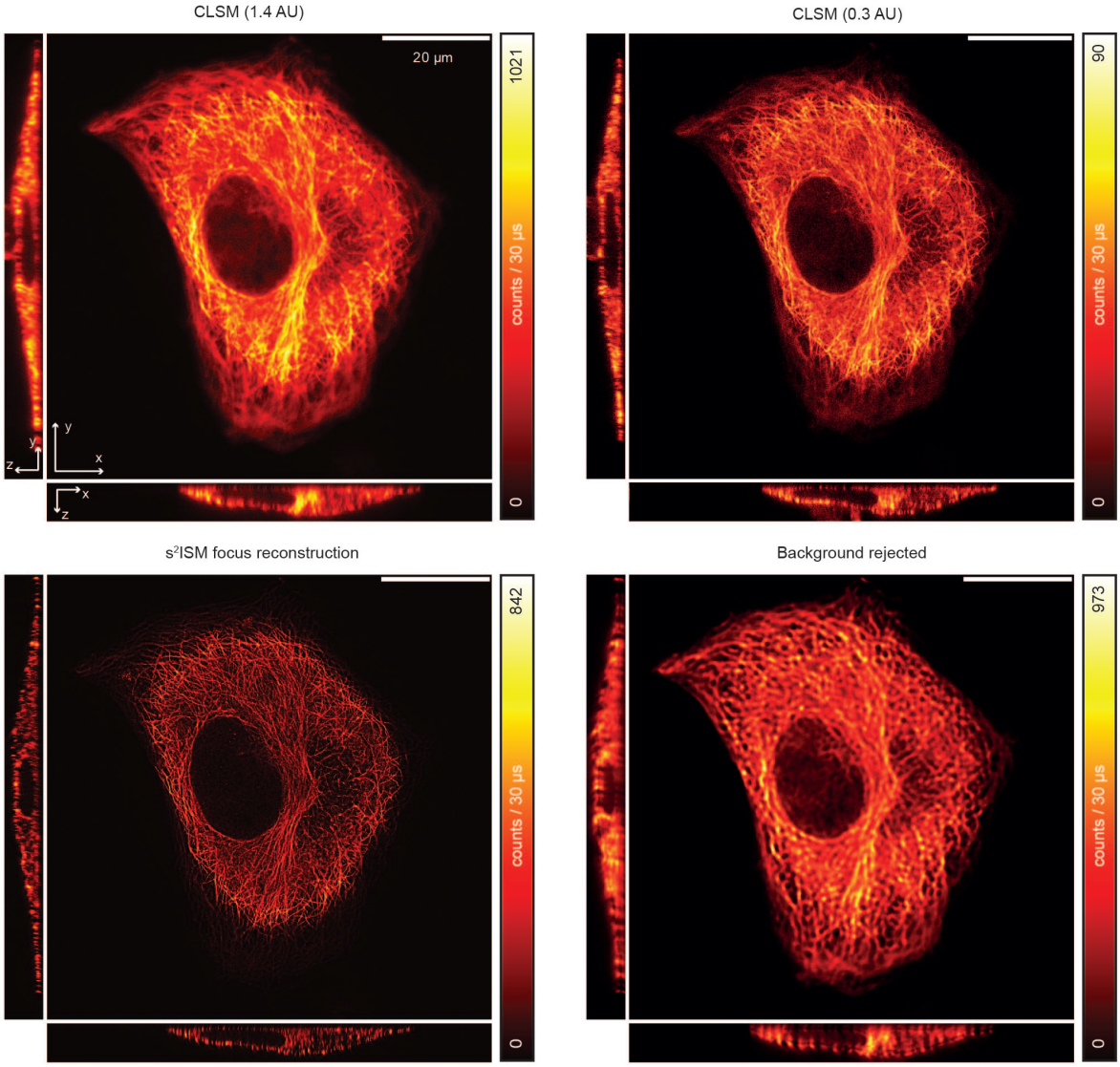

Figure C.2: **Volumetric reconstruction: HeLa cell.** Stack of the tubulin network of a HeLa cell, obtained with linear excitation at $\lambda_{\text{exc}} = 640\,\text{nm}$ and reconstructed plane-by-plane. Field-of-view: $7.2\,\mu\text{m} \times 86\,\mu\text{m} \times 86\,\mu\text{m}$ (zxy), image size: $56 \times 1075 \times 1075$ pixels (zxy), s^{2}ISM and Multi-Image deconvolution iterations $= 10$ per plane. The three slices (xy, xz, yz) are taken from the center of the stack.

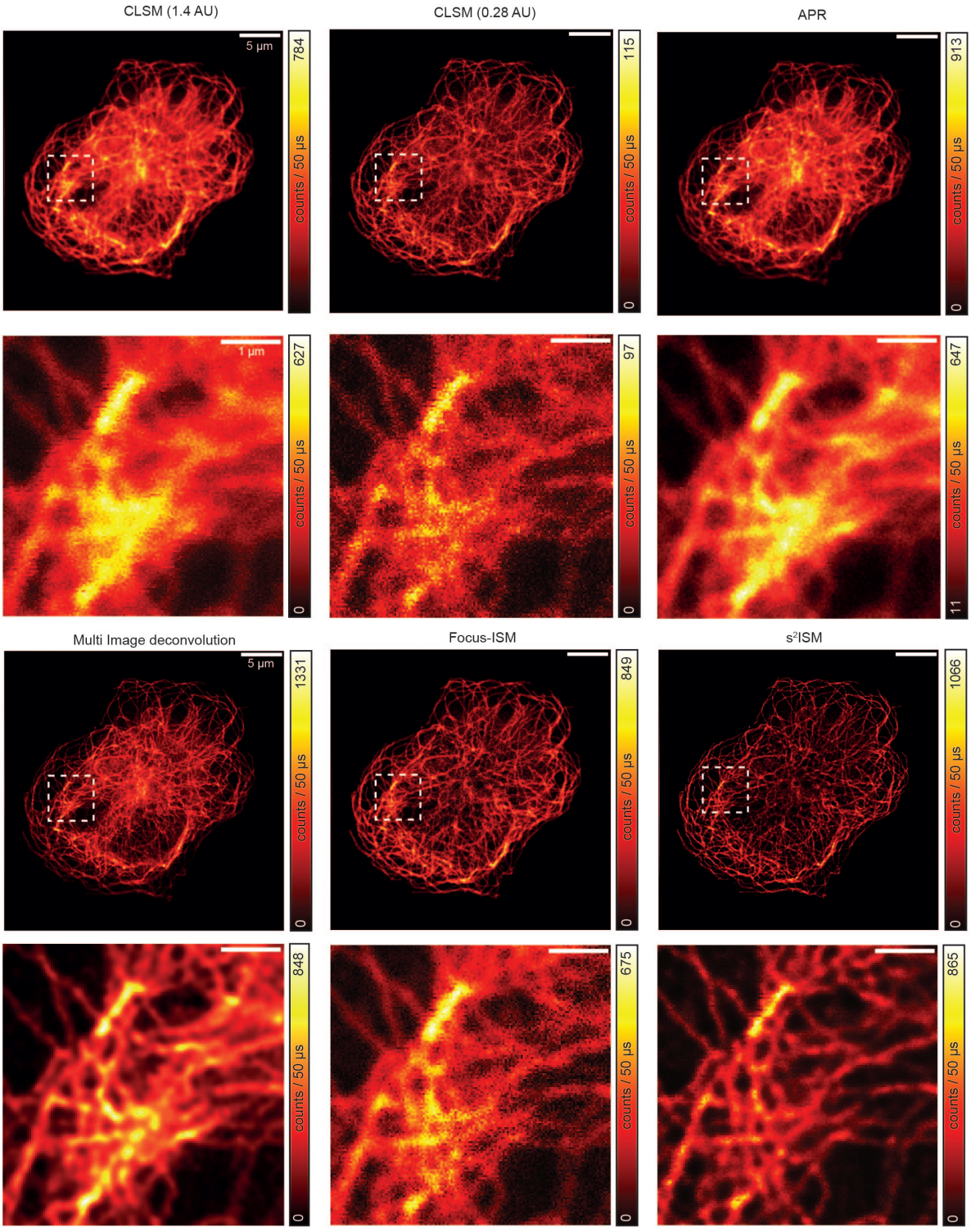

Figure C.3: **Extended data: tubulin network of a HeLa cell.** Extended images from Fig. **??**. Field-of-view: $35\,\mu\mathrm{m} \times 35\,\mu\mathrm{m}$. Image size: 875×875 pixels. Pixel dwell time: $50\,\mu\mathrm{s}$. Excitation laser: $\lambda = 640\,\mathrm{nm}$, CW. Average power at the sample plane: $1.2\,\mu\mathrm{W}$. s^{2}ISM and Multi-Image deconvolution iterations: 20.

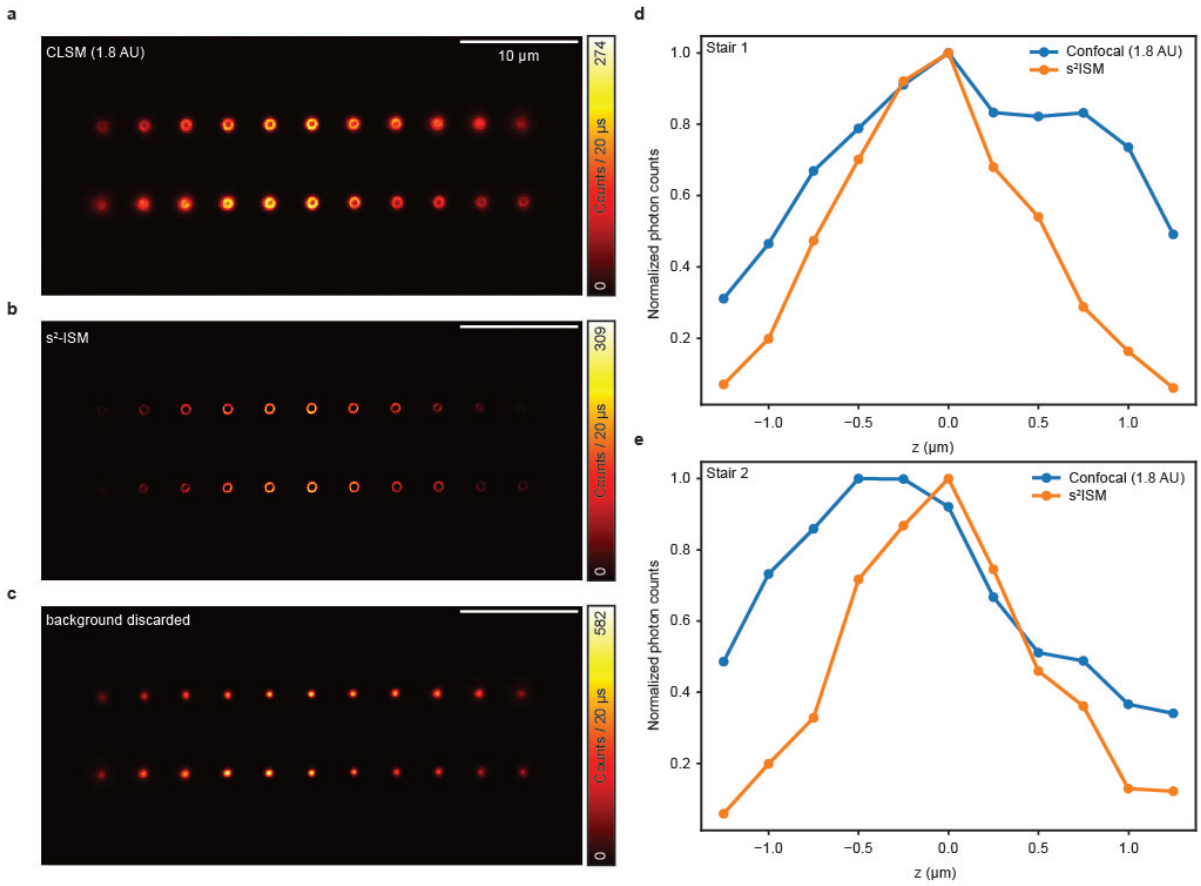

Figure C.4: **Extended data: axially spaced stairs.** Extended images from Fig. 6.3b. Field-of-view: $60\,\mu\mathrm{m} \times 60\,\mu\mathrm{m}$. Image size: 1250×1250 pixels. Pixel dwell time: $20\,\mu\mathrm{s}$. Excitation laser: $\lambda = 488\,\mathrm{nm}$, CW. Average power at the sample plane: $32.8\,\mu\mathrm{W}$. **a**, open-pinhole confocal image. **b**, in-focus s^{2}ISM reconstruction (20 iterations). **c**, discarded out-of-focus plane. **d**, normalized contrast of the steps of the upper stair. **e** normalized contrast of the steps of the lower stair.

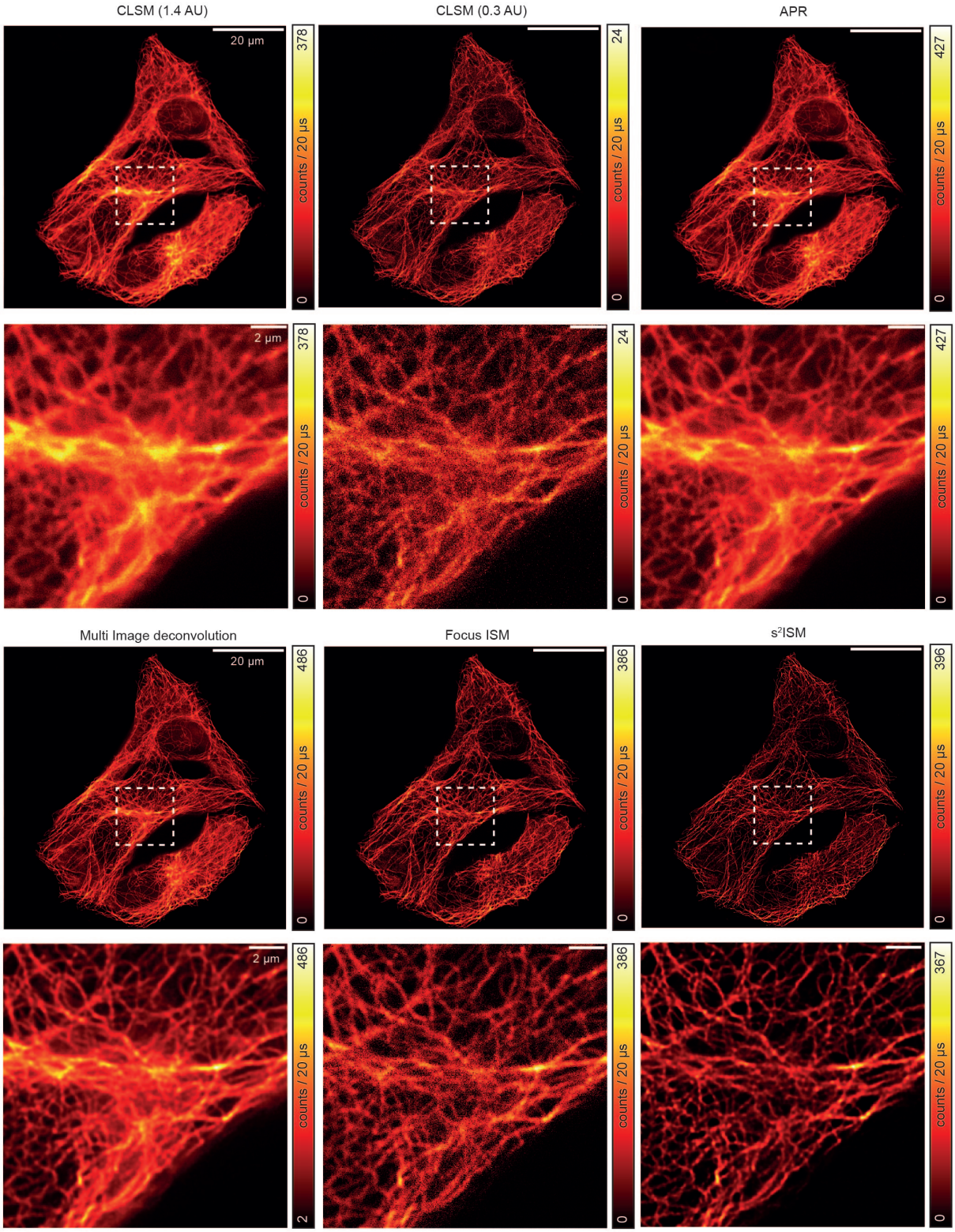

Figure C.5: **Extended data: tubulin network of a group of HeLa cells.** Extended images from Fig. 6.3a-b. Field-of-view: $80\,\mu\mathrm{m} \times 80\,\mu\mathrm{m}$. image size: 2000×2000 pixels. Pixel dwell time: $20\,\mu\mathrm{s}$. Excitation laser: $\lambda = 640\,\mathrm{nm}$, CW. Average power at the sample plane: $2.6\,\mu\mathrm{W}$. s^{2}ISM and Multi-Image deconvolution iterations: 20.

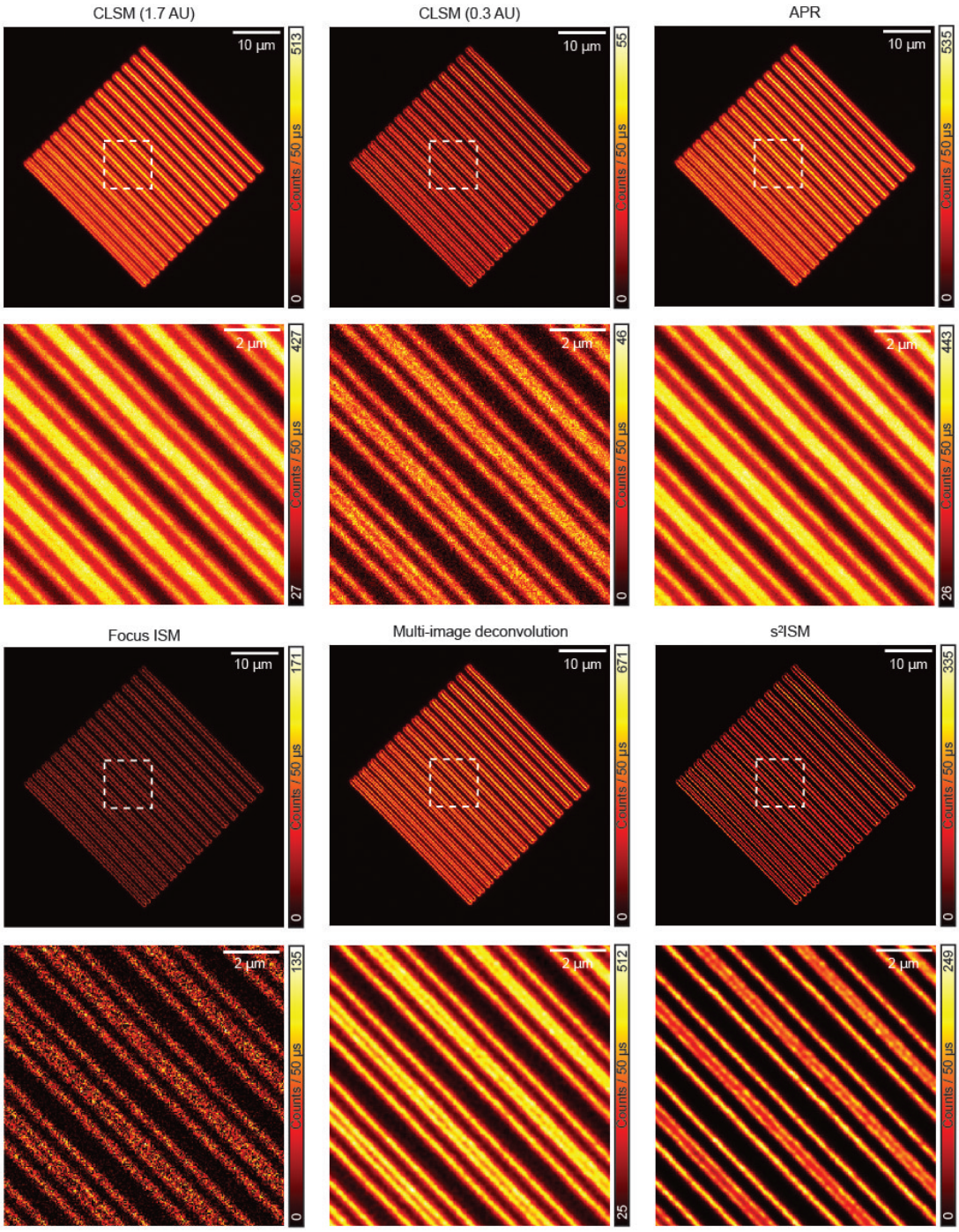

Figure C.6: **Extended data: gradually spaced lines.** Extended images from Fig. 6.3a. Field-of-view: $60\,\mu\mathrm{m} \times 60\,\mu\mathrm{m}$. Image size: 1500×1500 pixels. Pixel dwell time: $50\,\mu\mathrm{s}$. Excitation laser: $\lambda = 488\,\mathrm{nm}$, CW. Average power at the sample plane: $10.9\,\mu\mathrm{W}$. s^{2}ISM and Multi-Image deconvolution iterations: 20.

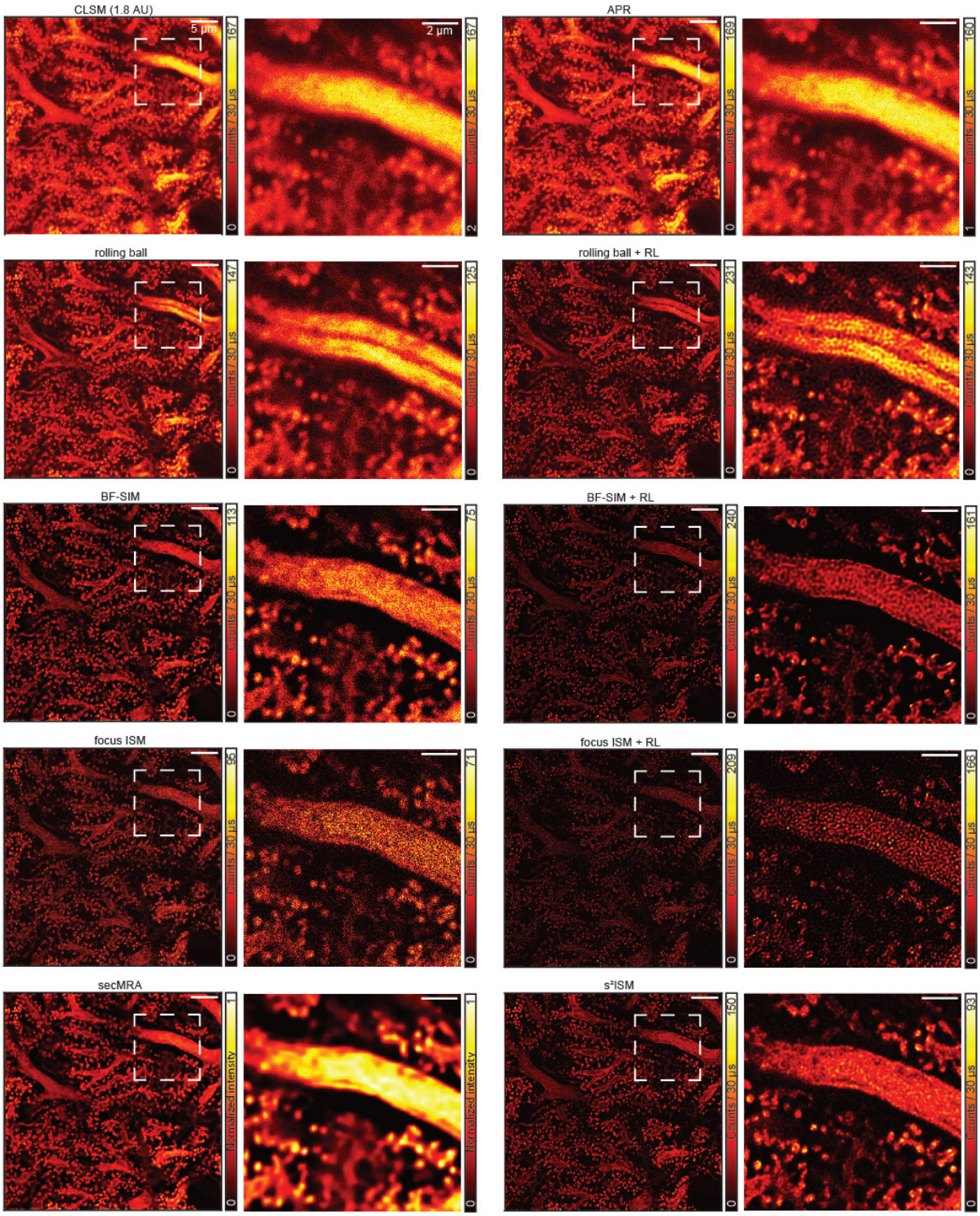

Figure C.7: **Comparison of background removal algorithms on experimental data.** Reconstruction of a single slice of the data from Supp. Fig. C.1. We used the same set of algorithms as in the previous figure, with an additional step of Richardson-Lucy (RL) deconvolution (20 iterations) on the output of the algorithms that do not inherently perform deconvolution. Non-ISM-specific algorithms cannot properly remove out-of-focus light and might generate artefacts. Notice the channel in the axon generated by the rolling ball algorithm as a reconstruction artefact.

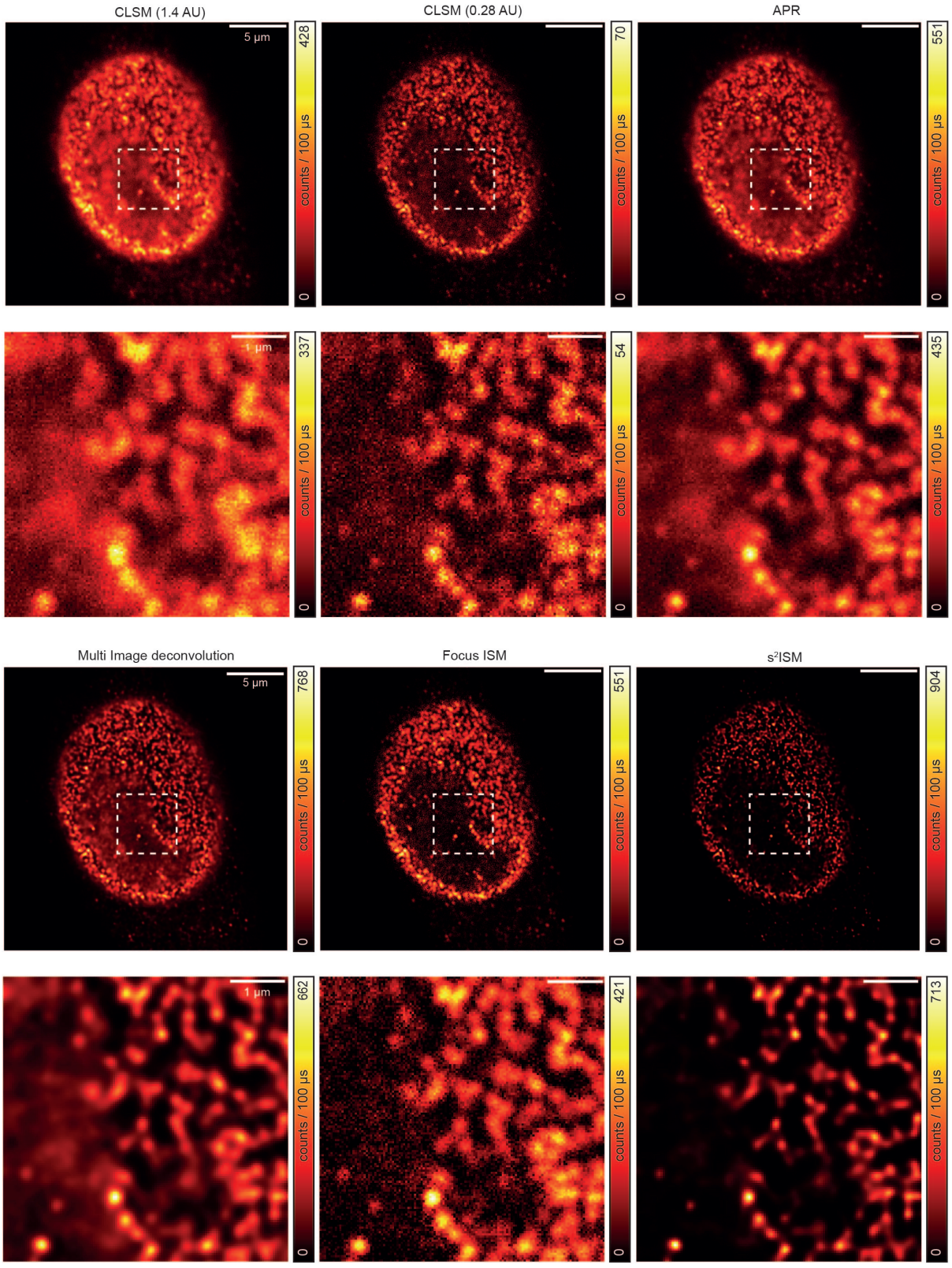

Figure C.8: **Extended data: nuclear pore complexes in a HeLa cell.** Extended images from Fig. 3.19b. Field-of-view: $25\,\mu\mathrm{m} \times 25\,\mu\mathrm{m}$. Image size: 625×625 pixels. Pixel dwell time: $100\,\mu\mathrm{s}$. Excitation laser: $\lambda = 640\,\mathrm{nm}$, CW. Average power at the sample plane: $14.7\,\mu\mathrm{W}$. s^{2}ISM and Multi-Image deconvolution iterations: 20.

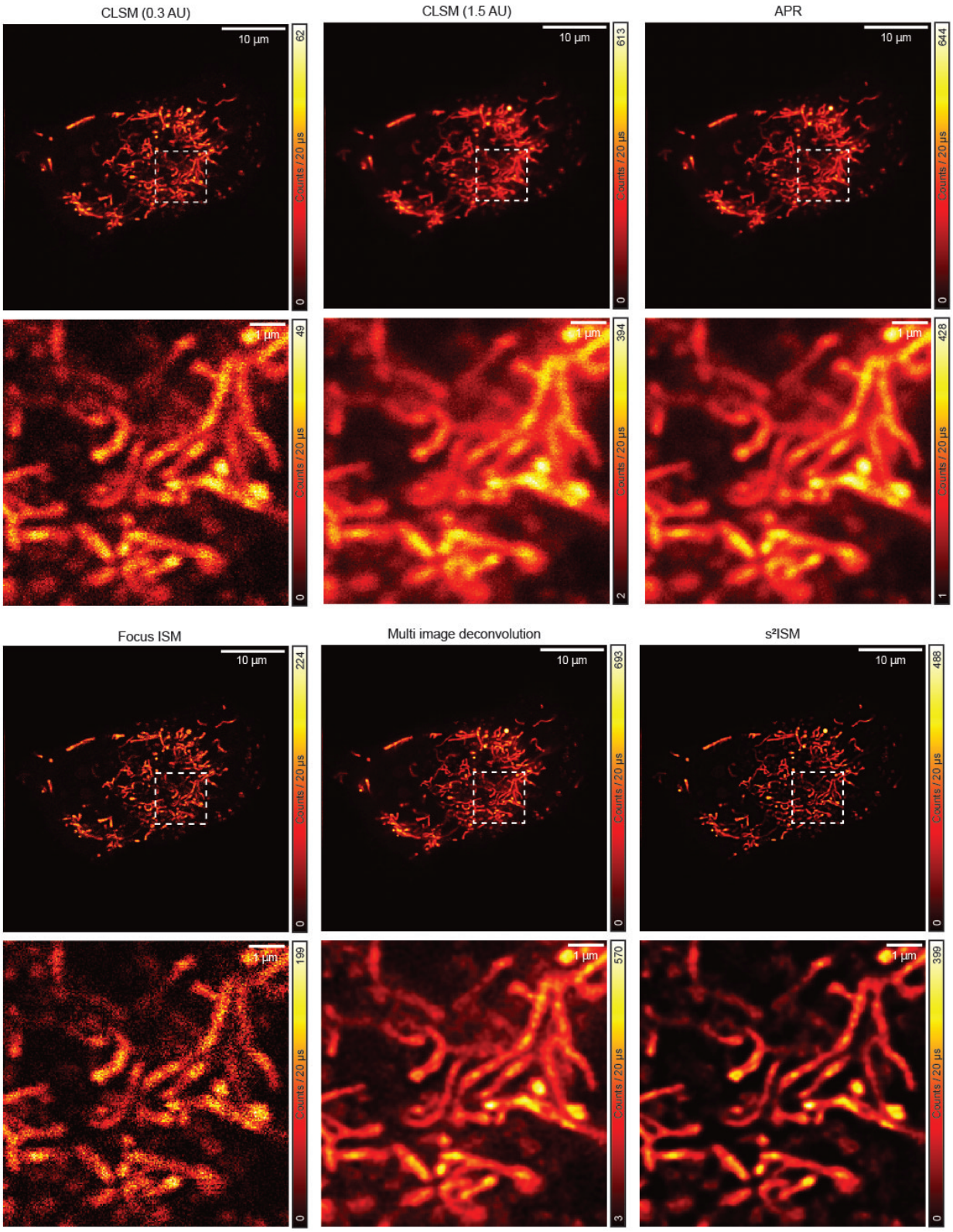

Figure C.9: **Extended data: live-cell imaging of mitochondria.** Extended images from a single frame of the sequence in Fig. 3.19c. Field-of-view: $60\,\mu\mathrm{m} \times 60\,\mu\mathrm{m}$. Image size: 1500×1500 pixels. Pixel dwell time: $20\,\mu\mathrm{s}$. Excitation laser: $\lambda = 561\,\mathrm{nm}$, CW. Average power at the sample plane: $1.8\,\mu\mathrm{W}$. Framerate: $25\,\mathrm{seconds/frame}$. s^{2}ISM and Multi-Image deconvolution iterations: 10.

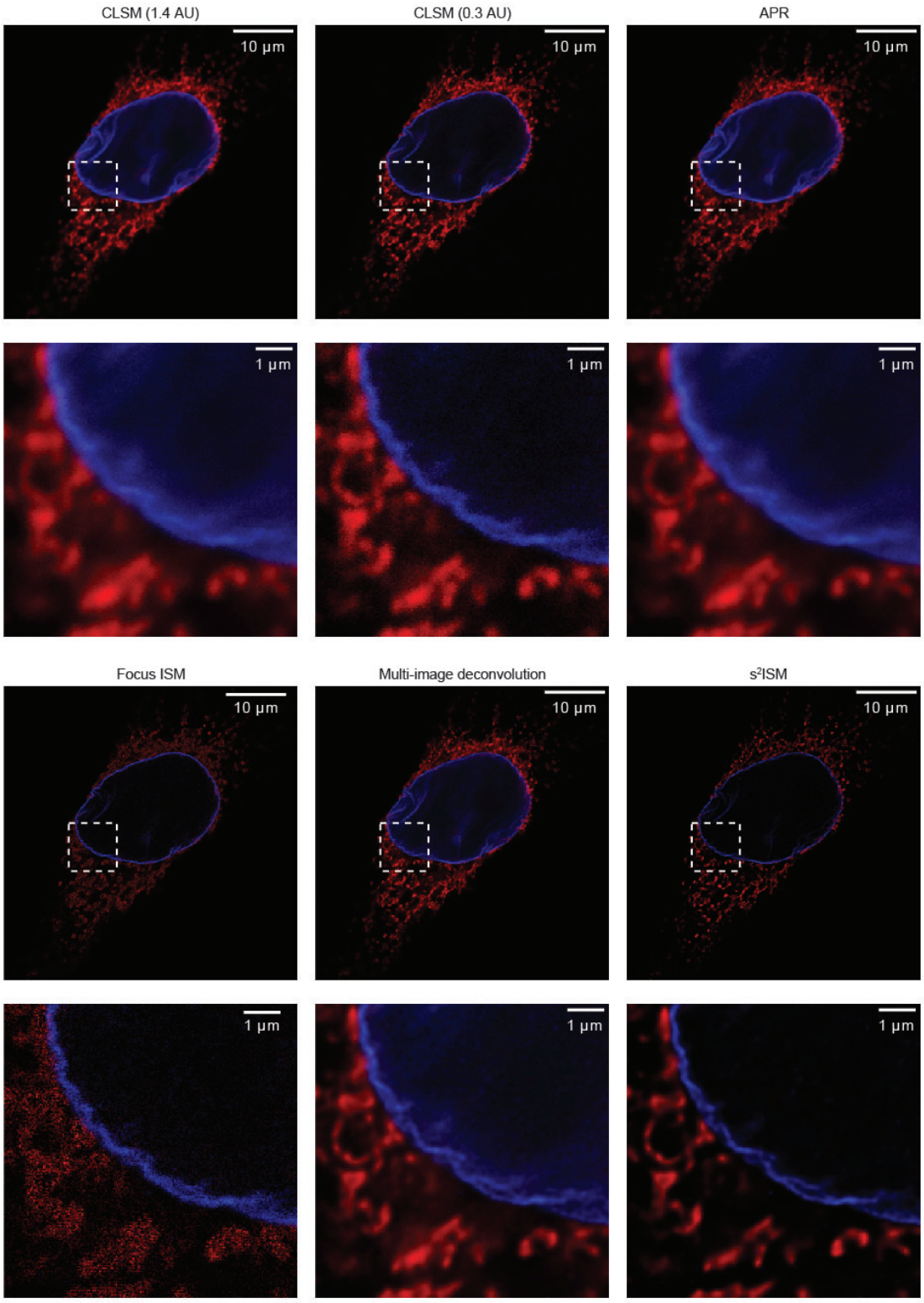

Figure C.10: **Extended data: multi-color imaging of a HeLa cell.** Extended images from Fig. 3.19d. Field-of-view: $65\,\mu\mathrm{m} \times 65\,\mu\mathrm{m}$. Image size: 1625×1625 pixels. Pixel dwell time: $20\,\mu\mathrm{s}$ for both channels. Blue excitation laser: $\lambda = 488\,\mathrm{nm}$, CW. Average power at the sample plane: $1.6\,\mu\mathrm{W}$. Red excitation laser: $\lambda = 640\,\mathrm{nm}$, CW. Average power at the sample plane: $4.2\,\mu\mathrm{W}$. s^{2}ISM and Multi-Image deconvolution iterations: 20.

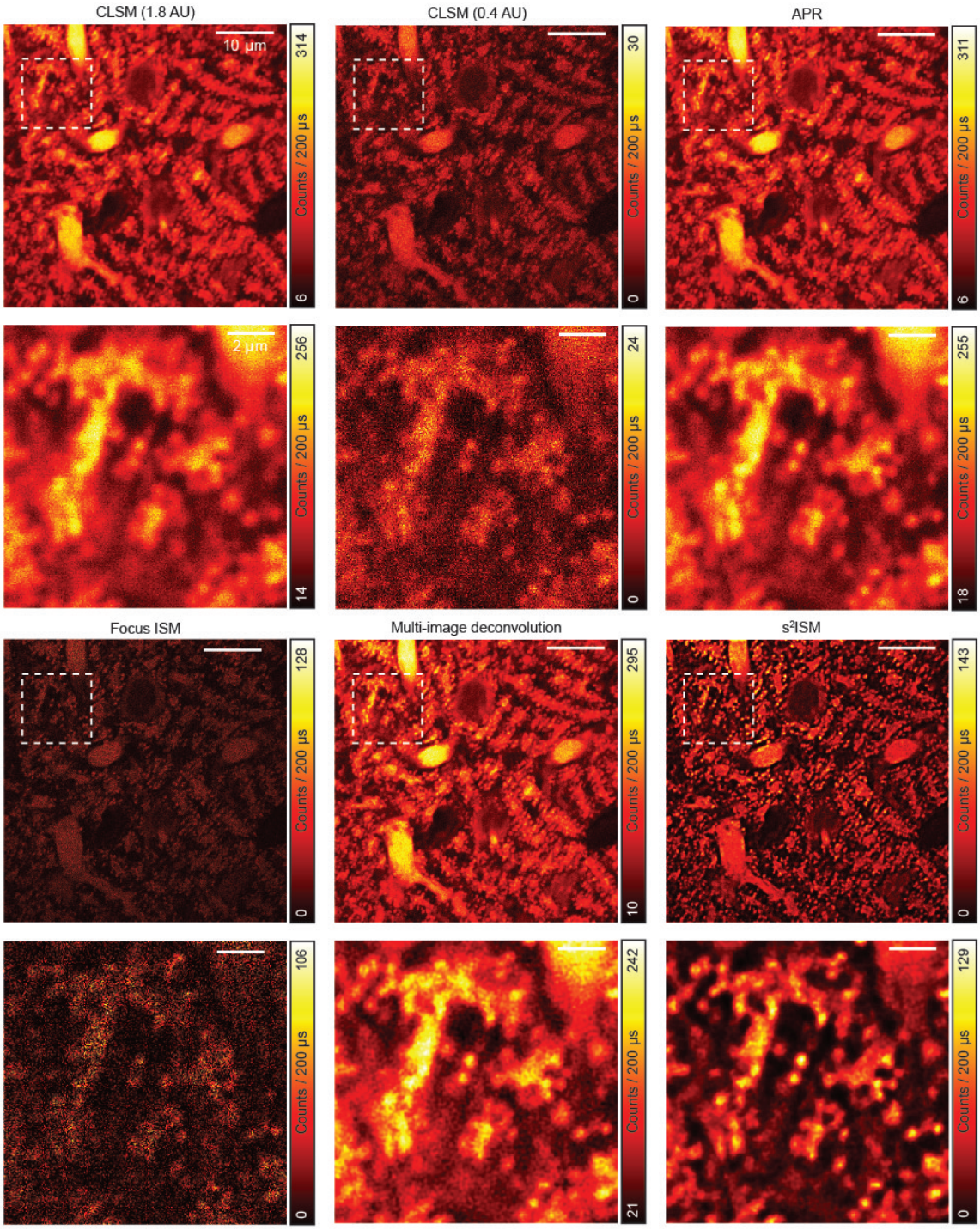

Figure C.11: **Extended data: two-photon excitation imaging of Purkinje cells in a cerebellum slice.** Extended images from Fig. 3.19e. Field-of-view: $50\,\mu\mathrm{m} \times 50\,\mu\mathrm{m}$. Image size: 1250×1250 pixels. Excitation laser: $\lambda = 900\,\mathrm{nm}$, pulsed at $80\,\mathrm{MHz}$. Average power at the sample plane: $10.8\,\mathrm{mW}$. s^{2}ISM and Multi-Image deconvolution iterations: 20.

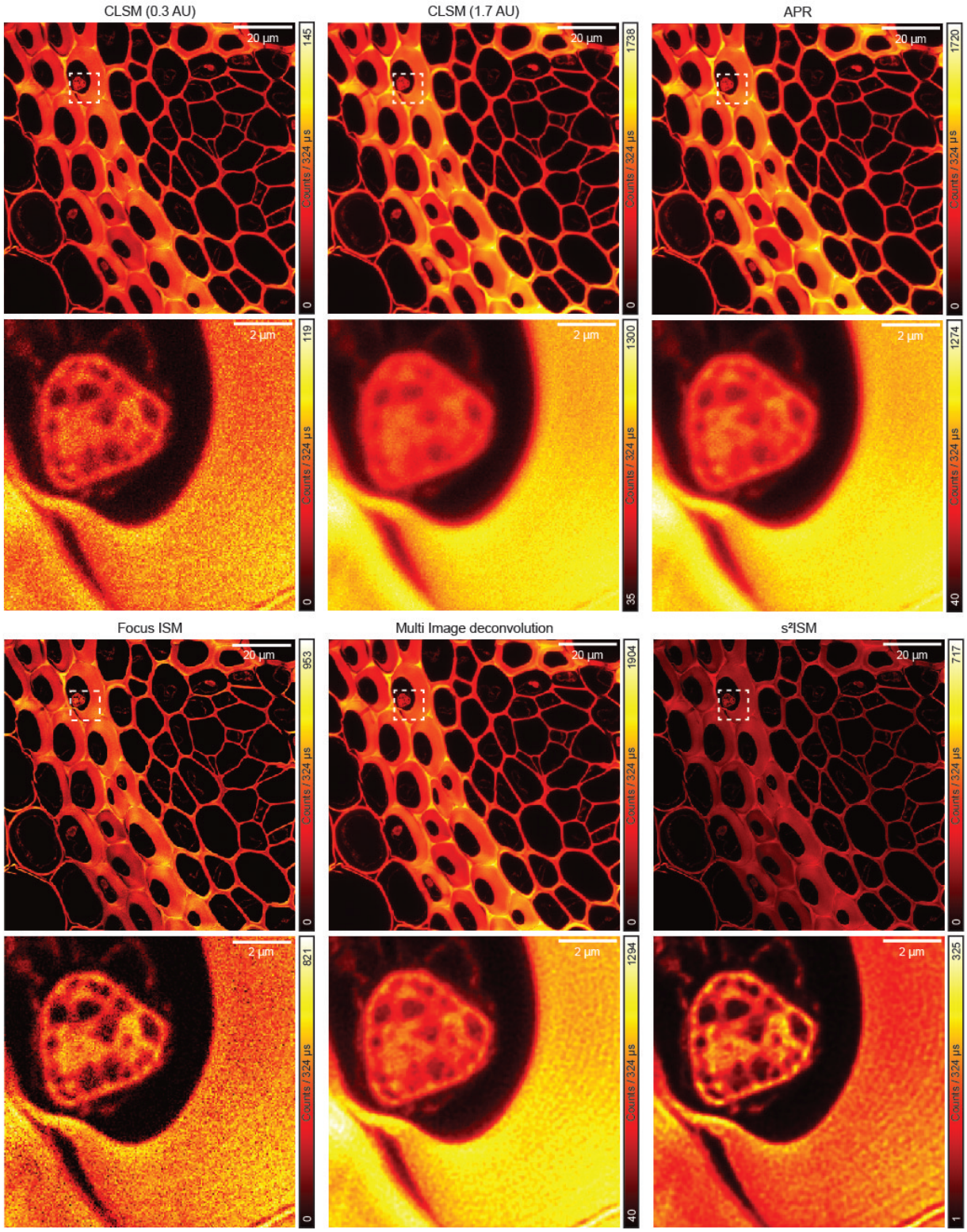

Figure C.12: **Extended data: Convallaria Majalis rhizome.** Extended images from Fig. 4.4b. Field-of-view: $100\,\mu\mathrm{m} \times 100\,\mu\mathrm{m}$. Image size: 2000×2000 pixels. Pixel dwell time: $324\,\mu\mathrm{s}$. Excitation laser: $\lambda = 488\,\mathrm{nm}$, pulsed at $40\,\mathrm{MHz}$. Average power at the sample plane: $9.1\,\mathrm{nW}$. s^{2}ISM and Multi-Image deconvolution iterations: 10.

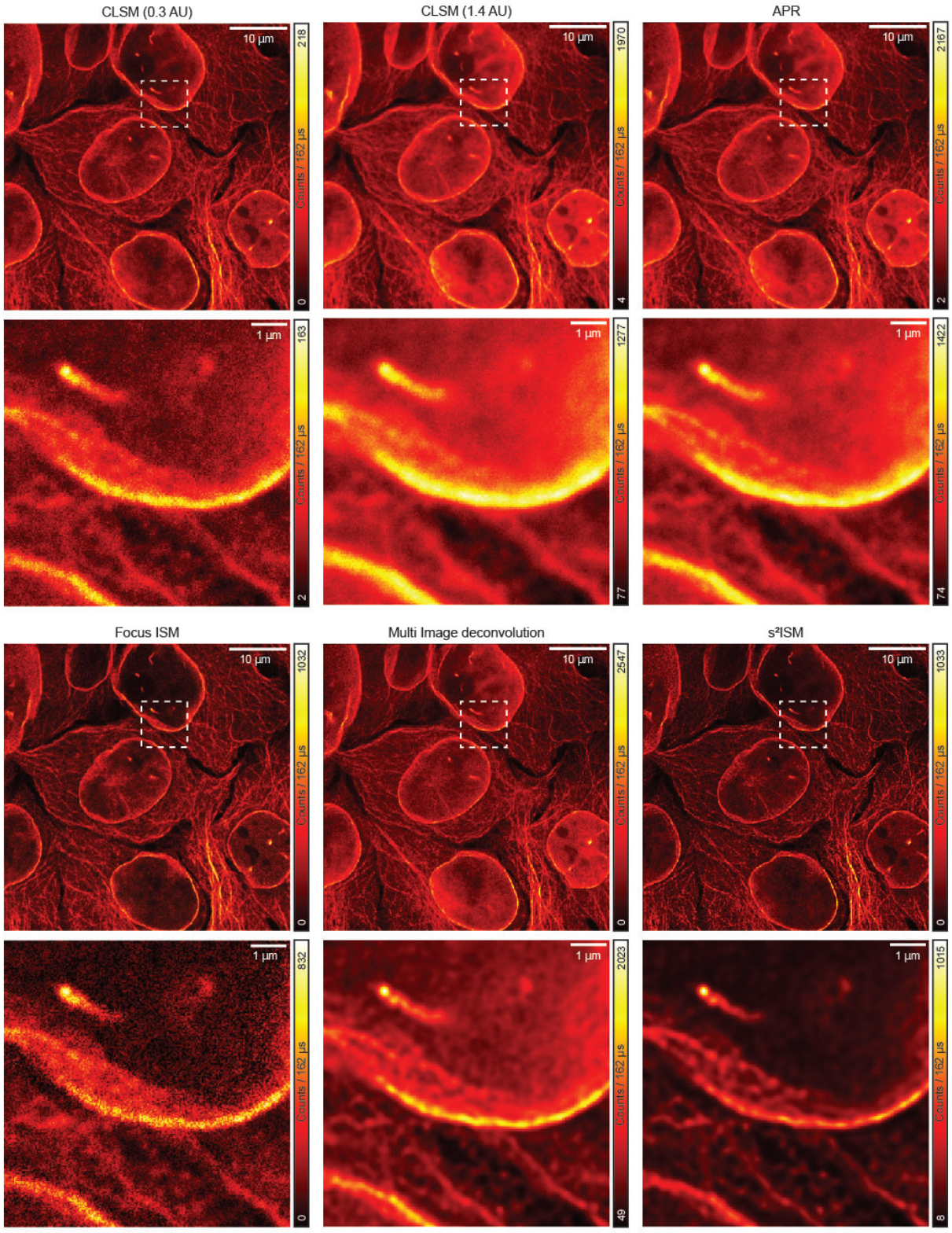

Figure C.13: **Extended data: nuclei and tubulin network in Hela cells.** Extended images from Fig. 4.4c. Field-of-view: $50\,\mu\mathrm{m} \times 50\,\mu\mathrm{m}$. Image size: 1250×1250 pixels. Pixel dwell time: $162\,\mu\mathrm{s}$. Excitation laser: $\lambda = 640\,\mathrm{nm}$, pulsed at $40\,\mathrm{MHz}$. Average power at the sample plane: $0.7\,\mu\mathrm{W}$. s^{2}ISM and Multi-Image deconvolution iterations: 10.

D

Non-ISM-specific state of the art algorithms for optical sectioning

D.1 Rolling ball

The rolling-ball algorithm provides a systematic method for estimating background intensity, particularly in settings where illumination is spatially non-uniform. Introduced by S. R. Sternberg in 1983 [1], it has since become a widely adopted tool in biomedical image processing for both background estimation and background suppression. In this framework, the image is interpreted as a three-dimensional surface: each pixel corresponds to a vertical column of unit height elements, and the height of each column is proportional to the pixel's intensity value.

To obtain the background intensity at a given pixel, one conceptually places a sphere (the "rolling ball") beneath the image surface at the corresponding location. As the sphere is raised until it is completely constrained by the surrounding intensity surface, the point on the sphere closest to the image surface determines the background estimate at that position. By continuously translating this sphere beneath the entire image, one obtains a smooth background surface that approximates the slowly varying intensity component.

The principal user-specified parameter required for the algorithm is the radius of the sphere (equivalently, the radius of the structuring element). This radius governs the scale at which variations are interpreted as background rather than signal, and therefore plays a central role in ensuring the correctness and stability of the background estimation process.

D.2 Sectioning MRA

Sectioning MRA (Multi-Resolution Algorithm) consists of a multi-step approach for image denoising and reconstruction [2]. It starts by applying a framelet transform on the raw data and then a thresholding on it to perform denoising. Thus, the output is reconstructed through a regularized iterative deconvolution implemented for Gaussian noise. The loss to minimize is defined as follows:

$$\arg\min_x \|f - Ax\|_2^2 + \lambda_1 \|Wx\|_1 + \lambda_2 \|Cx\|_1$$

where f is the input degenerated image, A denotes the blur kernel in matrix form, x is the recovered image, W denotes the framelet transform, C denotes the curvelet transform, λ_1 and λ_2 are two regularization parameters, $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the ℓ_1 and ℓ_2 norms, respectively. Reducing the value of the first term, also called the fidelity term, recovers high-frequency information. The second and third terms are the sparsity in the framelet and curvelet domains, respectively, which mainly control the noise. FISTA is employed to minimize this optimization problem.

References

[1] Sternberg. Biomedical image processing. *Computer*, 16, 1983.

[2] Baolei Liu and Fan Wang. Multi-resolution analysis for high-fidelity deconvolution microscopy. *Light: Science & Applications*, 14, 01 2025.

E Samples

E.1 Argolight calibration slide

We used the Argo-SIM v1 slide (Argolight). In detail, we imaged two patterns [1]. The first is the resolution target of gradually spaced lines rotated by $45°$. The second is a 3D crossing stair axially spaced by $250\,\mathrm{nm}$.

E.2 Cerebellum slice

Thy1-EGFP transgenic mouse brain is fixed in 4% paraformaldehyde in PBS (w/v) overnight at $4\,°\mathrm{C}$. 300–$350\,\mu\mathrm{m}$ vibratome sections were generated by sectioning tissues embedded in 2% agarose with a vibrating microtome (Leica) and permeabilized in 2% TritonX-100 in PBS (v/v) one day at $35\,°\mathrm{C}$. RapiClear 1.52 (SunJin Lab Co) was used to clear sections according to the manufacturer's instructions, and iSpacer (SunJin Lab Co) was applied to the sample mounting.

E.3 Cell culture

We cultured HeLa cells in Dulbecco's Modified Eagle Medium (DMEM, Gibco, ThermoFisher Scientific) supplemented with 10% fetal bovine serum (Sigma-Aldrich) and 1% penicillin/streptomycin (Sigma-Aldrich) at 37°C in 5% CO_2. The day before the staining, we seeded HeLa cells on coverslips in a 12-well plate (Corning Inc., Corning, NY) for immunostaining or a μ-Slide eight-well plate (Ibidi, Grafelfing, Germany) for live-cells imaging.

Fixed cells HeLa cells were fixed with either ice methanol, when cytoskeletal proteins were imaged, for 20 minutes at -20°C, or with a solution of 3.7% paraformaldehyde (Sigma-Aldrich) in phosphate-buffered saline (PBS, Gibco™, ThermoFisher Scientific) buffer for 15 min at room temperature. Cells were washed three times with PBS buffer and treated with blocking buffer (5% bovine serum albumin (BSA, Sigma-Aldrich) supplemented with 0.2% Triton X-100 in PBS buffer) for 1 hour at room temperature. Cells were incubated with primary antibodies diluted in the blocking buffer for 1 hour at room temperature. The primary antibodies used in this study were: monoclonal mouse anti-α-tubulin antibody (1:1000, Sigma-Aldrich), rabbit polyclonal anti-lamin B1 antibody (Abcam, ab16048, 1:500), rabbit polyclonal Nup-153 antibody (Abcam, ab84872, 1:500) and mouse monoclonal anti-ATP Synthase β antibody (Sigma, A9728, 1:250). After incubation with the antibody, cells were washed three times with blocking buffer and incubated with a secondary antibody diluted into blocking buffer for 1 hour at room temperature. The secondary antibodies used in this study were: Anti-Mouse IgG-Abberior STAR Red (Abberior, 1:1000), Anti-Mouse IgG-Abberior STAR 635P (Abberior, 1:1000), anti-Rabbit IgG Alexa-488 (ThermoFisher Scientific, 1:1000), Anti-Mouse IgG-Alexa647 (ThermoFisher Scientific, 1:500). We rinsed HeLa cells three times in PBS for 15 min. Finally, we mounted the coverslips onto microscope slides (Avantor, VWR International) with ProLong Diamond Antifade Mountant (Invitrogen, ThermoFisher Scientific).

Live cells For the mitochondrial staining in living cells, seeded HeLa cells were incubated with MitoTracker™ Orange (ThermoFisher Scientific) at a concentration of 100 nM in DMEM supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin for 10 minutes at 37°C in 5% CO_2.After the incubation, cells were washed three times with PBS and placed in live cells imaging solution (LICS, ThermoFisher Scientific) immediately before the measurement.

E.4 Zebrafish embryo handling and preparation

The zebrafish line used has a fluorescent protein (eGFP) encoded at the 3' end of the lap3b gene which is expressed under the control of the β-acting promoter (β-actin:LAP2B-EGFP). This results into a labeled nuclear envelope. Zebrafish were maintained and raised under standard conditions, and according to Swiss regulations (canton Vaud, license number VD-H28). Embryos were collected immediately upon fertilization. The chorion was chemically removed by incubating the embryos for 3 minutes in Pronase ($1.5\,\mathrm{mg/mL}$, Sigma-Aldrich, 107433), diluted in 0.3× Danieau's buffer, after which embryos were allowed to develop to high stage at 28 °C in 0.3× Danieau's buffer. Embryos were fixed in 4% (w/v) paraformaldehyde (PFA) in PBST (PBS 1X supplemented with Tween20 pH 7.4, Sigma-Aldrich) at 4°C overnight. Fixed embryos were washed 3 times with PBST and deyolked manually with forceps. The embryos were kept in a blocking buffer solution (4% BSA in PBST) for 1 hour at room temperature. Embryos were then dehydrated by incubating them in 50% and then 100% methanol in PBST and incubated overnight at $-20\,^{\circ}\mathrm{C}$. To continue, embryos were rehydrated gradually into blocking buffer. Embryos were finally mounted in Vectashield Antifade Mounting Medium with 4,6-diamidino-2-phenylindole (Vector Laboratories, VC-H-1200-10) before imaging. The β-actin:LAP2B-EGFP

fish line has been a gift from the laboratory of Dr. Caren Norden [2].

References

[1] Argolight. Argolight help center.

[2] Katrin Daniel, Jaroslav Icha, Cindy Horenburg, Doris Müller, Caren Norden, and Jörg Mansfeld. Conditional control of fluorescent protein degradation by an auxin-dependent nanobody. *Nature Communications*, 9(1), August 2018.

F

J-invariant framework

We model the signal y and its noisy measurement x as a pair of random variables in $\mathbb{R}^m$. If $J \subset \{1, \ldots, m\}$ is a subset of the dimensions, we write x_J for x restricted to J.

Definition. Let $\mathcal{J}$ be a partition of the dimensions $\{1, \ldots, m\}$ and let $J \in \mathcal{J}$. A function $f : \mathbb{R}^m \to \mathbb{R}^m$ is *J-invariant* if $f(x)$ does not depend on the value of x_J. It is *$\mathcal{J}$-invariant* if it is J-invariant for each $J \in \mathcal{J}$.

We propose minimizing the self-supervised loss

$$\mathcal{L}(f) = \mathbb{E}\|f(x) - x\|^2, \tag{1}$$

over $\mathcal{J}$-invariant functions f. Since f has to use information from outside of each subset of dimensions J to predict the values inside of J, it cannot merely be the identity.

Proposition 1. *Suppose x is an unbiased estimator of y, i.e. $\mathbb{E}[x|y] = y$, and the noise in each subset $J \in \mathcal{J}$ is independent from the noise in its complement J^c, conditional on y. Let f be $\mathcal{J}$-invariant. Then*

$$\mathbb{E}\|f(x) - x\|^2 = \mathbb{E}\|f(x) - y\|^2 + \mathbb{E}\|x - y\|^2. \tag{2}$$

That is, the self-supervised loss is the sum of the ordinary supervised loss and the variance of the noise. By minimizing the self-supervised loss over a class of $\mathcal{J}$-invariant functions, one may find the optimal denoiser for a given dataset.

Proof 1. In an ideal situation for signal reconstruction, we have a prior $p(y)$ for the signal and a probabilistic model of the noisy measurement process $p(x|y)$. After observing some measurement x, the posterior distribution for y is given by Bayes' rule:

$$p(y|x) = \frac{p(x|y)p(y)}{\int p(x|y)p(y)\,dy}$$

In practice, one seeks some function $f(x)$ approximating a relevant statistic of $y|x$, such as its mean or median. The mean is provided by the function minimizing the loss:

$$E_x \left[f(x) - y\right]^2$$

(The L^1 norm would produce the median) (Murphy, 2012).

Fix a partition J of the dimensions $\{1, \ldots, d\}$ of x and suppose that for each $J \in J$, we have

$$p(x|y) = p(x_J|y)p(x_{\bar{J}}|y),$$

i.e., x_J and $x_{\bar{J}}$ are independent conditional on y. We consider the loss

$$E_x \left[||f(x) - x||^2\right] = E_x \left[||f(x) - y||^2\right] + ||f(x) - y||^2 - 2\left(f(x), x - y\right)$$

If f is J-invariant, then for each j the random variables $f(x_j)$ and x_j are independent. The third term reduces to

$$\sum_j E_x \left(E_{y|x}[f(x_j) - y_j]\right) \left(E_{y|x}[x_j - y_j]\right),$$

which vanishes when $E[x_j|y] = y_j$.

For example, if the signal is an image with independent, mean-zero noise in each pixel, we may choose

$$\mathcal{J} = \{\{1\}, \ldots, \{m\}\}$$

to be the singletons of each coordinate. Then "donut" median filters, with a hole in the center, form a class of $\mathcal{J}$-invariant functions, and by comparing the value of the self-supervised loss at different filter radii, we can optimize the optimal radius for denoising the image at hand.

We may search over all J-invariant functions for the global optimum:

Proposition 2. The J-invariant function f_J^* minimizing (1) satisfies

$$f_J^*(x)_J = \mathbb{E}[y_j|x_J]$$

for each subset $J \in \mathcal{J}$.

That is, the optimal J-invariant predictor for the dimensions of y in some $J \in \mathcal{J}$ is their expected value conditional on observing the dimensions of x outside of J.

F.1 Blindly calibrating denoisers

Many denoising models have a hyperparameter controlling the degree of the denoising—the size of a filter, the threshold for sparsity, the number of principal components. If ground truth data were available, the optimal parameter θ for a parameter-dependent denoiser said to be f_θ can be retrieved by calculating

$$\|f_\theta(x) - y\|^2 \tag{F.1}$$

Without ground truth, we may nevertheless compute the self-supervised loss

$$\|f_\theta(x) - x\|^2. \tag{F.2}$$

For general denoisers this is not related to the ground-truth error but for J-invariant denoisers this error is equal to the ground-truth error plus the noise variance as proved before.

More generally, let g_θ be any classical denoiser, and let $\mathcal{J}$ be any partition of the pixels such that neighboring pixels are in different subsets. Let $s(x)$ be the function replacing each pixel with the average of its neighbors. Then the function f_θ defined by

$$f_\theta(x) := g_\theta\left(1_{\mathcal{J}} \cdot s(x) + 1_{\mathcal{J}^c} \cdot x\right)_{\mathcal{J}},$$

for each $\mathcal{J} \in \mathcal{J}$, is a $\mathcal{J}$-invariant version of g_θ. Indeed, since the pixels of x in $\mathcal{J}$ are replaced before applying g_θ, the output cannot depend on $x_{\mathcal{J}}$.

G

Richardson-Lucy algorithm

The Richardson-Lucy algorithm is just an Expectation-Maximization algorithm for Poisson data, when the linear relation between data and ground-truth is expressed as a convolution operator with the respective blurring kernel.

The expectation-maximization algorithm is an iterative method for solving the problem of maximizing a given probability density. The first application of this method is maximum likelihood estimation, but EM applies equally to the maximization of the posterior distribution in a Bayesian framework. Let Y be the random vector corresponding to the observed data y_η, with probability distribution $p_Y(y; x)$, where $x = (x_1, \ldots, x_m)$ is the vector of unknown parameters in the parameter space $\Omega \subseteq \mathbb{R}^m$.

The EM algorithm is an algorithm with numerous applications that provides an iterative procedure for computing the maximum likelihood (ML) estimate in situations where, if all the data were available, the ML estimate would be simple. Therefore, in this context, the observed data vector is considered incomplete and is regarded as an observable function of the so-called complete data. The notion of incomplete data includes the conventional meaning of missing data, but it also applies to situations in which the complete data represent what would be available under an ideal experiment. In the latter case, the complete data may contain some variables that are never strictly observable.

In this framework, we denote by z the vector containing the complete data, also called augmented data, and we denote by w the vector containing the additional data, thought of as unobservable or missing data. Thus $z = (y_\eta, w)$. As will become evident from the example of maximum likelihood estimation for a multivariate Poisson variable, even when a problem does not appear to be incomplete, the computation of the ML estimate is often considerably facilitated by its artificial formulation. This is because the EM algorithm exploits the reduced complexity of ML estimation for complete data. For many statistical problems, the complete-data probability has a readily usable form.

Let $p_Z(z;x)$ be the probability distribution of the random vector Z corresponding to the complete data vector z. The complete-data log-likelihood that could be formed for x if z were fully observable is given by

$$\log L_c(x,z) = \log p_Z(z;x).$$

Formally, we have two sample spaces Z and Y and a many-to-one mapping $h : Z \to Y$. Instead of observing the complete data vector z in Z, we observe the incomplete data vector $h(z) = y$ in Y. It follows that the probability of observing y is given by

$$p_Y(y;x) = \int_{h^{-1}(y)} p_Z(z;x)\, dz$$

where $h^{-1}(y) = \{z : h(z) = y\}$ is the subset of Z determined by the equation $h(z) = y$.

The EM algorithm addresses the problem of indirectly solving the incomplete-data likelihood equation by proceeding iteratively and maximizing an estimate of the complete-data log-likelihood function, $\log L_c(x,z)$ (based on the random variable depending on Z, which is not observable). The estimate is obtained by replacing the unknown function $\log L_c(x,z)$ with its conditional expectation given Y using the current estimate of x, that is,

$$\mathbb{E}_{Z|Y}(x^{(k)}) \left\{\log L_c(x,z)\right\}.$$

More precisely, let $x^{(k)}$ be the current estimate (the k-th iteration) of the variable x. The Expectation-Maximization algorithm proceeds as follows.

Expectation step (E-step): requires the computation of

$$Q(x;x^{(k)}) = \mathbb{E}_{Z|Y}\big(x^{(k)}\big)\{\log L_c(x,z)\},$$

where the conditional expectation depends on the parameters $x^{(k)}$ at the current iteration.
Maximization step (M-step): requires the maximization of Q with respect to x over the parameter space Ω. That is, we choose $x^{(k+1)}$ such that

$$Q\big(x^{(k+1)};x^{(k)}\big) \geq Q\big(x;x^{(k)}\big)$$

for every $x \in \Omega$.
Theory proved that the sequence of likelihood values

$$\{L(x^{(k)}, y_\eta)\}_{k=0,1,2,\ldots}$$

generated by the EM method is monotonically increasing. Since the likelihood sequence is monotone non-decreasing and upper-bounded, and $\mathbb{R}$ with the Euclidean distance is complete; it follows that the likelihood converges.

Expectation Maximization algorithm for Poisson data

We begin by considering the observable variables: the counts $y_{\eta_1}, \ldots, y_{\eta_n}$ are conditionally independent according to a Poisson distribution, that is,

$$y_{\eta_i} \sim \text{Poisson}\big((Tx)_i\big) \tag{35}$$

where

$$(Tx)_i = \sum_{j=1}^{m} T_{ij} x_j,$$

and T_{ij} is interpreted as the conditional probability that a photon is counted by the j-th detector and emitted from within the i-th pixel.

The advantage of this statistical approach is that it takes into account the Poisson nature underlying the data. The coefficients in the Poisson regression model, T_{ij}, appear as conditional probabilities that depend on the geometry and physics of the detection system.

An obvious choice for the complete data vector in this example is the vector z, consisting of the counts of the unobservable data z_{ij}, where z_{ij} is defined as the number of photons emitted within the i-th pixel and recorded by the j-th detector, with $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Thus,

$$Z_{ij} \sim \text{Poisson}(T_{ij} x_j),$$

that is, given x, the Z_{ij} are conditionally independent.

Since

$$y_{\eta_i} = \sum_{j=1}^{m} z_{ij},$$

it is clear that these assumptions for the unobservable data z_{ij} imply model (35) for the incomplete data y_{η_i}.

To recover the general description, we may interpret the complete data vector as a completion of the observable data vector y_η through the missing data w, namely $z = (y_\eta, w)$.

The complete-data log-likelihood is given by

$$\log L_c(x, z) = \sum_{j=1}^{m} \sum_{i=1}^{n} \left\{ -T_{ij} x_j + z_{ij} \log(T_{ij} x_j) - \log(z_{ij}!) \right\}.$$

Since the Poisson distribution belongs to the linear exponential family, the likelihood is linear in the unobservable data z. Therefore, the E-step (at iteration $k+1$) simply requires computing the conditional expectation of Z given the observed quantities y_η, evaluated at the current estimate $x^{(k)}$.

It can be shown that the conditional distribution of Z given y and $x^{(k)}$ is binomial. Indeed,

$$P(Z_{ij} = z_{ij} \mid Y_i = y_{\eta_i}) = \frac{P(Z_{ij} = z_{ij}, \sum_{k=1}^{m} Z_{ik} = y_{\eta_i})}{P(Y_i = y_{\eta_i})}.$$

Using independence and the Poisson distributions, one obtains

$$P(Z_{ij} = z_{ij} \mid Y_i = y_{\eta_i}) = \frac{y_{\eta_i}!}{z_{ij}!(y_{\eta_i} - z_{ij})!}(1 - p_{ij})^{y_{\eta_i} - z_{ij}} p_{ij}^{z_{ij}},$$

where the binomial parameter is

$$p_{ij} = \frac{T_{ij}x_j}{(Tx)_i}.$$

The expected value of the binomial distribution is $p_{ij}y_{\eta_i}$, and therefore

$$\mathbb{E}_{Z|Y}(x)\big(Z_{ij} \mid y_\eta\big) = \frac{x_j T_{ij} y_{\eta_i}}{(Tx)_i}.$$

Evaluating the conditional expectation at $x^{(k)}$, we define

$$z_{ij}^{(k)} := \mathbb{E}_{Z|Y}(x^{(k)})\big(Z_{ij} \mid y_\eta\big) = \frac{x_j^{(k)} T_{ij} y_{\eta_i}}{(Tx^{(k)})_i}.$$

To perform the E-step of the EM algorithm we compute the conditional expectation of the log-likelihood:

$$Q(x; x^{(k)}) = \mathbb{E}_{Z|Y}(x^{(k)}) \log L_c(x)$$

$$= \sum_{j=1}^{m}\sum_{i=1}^{n}\left\{-T_{ij}x_j + z_{ij}^{(k)} \log(T_{ij}x_j) - \mathbb{E}_{Z|Y}(x^{(k)}) \log(z_{ij}!)\right\}.$$

Now we apply the M-step. We compute

$$\frac{\partial Q(x; x^{(k)})}{\partial x_p} = \sum_{j=1}^{m}\sum_{i=1}^{n}\left\{-T_{ij}\delta_{pj} + \frac{z_{ij}^{(k)} T_{ij}\delta_{pj}}{T_{ij}x_j}\right\}.$$

Hence,

$$\sum_{i=1}^{n}\left\{-T_{ip} + \frac{z_{ip}^{(k)}}{x_p}\right\} = -(T^*\mathbf{1})_p + \frac{\sum_{i=1}^{n} z_{ip}^{(k)}}{x_p} = 0.$$

Therefore,

$$x_p = \frac{\sum_{i=1}^{n} z_{ip}^{(k)}}{(T^*\mathbf{1})_p},$$

which would be the simple estimate from the complete-data log-likelihood if the variables $z^{(k)}$ were known.

Instead, we substitute $z^{(k)}$ with their conditional expectations and obtain

$$x_j^{(k+1)} = \frac{1}{(T^*\mathbf{1})_j}\sum_{i=1}^{n}\mathbb{E}_{Z|Y}(x^{(k)})\big(Z_{ij} \mid y_\eta\big) = \frac{1}{(T^*\mathbf{1})_j}\sum_{i=1}^{n}\frac{x_j^{(k)} T_{ij} y_{\eta_i}}{(Tx^{(k)})_i}$$

$$= \frac{x_j^{(k)}}{(T^*\mathbf{1})_j} \left(T^* \frac{y_\eta}{Tx^{(k)}} \right)_j ,$$

which is the Expectation Maximization algorithm for Poisson data. When the forward operator T is chosen as the convolution operator, we obtain the famous Richardson-Lucy algorithm. In the context of microscopy, where the convolution is often intended as the blurring of the ground-truth, the operator acts as a convolution between ground truth x and the PSF h, namely

$$y(s) = (Tx)(s) = (h * x)(s) = \int_{\mathbb{R}^n} h(s-y)\, x(y)\, dy. \tag{G.1}$$

shaping the Richardson-Lucy algorithm as

$$x^{(k+1)} = \frac{x^{(k)}}{h^* * \mathbf{1}} \left(h^* * \frac{y_\eta}{h * x^{(k)}} \right).$$

H

Practical Adoption Guide

User guide pipeline

Here we show the pseudo Python code pipeline showing all the functions embedded in the package. To minimize the burden of work for the user and optimize reconstruction capabilities through physically reliable parameter estimation:

Algorithm 1 Data-Driven PSF Simulation

procedure SIMULATEPSFWITHALIGNMENT(data, exPar, emPar, grid, geometry='rect')
    $N \leftarrow \sqrt{\text{number of detector channels}}$ ▷ 1. Estimate optimal out-of-focus depth
    if z_{out} not provided **then**
        $(\text{pxsize}_z, \_) \leftarrow$ FindOutOfFocusFromParameters(grid.pxsizex, exPar, emPar)
    else
        $\text{pxsize}_z \leftarrow z_{\text{out}}$
    end if ▷ 2. Estimate geometric parameters from dataset
    grid_sim ← Initialize GridFinder(grid)
    grid_sim.Estimate(data, exPar.wl, emPar.wl, emPar.na) ▷ 3. Determine upsampling factor for PSF precision
    if downsample is True **then**
        $ups \leftarrow$ FindUpsampling(grid_sim.pxsizex)
    else
        $ups \leftarrow 1$
    end if
    ▷ 4. Crop central 3x3 detector ring for alignment
    shift_exp ← CropShift(grid_sim.shift, geometry)
    ▷ 5. Get theoretical shift vectors
    shift_theor ← ShiftMatrix(geometry)
    ▷ 6. Minimize loss to find magnification, rotation, mirror
    $(\alpha, \theta, mirror) \leftarrow$ LossMinimizer(shift_exp, shift_theor) ▷ 7. Update grid parameters according to optimized magnification.
    grid_sim.M ← α ▷ 8. Simulate 3D SPAD PSFs using updated parameters
    $(Psf, detPsf, exPsf) \leftarrow$ SPAD_PSF_3D(grid_sim, exPar, emPar)
    ▷ 9. Downsample PSFs to original pixel size
    if downsample is True **then**
        $Psf \leftarrow$ DownSample(Psf, ups)
        $detPsf \leftarrow$ DownSample(detPsf, ups)
        $exPsf \leftarrow$ DownSample(exPsf, ups)
    end if
    return $Psf, detPsf, exPsf, \alpha, \theta, mirror$
end procedure

Algorithm 2 Maximum Likelihood Reconstruction

function MaxLikelihoodReconstruction(dset, psf, max_iter, stop, threshold, initialization, rep_to_save)
    Convert `dset` and `psf` to internal tensor format: `data`, `h`
    Handle even/odd image dimensions by cropping the first row/column if needed
    Normalize PSF: $h \leftarrow h$/sum over lateral axes for every axial plane
    Flip PSF: $ht \leftarrow$ flip(h) along lateral axes
    if initialization $=$ `flat` **then**
        $O \leftarrow$ constant value $\propto$ total flux / number of pixels
    else if initialization $=$ `sum` **then**
        $O[z, ...] \leftarrow$ sum over input channels / number of planes
    end if
    Precompute FFTs: $h\_fft \leftarrow$ FFT(h), $ht\_fft \leftarrow$ FFT(ht)
    $k \leftarrow 0$, $pre\_flag \leftarrow$ True, $flag \leftarrow$ True
    Initialize photon counts and derivatives arrays: `counts`, `diff`
    while flag **do**
        Forward model: img_estimate $\leftarrow$ Convolve_FFT(h_fft, O)
        Compute ratio: fraction $\leftarrow$ img/img_estimate, set to 0 if denominator $< \epsilon$
        Backprojection update: update $\leftarrow$ Convolve_FFT$(ht_fft,$ fraction$)$
        Update object: $O\_new \leftarrow O *$ update
        Update stopping criteria: $pre\_flag, flag, counts[:, k], diff[:, k] \leftarrow$ CheckStopping$(O, O_new, pre_flag, flag, stop, threshold)$
        if rep_to_save $=$ all OR k $\in$ rep_to_save **then**
            Save O to output array
        end if
        Prepare next iteration: $O \leftarrow O\_new$, $k \leftarrow k + 1$
    end while
    return O, counts, diff, k
end function

There is also one function merging the two presented above:

Algorithm 3 Data-Driven ISM Reconstruction with PSF Alignment

procedure DATADRIVENRECONSTRUCTION(dset, gridPar, exPar, emPar, max_iter, stop, threshold, downsample, z_out_of_focus)

▷ 1. Check PSF availability and simulate if needed

$(psf, detPsf, exPsf) \leftarrow$ PSFEstimatorFromData(dset, exPar, emPar, gridPar, downsample=downsample, z_out_of_focus=z_out_of_focus) ▷ 2. Crop 3x3 central ring for alignment

shift_exp $\leftarrow$ CropShift(psf.shift, geometry=gridPar.geometry) ▷ 3. Generate theoretical shift vectors for the central ring

shift_theor $\leftarrow$ ShiftMatrix(geometry=gridPar.geometry) ▷ 4. Optimize magnification, rotation, and mirror

$(\alpha, \theta, mirror) \leftarrow$ LossMinimizer(shift_exp, shift_theor)

gridPar.M $\leftarrow \alpha$ ▷ update magnification in grid parameters ▷ 5. Launch iterative maximum-likelihood reconstruction

Initialize reconstructed object O_0 based on `initialization` ('flat' or 'sum')

for $i = 1$ to max_iter **do**

  $O_i \leftarrow$ UpdateObjectMLE(O_{i-1}, dset, psf)

  $counts_i \leftarrow$ ComputePhotonCounts(O_i, psf)

  $diff_i \leftarrow$ DerivativePhotonCounts($counts_i$)

  if stop == 'auto' **and** max($|diff_i|$) < threshold **then**

    break

  end if

end for

▷ 6. Return reconstructed object, photon statistics, and PSFs

return $O_i, counts, diff, k = i, psf$

end procedure

If the memory required to store the raw dataset and all associated computational structures exceeds the available system resources, the reconstruction can be performed using the batch-wise approach. This method partitions the raw dataset into smaller, manageable sub-volumes, which are then processed sequentially. By iterating over these sub-volumes and merging the results, the algorithm efficiently reconstructs the full dataset while respecting the memory limitations of the host system.

Algorithm 4 Batch-Wise Data-Driven ISM Reconstruction

procedure BATCHRECONSTRUCTION(dset, psf, batch_size, overlap, max_iter, stop, threshold, initialization, process)
  $(wx, wy) \leftarrow$ batch_size
  Initialize reconstruction array $O \leftarrow$ zeros(Nz, Nx, Ny, Nt, Nch)
  $nx \leftarrow$ number of batches along x-axis
  $ny \leftarrow$ number of batches along y-axis
  $k\_iter \leftarrow 1$
  for $i = 1$ to nx **do**
    for $j = 1$ to ny **do**
      Print "Processing batch $k\_iter/(nx \cdot ny)$"
      ▷ Determine overlap offsets
      $overlap_x \leftarrow$ overlap if $i > 1$ else 0
      $overlap_y \leftarrow$ overlap if $j > 1$ else 0
      ▷ Crop current batch from dataset
      $batch\_dset \leftarrow CropDataset(dset, i, j, wx, wy, overlap_x, overlap_y)$
      ▷ Perform maximum-likelihood reconstruction on batch
      $batch\_recon, *_{\leftarrow}$ MaxLikelihoodReconstruction(batch_dset, psf, stop=stop, max_iter=max_iter, threshold=threshold, initialization=initialization, process=process)
      ▷ Place reconstructed batch into the global reconstruction volume
      $MergeBatchIntoGlobal(O, batch\_recon, i, j, overlap_x, overlap_y)$
      $k\_iter \leftarrow k\_iter + 1$
    end for
  end for
  return O
end procedure

Computational scaling and complexity analysis

In our simulations, we typically choose the number of axial reconstruction planes to be $Z = 2$, corresponding to ideal optical setup conditions in most cases. For a dataset of size $400 \times 400 \times 25$ ($N_x \times N_y \times Ch$), this configuration yields approximately 3.8 iterations per second on a workstation equipped with a 10-core CPU (3.6 GHz), 16 GB of RAM, and a 512 GB SSD. The computational cost of the algorithm scales linearly with the number of axial planes, as each plane must be updated independently during every iteration. Consequently, doubling the number of axial planes roughly halves the number of iterations that can be processed per second. The lateral dimensions of the dataset play an even more significant role in the computational burden, since the dominant operations—such as pixel-wise updates and convolutions—scale with the number of lateral pixels. For a dataset of lateral size $N_x \times N_y$, the per-iteration cost behaves as $\mathcal{O}(N_x \times N_y)$; doubling the lateral dimension, $N_{x,y} \to 2N_{x,y}$, increases the cost by a factor of four, while the achievable number of iterations per second decreases accordingly. In addition, the total computational effort scales linearly with the number of iterations I, so that combining the effects of axial planes, lateral dimensions, and iterations, the overall computational complexity can be expressed as $\mathcal{O}(I\, Z\, N_x N_y Ch)$.

Within the s^{2}ISM pipeline, the optional simulation of synthetic PSFs introduces only a marginal overhead relative to the full iterative reconstruction. Specialized functions have been implemented to minimize the temporal cost of each step, accepting increased memory usage as a deliberate trade-off to improve execution speed. A particularly important factor in performance arises from the square geometry of the SPAD array detector. To accurately simulate rotations of the detector within the custom optical setup, the square pinhole configuration is first constructed on a downsampled lateral grid, then rotated, and finally upsampled back to the original resolution. Let k denote the upsampling factor. Since these operations affect only the two lateral dimensions, the effective grid size becomes $(kN_x) \times (kN_y)$, and the computational cost scales quadratically with k, supposing squared data, $\mathcal{O}((kN_x)^2) = \mathcal{O}(k^2 N_x^2)$. For example, choosing $k = 10$ increases the computational cost by a factor of $10^2 = 100$, significantly impacting total runtime.

When extending the dataset to include a temporal dimension, such as moving to a $400 \times 400 \times 91 \times 25$ ($N_x \times N_y \times t \times Ch$) tensor, the computational cost scales linearly with the number of temporal frames t. Each frame must be processed independently during each iteration, so that the overall complexity becomes $O(ItZN_xN_yCh)$. Consequently, doubling the number of temporal frames approximately halves the number of iterations that can be processed per second, and memory usage increases proportionally with t.