

Single-mask sphere-packing with implicit neural representation reconstruction for ultrahigh-speed imaging

NELSON DIAZ^{1†}, MADHU BENIWAL^{2†}, MIGUEL MARQUEZ², FELIPE GUZMAN¹, CHENG JIANG², JINYANG LIANG², AND ESTEBAN VERA^{1,*}

¹School of Electrical Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

²Centre Énergie Matériaux Télécommunications, Institut National de la Research Scientifique, Varennes, Québec J3X1S2, Canada

† Contributed equally to this work.

** esteban.vera@pucv.cl

Abstract:

Single-shot, high-speed 2D optical imaging is essential for studying transient phenomena in various research fields. Among existing techniques, compressed optical-streaking ultra-high-speed photography (COSUP) uses a coded aperture and a galvanometer scanner to capture non-repeatable time-evolving events at the 1.5 million-frame-per-second level. However, the use of a random coded aperture complicates the reconstruction process and introduces artifacts in the recovered videos. In contrast, non-multiplexing coded apertures simplify the reconstruction algorithm, allowing the recovery of longer videos from a snapshot. In this work, we design a non-multiplexing coded aperture for COSUP by exploiting the properties of congruent sphere packing (SP), which enables uniform space-time sampling given by the synergy between the galvanometer linear scanning and the optimal SP encoding patterns. We also develop an implicit neural representation—which can be self-trained from a single measurement—to not only largely reduce the training time and eliminate the need for training datasets, but also reconstruct far more ultra-high-speed frames from a single measurement. The advantages of this proposed encoding and reconstruction scheme are verified by simulations and experimental results in a COSUP system.

© 2025 Optical Society of America

1. Introduction

Single-shot ultrafast imaging has represented a groundbreaking approach for capturing ultrafast dynamic phenomena occurring in two-dimensional (2D) space [1], attracting substantial interest in physics [2], biology [3], chemistry [4], materials science [5], and engineering [6]. Among existing approaches, snapshot compressive temporal imaging (SCTI) physically captures a transient scene in a single measurement followed by computational reconstruction to retrieve high-speed frames. The optical implantation of compressed sensing (CS) overcomes conventional limitations in ultrafast optoelectronic devices [1]. It relies on encoding the dynamic scene—with either a single static coded aperture (CA) or multiple dynamic CAs [7]—during data acquisition. This encoding operation samples the spatio-temporal scene, which facilitates the reconstruction algorithm to correctly extract spatiotemporal features of the transient scene. For dynamic encoding, the imaging speed is limited by the pattern refreshing rate of the deployed spatial light modulator. In contrast, static encoding circumvents this restriction so that the imaging speed is determined by the shearing unit.

Over the past decade, SCTI systems based on single static encoders and various temporal shearing units have witnessed significant advancements [8, 9]. Compressed ultrafast spectral photography (CUSP) [10–13] and its variants have leveraged chirped ultrashort pulses and light

dispersion to achieve imaging speeds of up to 2.56×10^{14} fps. However, due to their reliance on active illumination, these systems are unable to capture self-luminescent scenes, such as dynamic scattering, photoluminescence intensity decay, and plasma emission. This limitation can be addressed through passive detection, where receive-only ultrafast detectors record emitted and/or scattered photons from the scene. To achieve this, various CUP systems [2, 14, 15] have utilized streak cameras to passively image dynamic scenes at frame rates ranging from 10^{10} to 10^{13} fps. However, streak cameras suffer from electronic jitter, low quantum efficiency, spectral limitations imposed by the photocathode, and space-charge effects, all of which degrade temporal resolution and sensitivity. Consequently, a recent research trend has focused on all-optical shearing units for SCTI. Among existing techniques, compressed optical-streaking ultrahigh-speed photography (COSUP) [16] employs a galvanometer scanner as a mechanical shearing unit, enabling all-optical image acquisition at speeds up to 1.5×10^6 fps. Overall, SCTI based on a single static encoder has demonstrated promising applications, including luminescence lifetime thermometry [17].

So far, in single-mask SCTI, the most commonly used CA is the binary mask with random Bernoulli distributed entries and a transmittance of nearly 50% [18, 19]. The main challenges of random binary CAs are the suboptimal performance in reconstruction algorithms, such as low convergence speed, because of the information loss and their stochastic nature, they induce artifacts in the recovered video. In contrast, there has been a rapid rise in the use of optimized binary patterns for SCTI. For example, some approaches exploit genetic algorithms [20], which is a heuristic method used to optimize multi-objective cost functions. In particular, the genetic algorithm optimizes the mask entries, although heuristic methods usually obtain local instead of global minimum. Other approaches design the binary masks by minimizing the Frobenius norm of the Gram matrix of the column-normalized random and unit matrix, which reduces the mutual coherence [21]. However, these computations involve enormous matrices that restrict methods to specific conditions, such as resolution and transmittance. A recent approach exploits deep learning to optimize a cost function that enhances the mask and the reconstruction for femto-photography [22]. However, data-driven reconstruction algorithms demand tremendous amounts of data and specific computational hardware. Furthermore, these efforts stick to the spatiotemporal multiplexing induced by this type of pattern in the measurement, leading to a compressive inference problem with an ill-posed inverse problem to reconstruct the frames, either using traditional iterative algorithms [23] or sophisticated deep-learning reconstruction techniques [24–27].

Recently, novel design patterns such as the sphere packing coded aperture (SPCA) have been introduced [28–30]. SPCA leverages sphere packing optimization to design the entries of the CA by considering the entries of the CA as centers of spheres in the 3D space to maximize the packing density [28–30]. SP optimization relaxes the inverse problem, accelerating the convergence of the reconstruction algorithm and reducing the number of reconstructed artifacts [28–30], as previously shown in the shuffling rolling shutter design that can remove wobble artifacts in the CMOS sensor [28, 29], as well as in the demosaicking of multispectral filter arrays [30]. In the case of spatiotemporal datacubes, the resulting non-multiplexed SPCA design transmittance will be inversely proportional to the number of frames. On the other hand, conventional data-driven retrieval algorithms for SCTI—such as deep-learning approaches [26]—require significant data to train a model on top of demanding enormous computational resources.

Novel approaches such as neural radiance field (D-NeRF) [31] and 4D Gaussian splatting (4D-GS) [32] might provide spatio-temporal advantages. However, D-NeRF requires images of objects under rigid and non-rigid motions from a single camera moving around the scene as inputs. Moreover, 4D-GS relies on background references and precise camera pose during large motions to optimize 4D Gaussians. Since in our particular context the COSUP system is static, it restricts the potential use of D-NeRF and 4D-GS approaches. Contrary to data-driven methods, novel implicit neural representations (INRs) [33] exploit neural networks that parametrize a

high-dimensional space, usually in a coordinate-based manner. The INR neural network structure is a multilayer perceptron (MLP) with periodic activation functions enabling the representation of complex natural signals, reducing the data for training to actual samples while minimizing the computational burden. Therefore, the INR can be conveniently combined with coordinate sampling approaches such as SPCA for the efficient reconstruction of under-sampled ultrahigh-speed space-time datacubes. By using INR, we not only reduce the need for training datasets beyond the measurements themselves, but we also avoid possible data hallucinations that may appear in data-driven approaches in severe under-sampled situations [34]. To overcome the challenges provided by traditional binary mask designs and data-driven dependencies in recovery algorithms, in this work, we propose a novel approach for SCTI that exploits non-multiplexed static binary CAs designed using SP, combined with efficient INRs as the reconstruction algorithm. To the best of our knowledge, this is the first time that such SPCA has been applied to SCTI. Specifically, the entries of the binary CA designed by solving a sphere packing problem promote a uniform subsampling of the transient event, exploiting the shearing of the galvanometer in the COSUP to complete the scan of the scene. Moreover, the proposed INR reconstruction approach leverages the subsampling of the SPCA—that obtains a non-multiplexed single measurement of the COSUP, which is the input of the INR—recovering the underlying transient events without requiring vast volumes of data and using considerably fewer hardware resources.

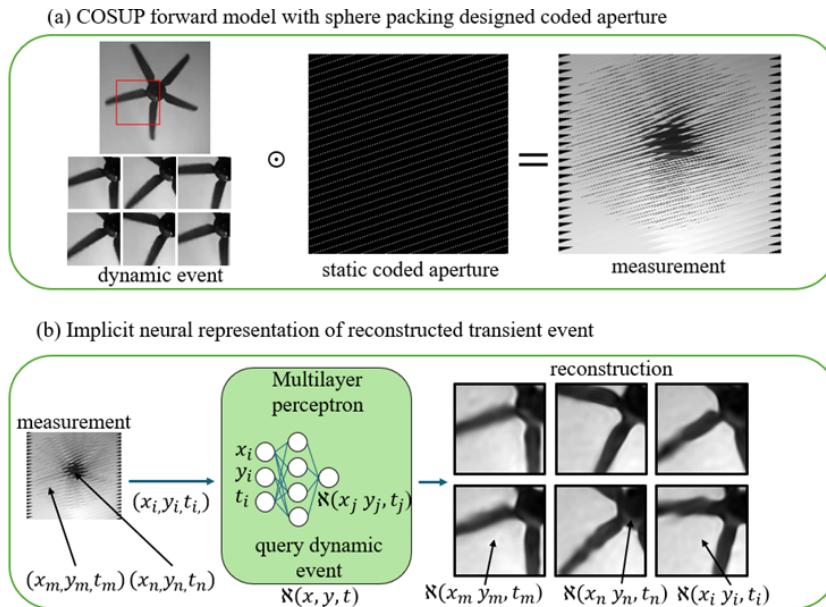


Fig. 1. Implicit neural representation (INR) for COSUP temporal imaging reconstruction. (a) The forward model comprises a non-multiplexed static binary CA that encodes the incoming temporal frames; the galvanometer shears these encoded frames, and the resulting measurement is captured in a CMOS sensor. (b) The INR is trained using only the COSUP single measurement coordinates (x, y, t) , and the output is the intensity value of the transient event at these coordinates, resulting in an agile reconstruction approach.

Figure 1 shows our sensing model using SPCA and our novel INR, which is trained using only a single INR measurement. Specifically, Fig. 1(a) depicts the COSUP forward model, where the static CA designed using SP promotes the uniform subsampling of the transient event. Then, the shifting produced by the galvanometer scanner of the COSUP completes the 3D space-time subsampling of coordinates (x, y, t) . Moreover, Fig. 1(b) shows the INR pipeline used to recover

the transient event from a single COSUP measurement. The 3D coordinates (x, y, t) of the single measurement are the training inputs of the INR. Then, the trained network infers all the missing values by querying all 3D entries of the transient event $\mathbf{N}(x, y, t)$. The key advantage of INR is the fast recovery of the transient events, i.e., since training using only a single measurement reduces the training and inference time and paves the way for agile applications without additional training data or time.

2. Forward model

The corresponding discrete model of COSUP is as follows:

$$\mathbf{Y} = \sum_{k=0}^{T-1} \varphi_k (\mathcal{X}_{\cdot,\cdot,k} \odot \mathbf{C}) + \boldsymbol{\Omega}. \quad (1)$$

Here $\mathcal{X} \in \mathbb{R}^{M \times N \times T}$ is the tensor that represents the 3D datacube, whose spatial resolution is $M \times N$ and the number of frames is T . $\mathbf{C} \in \{0, 1\}^{M \times N}$ denotes the tensor of the 2D-CA. The operator $\varphi(\cdot) : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}^{M \times (N+T-1)}$ introduces a right-horizontal circular shifting of k pixels to a right-zero-padded version of the compressed measurement (i.e., $[\mathbf{X}_{\cdot,\cdot,k}, \mathbf{0}]$ with $\mathbf{0} \in \mathbb{R}^{M \times (N+T-1)}$). $\mathbf{Y} \in \mathbb{R}^{M \times (N+T-1)}$ represents the compressed measurement. $\boldsymbol{\Omega} \in \mathbb{R}^{M \times (N+T-1)}$ is assumed to be additive Gaussian noise, although other types of noise, such as Poisson noise, might be considered.

2.1. Sphere Packing Coded Aperture Design

Fundamentally, the optimal SP density in the 3D Euclidean space is determined by the face-centered cubic (FCC) lattice, previously known as the Kepler conjecture [35], and is now called the Hales theorem [36, 37]. The FCC is denoted by the lattice $\Lambda_3 = [110; 101; 011]$, whose density is $\rho(\Lambda_3) = \frac{\pi}{\sqrt{18}} \approx 0.74$, which sets the upper bound in 3D for CAs sampling density by jointly considering the Hales theorem and the restriction imposed by CA's chessboard structure. SP considers non-overlapping spheres of an identical size in a cubic container. The volume of the cubic container depends on the CA's spatial resolution and the number of sub-sampled frames, assuming that the CA is square and that the number of spheres is equal to the total number of active entries of the CA.

The SP-based CA is designed for digital micromirror devices (DMDs)—one of the most popularly used encoders in SCTI. In this context, the spatial resolution of the DMD determines the size of the cubic container, and the number of pixels of DMD is equivalent to the number of packed spheres V . We consider DMDs with square pixels Fig. 2(a) and with diamond pixels [Fig. 2(b)]. The SP density is computed via $\rho(\mathbf{K}) = \frac{4V\pi(r(\mathbf{K}))^3}{3(\sqrt{V}+r(\mathbf{K}))^3}$, where $r(\cdot) : \mathbb{R}^{T \times T} \rightarrow \mathbb{R}$, $\mathbf{K} \in \mathbb{R}^{T \times T}$ is the either square or diamond DMD lattice, r is the radius of the sphere and V is the number of packed spheres, such that $V = T^2$.

The calculation of the optimal SP density using the full set of coordinates of the DMD is a nondeterministic polynomial time problem; to address this issue, we applied the divide and conquer strategy, which solves a relaxed version by computing the SP of a square kernel instead of all entries of the DMD. The size of the kernel is a function of the number of frames T that we aim to encode. Our designed kernel is given by

$$\mathbf{K} = ((\alpha \mathbf{W} + \beta \mathbf{H}) \mod T) + 1, \quad (2)$$

where $\mathbf{W} \in \mathbb{R}^{T \times T}$ and $\mathbf{H} \in \mathbb{R}^{T \times T}$ are vertical and horizontal translation matrices, and $\mathbf{1} \in \mathbb{R}^{T \times T}$ is an all-ones matrix. Note that the parameters $\alpha > 0$ and $\beta > 0$ are positive integers. In order to obtain the parameters optimal of α and β , we use the algorithm proposed by Diaz et al. [30]. First, we reorganize the kernel to obtain a set of 3D coordinates $p_i = (x, y, k_{(x,y)})$ $i \in \{0, \dots, V - 1\}$,

where $k_{(x,y)}$ stands for the kernel value at the x and y position, then we compute optimal radius $r(\mathbf{K})$ of the SP by maximizing the minimum distance between all the spheres, i.e. we compute all pairwise distances of the kernel coordinates

$$r(\mathbf{K}) = \max \left(\min_{1 \leq i < j \leq V-1} \|\mathbf{p}_i - \mathbf{p}_j\|_2^2 \right). \quad (3)$$

Note that $r(\mathbf{K})$ is a function of the number of spheres, and its position is in the 3D Euclidean space. Once the kernel is computed, it is translated along the DMD space:

$$\mathbf{E} = \mathbf{A} \otimes \mathbf{K}, \quad (4)$$

where $\mathbf{E} \in \mathbb{R}^{M \times N}$, $\mathbf{A} \in \mathbb{R}^{q \times s}$ with $q = \frac{M}{T}$ and $s = \frac{N}{T}$, and \otimes is the Kronecker product. The resulting optimal SPCA is:

$$C_{m,n} = \begin{cases} 1 & \text{if } 1 = E_{m,n} \\ 0 & \text{if } 1 \neq E_{m,n}, \end{cases} \quad (5)$$

where $m \in \{0, \dots, M-1\}$ and $n \in \{0, \dots, N-1\}$ index the spatial coordinates. For example, we set $T^2 = 256$ pixels for each case, which yields an SP density of 0.31 for the square-pixel DMDs. For the diamond-pixel DMDs, the density was calculated to be 0.31 by computing the ratio between the volume of 256 spheres and the volume container of the diamond DMD, which is equal to that obtained by the square pixel DMD due to its hexagonal packing property. Figures 2(a) and 2(b) depict the kernel \mathbf{K} when the number of frames is 16 for the square, and diamond DMD, respectively. To generate diamond DMD, we displace half of a unit to the left of the odd rows of the square kernel. The corresponding SP of these kernels are shown in Figs. 2(c) and 2(d). Since the entries of the diamond DMD are rotations of the square DMD the resulting density is the same because SP density is invariant to rotation, a comprehensive study of the SP density for square DMD can be found in [26, 28].

Moreover, Fig. 3(a) depicts the designed CA with a spatial resolution of 256×256 to capture 31 frames. The uniformity of our approach is demonstrated by summing the number of times that a CA scans the measurement, Fig. 3(b) shows that each pixel is sampled only once, which guarantees uniform sampling. In addition, Fig. 3(c) depicts a conventional static random coded aperture, and Fig. 3(d) shows each pixel is sampled a different number of times resulting in a non-uniform sampling approach. The proposed SPCA is not compressive as CACTI given the non-overlapping of the spatio-temporal sampling positions achieved when combining the coded aperture and the shearing induced by the galvanometer, leading to a mutual coherence of zero. Although this simplifies the reconstruction process, it also means we are giving up on the light transmission since it is inversely proportional to the number of frames $\frac{1}{T}$.

2.2. Reconstruction

We start by expanding the measurement \mathbf{Y} into the datacube by using the CA such that

$$\bar{\mathcal{X}}_{(:,:,k)} = \mathbf{C} \odot \mathbf{Y}_{:,(k:k+N)}, \quad (6)$$

with $k = \{0, \dots, T-1\}$. Then, we use an interpolation algorithm that recovers the underlying datacube via nearest neighbor interpolation (NNI) [38] or INR [33], whose input is the subsampled $\bar{\mathcal{X}} \in \mathbb{R}^{M \times N \times T}$ and whose output is the reconstructed datacube $\hat{\mathcal{X}} \in \mathbb{R}^{M \times N \times T}$. Furthermore, when we use INR to improve the reconstruction quality, these INR models are fully connected layers, and they are trained using only a single measurement [16].

INR refers to a method of representing continuous signals, such as images, audio, or 3D shapes, using a neural network. Instead of storing explicit data such as pixel values or point

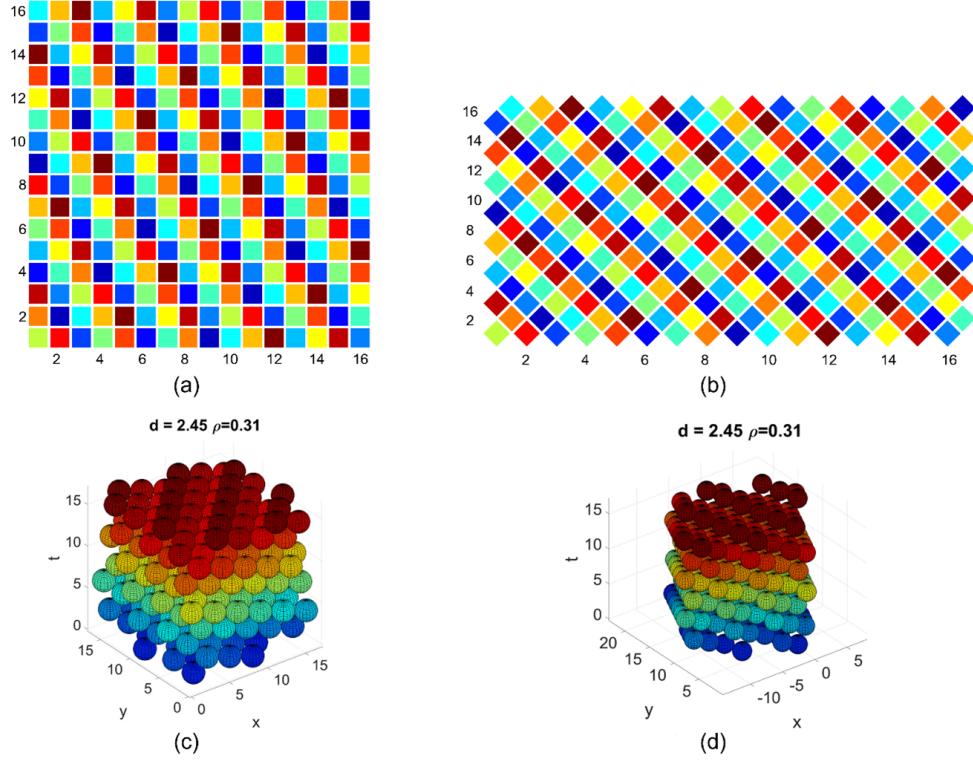


Fig. 2. Comparison of sphere-packing (SP) ability of two DMD structures. (a) Arrangement of a (a) square-pixel and (b) diamond-pixel DMD. SP of the (c) square-pixel and (d) diamond-pixel DMD.

clouds, an INR uses the weights of a neural network to encode the information. The network takes coordinates as inputs and outputs the corresponding signal value at that location. This approach allows for efficient and continuous representation of high-dimensional data and can provide benefits such as smooth interpolation and compact storage. The forward process of the neural networks for each intensity is composed of a multilayer perceptron (MLP) as follows:

$$p_n = \mathbf{N}(x, y, t), \quad (7)$$

where the inputs of the network are the normalized spatial coordinates $x = \frac{2*(i-1)}{(M-1)} - 1$, $y = \frac{2*(j-1)}{(N-1)} - 1$, and the temporal axis $t = \frac{2*(k-1)}{T-1} - 1$, and the outputs are the intensity values of the datacube $\hat{\mathbf{X}}_{(i,j,k)}$ in the given position. The loss function is the ℓ_1 norm,

$$Loss = \|X_{i,j,k} - \mathbf{N}(x, y, t)\|_1, \quad (8)$$

where $X_{i,j,k}$ is computed with real measurements at points (i, j, k) , $\mathbf{N}(x, y, t)$ is the MLP output that represents the intensity value of that position, $\|\cdot\|_1$ is the ℓ_1 norm.

3. Simulations

To prove the advantage of the proposed approach, we use a set of 10 videos with a spatial resolutions of 256×256 and 31 frames [28]. The randomly permuted CA and our SPCA have a light transmittance proportional to the number of frames, $\frac{1}{T}$. To simulate the discrete acquisition

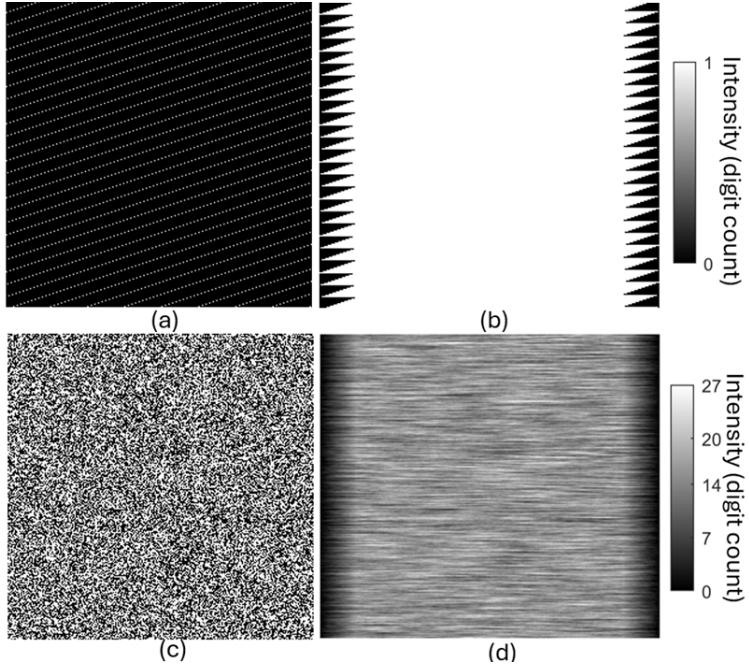


Fig. 3. Uniform sampling promoted by the SPCA design against conventional random approach. (a) Single SP binary mask. (b) Spatial distribution of the sampled pixels according to the mask sequence. (c) As (a), but for a random mask. (d) As (b), but for the random mask in (c).

process, we use Eq. (1). In Fig. 4, we compare the reconstruction quality of the rotating ceiling fan video, where a single measurement is shown in Fig. 4(a). The first, second, third and fourth rows of Fig. 4(b) depict the ground truth, and the frame reconstruction via the interpolation algorithm NNI, generalized alternating projection based total variation (GAP-TV) [23], and the implicit representation INR, of the 1st, 6th, 11th, 16th, 21th, 26th and 31st frames of a rotating ceiling fan video. The spatial fidelity is measured using the peak-signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [39]. The corresponding PSNR and SSIM of the rotating fan video are 26.22 dB and 0.94, respectively, by averaging all the frames. Note that although the INR's spatial fidelity is lower, the visual quality is superior. Moreover, the reconstruction time of GAP is significant, 79, 142, and 162 seconds for 16, 25, and 31 frames, respectively. Our simulation results are summarized in Table 1, and we compared the random permutation of frames against our SP approach. The results presented in Table 1 use a MLP with 4 hidden layers, each one with 256 neurons. The initial learning rate is 0.0005, and the model is trained on each input video for 1500 iterations. Table 2 introduces the computational time in training and testing for 16, 25 and 31 frames.

Figure 5 displays one of the advantages of using the proposed sphere-packing coded aperture and random approach with INR reconstruction, which is the flexibility of designing the system for an arbitrary number of reconstructed frames in Visualization 1. We perform simulations using the scenes of strawberries falling in milk with 32, 64, and 128 frames and, as a recovery algorithm, an INR of 1500 iterations. Fig. 5(a) depicts the simulated COSUP measurements and Fig. 5(b) shows the groundtruth and reconstructed frames. To facilitate the comparison of the spatial fidelity, we include the PSNR of each frame and note that the PSNR decreases when the number of frames increases.

Algorithm	CA	Number of frames		
		$T = 16$	$T = 25$	$T = 31$
NNI	Random CA PSNR (dB)	28.03	26.17	26.19
	Random CA SSIM	0.85	0.82	0.81
	SPCA PSNR (dB)	28.83	27.48	25.97
	SPCA SSIM	0.86	0.84	0.81
GAP	Random CA PSNR (dB)	28.61	26.68	26.58
	Random CA SSIM	0.87	0.82	0.81
	SPCA PSNR (dB)	29.11	27.81	26.28
	SPCA SSIM	0.87	0.84	0.81
INR	Random CA PSNR (dB)	26.44	25.30	25.63
	Random CA SSIM	0.85	0.79	0.80
	SPCA PSNR (dB)	26.74	26.25	25.57
	SPCA SSIM	0.85	0.82	0.78

Table 1. Performance of random permutation CAs against the proposed SPCA method on the 10 video dataset using the NNI interpolation algorithm and the implicit network INR, and GAP where $N = M = 256$ and $T = 16, 25$ and 31 frames.

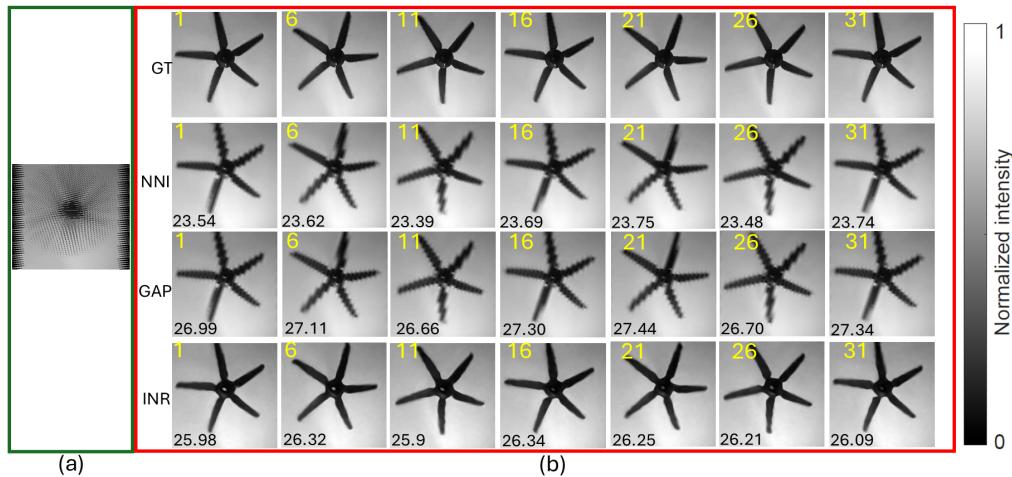


Fig. 4. Comparison of the simulation reconstructions. (a) Simulated single measurement using COSUP with SPCA. Groundtruth (top row) and reconstructions using NNI (second row), GAP (third row), and INR (bottom row) 1st, 6th, 11th, 16th, 21st, 26th and 31st frames of the rotating fan, respectively, where each black the number on the bottom is the PSNR in dB.

Stage / T	T=16 (simulation)	T=25 (simulation)	T=31 (simulation)	T=31 (readdata)
Training	6.2 s	6.1 s	6.1 s	1.7 s
Testing	1.00 ms	1.25 ms	1.30 ms	2.55 ms

Table 2. Computational time for training (seconds) and testing (milliseconds) in simulation and experimental evaluation using as a reconstruction algorithm INR.

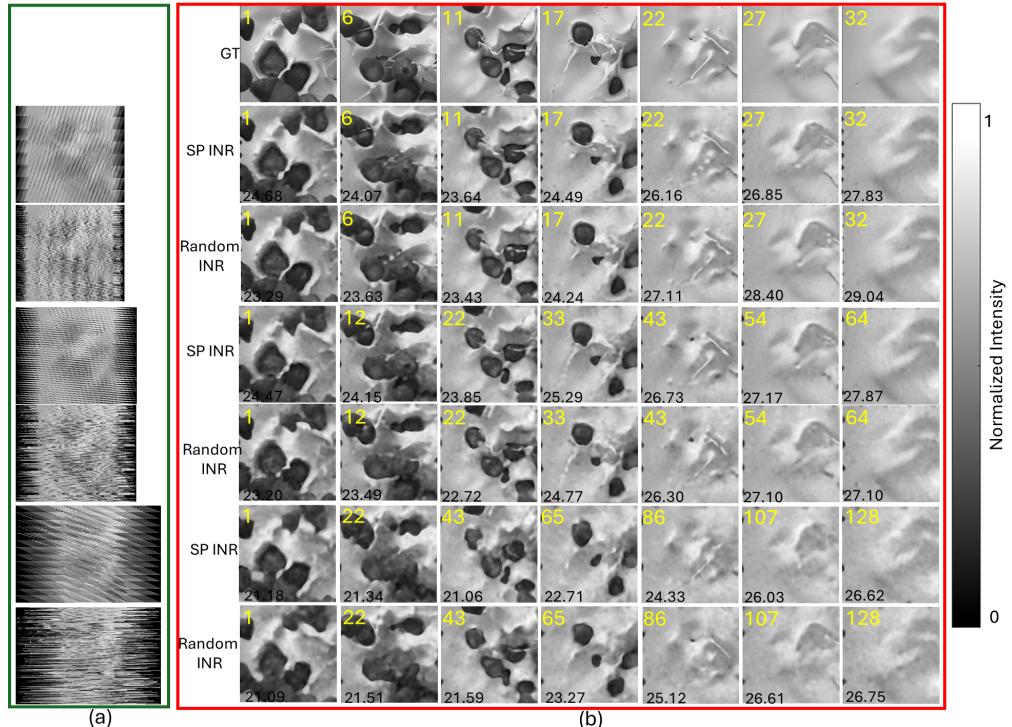


Fig. 5. Comparison of image reconstruction between sphere packing and randomly generated coded apertures using INR for more frames (see Visualization 1). (a) Single measurements for 32, 64, and 128 frames, on top INR SP, then INR random, and so forth; (b) The groundtruth (GT) of the –strawberries falling in the milk– scene and the reconstruction of seven frames using the SP and random coded and 1500 iterations of INR.

4. Experimental Results

4.1. Optical Implementation

To demonstrate the SP capability of acquiring and reconstructing transient events, we conducted experimental validations whose setup is schematically shown in Fig. 6. A system schematic of the COSUP is shown in Fig. 6(a). A green LED (CBT-90-G-L11-CK100, Luminus Devices) was used as the light source. The light is directed onto a digital mirror device (DMD, $10.8\mu m \times 10.8\mu m$ pixel size, AJD 4500, Ajile Light Industries) at an incident angle 24° to its surface normal. Ultrahigh-speed animations, multiplied with the optimized coded aperture designed by the SP approach (see Fig.3(a)), were loaded onto the DMD and displayed at 10 kHz. The spatially encoded scene is imaged by a 4f imaging system consisting of two achromatic lenses [Lens 1 (AC508-075-A,

Thorlabs) and Lens 2 (AC127-050-A, Thorlabs)] with a magnification ratio of 0.66. The intermediate image, located on the back focal plane of Lens 2, is relayed onto a CMOS camera ($5.5\mu\text{m} \times 5.5\mu\text{m}$ pixel size, GS3-U3-41C6C, Edmund Optics) by a second $4f$ system in which a galvanometer scanner (GS, QS30X-AG-ø30mm, Thorlabs) is placed at its Fourier plane for temporal shearing. The second $4f$ system is composed of two achromatic lenses [Lens 3 and Lens 4 (AC508-075-A, Thorlabs)] Finally, the spatially encoded and temporally sheared scene is spatiotemporally integrated by the CMOS camera to form a single measurement.

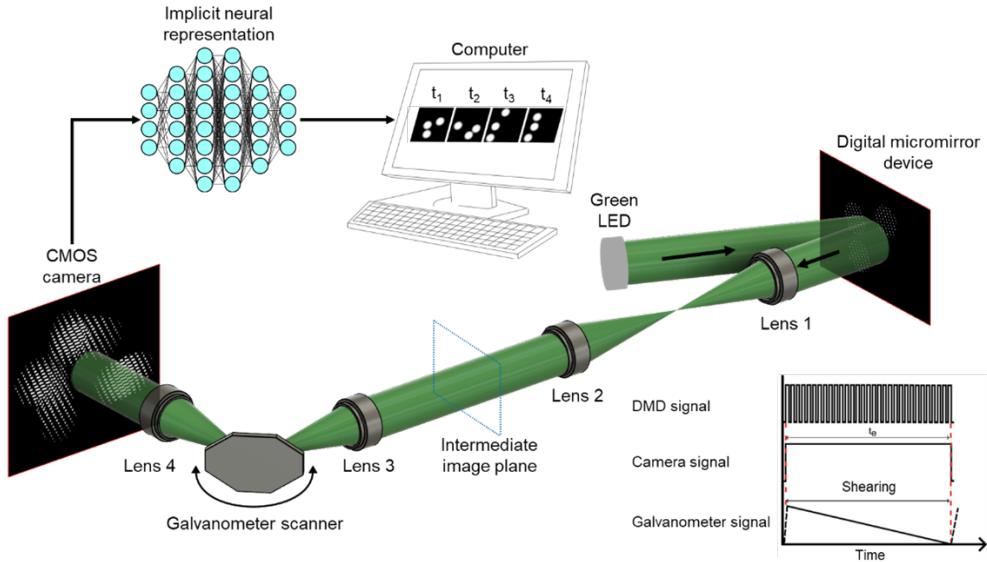


Fig. 6. Optical setup for experimental evaluation of CA design. Bottom-right inset: System schematic and timing diagram and acquisition sequence, where t_e is the exposure time of the camera.

To test the performance of our approach, we tested two dynamic scenes, namely propagating lightning rays and a rotating spinner, both with a frame size of 714×574 pixels and one sequence depth of 31 frames. The measurements are depicted in Fig. 6(b) and (e); the corresponding selected areas for the zoomed version of the two scenes are shown in Fig. 7(a) and (d). The first and second rows of Fig. 6(c) and (f) depict the reconstructions using NNI and INR of the lightning rays and the rotating spinner, respectively. The implicit representation used to reconstruct the underlying video was an MLP architecture with two hidden layers of 128 neurons, an initial learning rate of 0.0001, 800 iterations, and a batch size of 8000. The training time for each real measurement is 1.7 seconds, and the query time is 2.55 milliseconds. The computation time is reported in Table 2, running on an RTX 4070 Ti. The reason for the smaller number of layers in MLP of the experiment INR compared with the setup for the simulations is that the motions and textures of the real scenes are simpler than those used during the simulation. Although the interpolation algorithm recovers the motion in both scenes, the algorithm that relies on implicit representation outperforms the interpolation algorithm because its reconstruction is smoother.

Funding

ANID Fondecyt de Postdoctorado (3230489), Fondecyt de Exploracion (13240119), ANID Doctorado Nacional (2022- 21221399), ANID Anillos (ATE220022), Natural Sciences and Engineering Research Council of Canada (RGPIN-2024-05551).

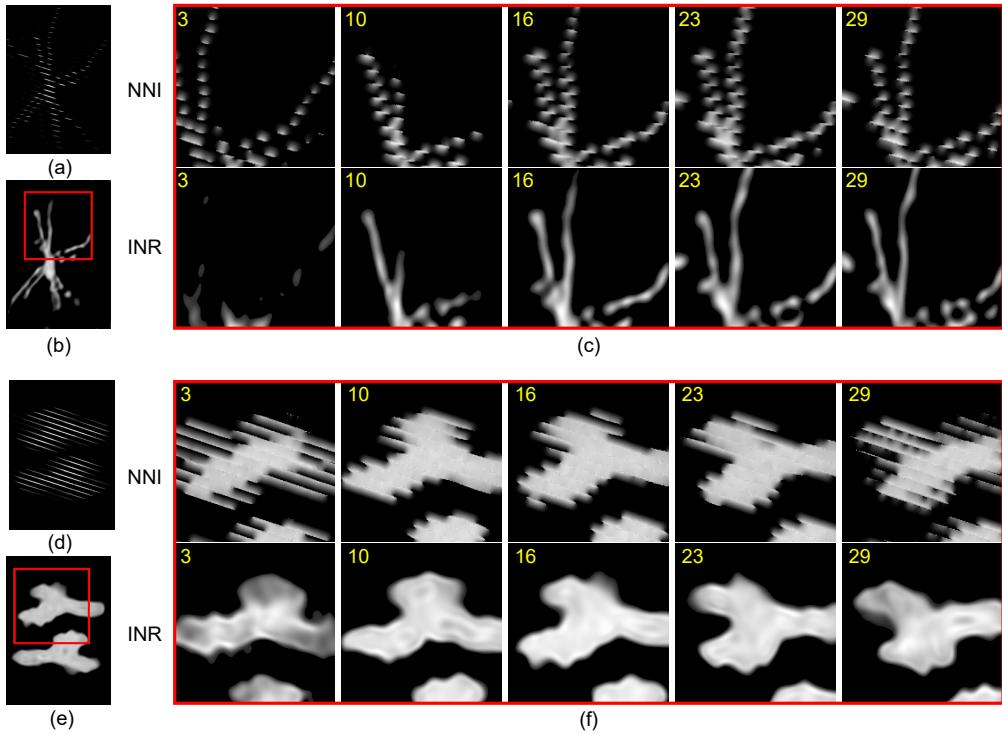


Fig. 7. Experimental results with two dynamic scenes. (a) and (d) measurements of a flash of lightning rays and a rotating spinner sampled using our optical COSUP setup; (b) and (e) red squares denoting the selected zoomed versions of the rays and spinner scenes depicted in (c) and (f), respectively. (c) and (f) denote the 3rd, 10th, 16th, 23rd, 29th frames of the reconstructed lightning rays and rotating spinner scenes using the interpolation and implicit representation algorithms.

Data availability.

Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Disclosures.

The authors declare that they have no conflicts of interest.

5. Conclusions

We have introduced a CA design approach that exploits optimal sphere packing for single-mask SCTI. Our approach can accurately recover a video from a single shot using the implicit neural representation as the reconstruction algorithm. The combination of the optimal sampling scheme with the proposed neural representation reduces the computational time for training and reconstructing the ultrahigh-speed videos. The proposed approach yields image reconstruction qualities of up to 28.70 dB for PSNR and 0.92 for SSIM. Furthermore, reconstructions using experimental data prove the feasibility of this approach for recovering the complex dynamics of ultrahigh-speed transient scenes, such as rotations and displacements, using both the conventional interpolation algorithm [38] and the state-of-the-art implicit representation [33], where the inference time is 2.55 ms, which paves the way for agile applications. In future work, we will extend this approach to improve the transmittance by designing SP-based CA multiplexing in

time [40], while also exploring the use of data-driven, physics-aware, deep-learning reconstruction algorithms.

References

1. L. Gao, J. Liang, C. Li, and L. V. Wang, “Single-shot compressed ultrafast photography at one hundred billion frames per second,” *Nature* **516**, 74–77 (2014).
2. N. H. Matis, S. Reed, S. S. Bulanov, V. Chvykov, G. Kalintchenko, T. Matsuoka, P. Rousseau, V. Yanovsky, A. Maksimchuk, S. Kalmikov *et al.*, “Snapshots of laser wakefields,” *Nat. Phys.* **2**, 749–753 (2006).
3. M. Y. Berezin and S. Achilefu, “Fluorescence lifetime measurements and biological imaging,” *Chem. reviews* **110**, 2641–2684 (2010).
4. B. J. Berne and R. Pecora, *Dynamic light scattering: with applications to chemistry, biology, and physics* (Courier Corporation, 2000).
5. M. Beresna, M. Gecevičius, and P. G. Kazansky, “Ultrafast laser direct writing and nanostructuring in transparent materials,” *Adv. Opt. Photon.* **6**, 293–339 (2014).
6. J. Liang, C. Ma, L. Zhu, Y. Chen, L. Gao, and L. V. Wang, “Single-shot real-time video recording of a photonic mach cone induced by a scattered light pulse,” *Sci. advances* **3**, e1601814 (2017).
7. J. Liang, *Coded Optical Imaging* (Springer Nature, 2024).
8. Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar, “Video from a single coded exposure photograph using a learned over-complete dictionary,” in *2011 International Conference on Computer Vision*, (IEEE, 2011), pp. 287–294.
9. P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, “Coded aperture compressive temporal imaging,” *Opt. express* **21**, 10526–10545 (2013).
10. Y. Lu, T. T. Wong, F. Chen, and L. Wang, “Compressed ultrafast spectral-temporal photography,” *Phys. review letters* **122**, 193904 (2019).
11. P. Wang, J. Liang, and L. V. Wang, “Single-shot ultrafast imaging attaining 70 trillion frames per second,” *Nat. communications* **11**, 2091 (2020).
12. J. Liu, M. Marquez, Y. Lai, H. Ibrahim, K. Légaré, P. Lassonde, X. Liu, M. Hehn, S. Mangin, G. Malinowski *et al.*, “Swept coded aperture real-time femtophotography,” *Nat. Commun.* **15**, 1589 (2024).
13. H. Tang, T. Men, X. Liu, Y. Hu, J. Su, Y. Zuo, P. Li, J. Liang, M. C. Downer, and Z. Li, “Single-shot compressed optical field topography,” *Light. Sci. & Appl.* **11**, 244 (2022).
14. Y. Lai, M. Marquez, and J. Liang, “Tutorial on compressed ultrafast photography,” *J. Biomed. Opt.* **29**, S11524–S11524 (2024).
15. Y. Lai, Y. Xue, C.-Y. Côté, X. Liu, A. Laramée, N. Jaouen, F. Légaré, L. Tian, and J. Liang, “Single-shot ultraviolet compressed ultrafast photography,” *Laser & Photonics Rev.* **14**, 2000122 (2020).
16. X. Liu, J. Liu, C. Jiang, F. Vetrone, and J. Liang, “Single-shot compressed optical-streaking ultra-high-speed photography,” *Opt. Lett.* **44**, 1387–1390 (2019).
17. X. Liu, A. Skripka, Y. Lai, C. Jiang, J. Liu, F. Vetrone, and J. Liang, “Fast wide-field upconversion luminescence lifetime thermometry enabled by single-shot compressed ultrahigh-speed imaging,” *Nat. communications* **12**, 6401 (2021).
18. D. J. Brady, *Optical imaging and spectroscopy* (John Wiley & Sons, 2009).
19. X. Yuan, D. J. Brady, and A. K. Katsaggelos, “Snapshot compressive imaging: Theory, algorithms, and applications,” *IEEE Signal Process. Mag.* **38**, 65–88 (2021).
20. C. Yang, D. Qi, X. Wang, F. Cao, Y. He, W. Wen, T. Jia, J. Tian, Z. Sun, L. Gao *et al.*, “Optimizing codes for compressed ultrafast photography by the genetic algorithm,” *Optica* **5**, 147–151 (2018).
21. X. Liu, S. Zhang, A. Yurtsever, and J. Liang, “Single-shot real-time sub-nanosecond electron imaging aided by compressed sensing: Analytical modeling and simulation,” *Micron* **117**, 47–54 (2019).
22. M. Marquez, G. Balistreri, R. Morandotti, L. Razzari, and J. Liang, “Metalens-based compressed ultracompact femtophotography: analytical modeling and simulations,” *Ultrafast Sci.* **4**, 0052 (2024).
23. X. Yuan, “Generalized alternating projection based total variation minimization for compressive sensing,” in *2016 IEEE International conference on image processing (ICIP)*, (IEEE, 2016), pp. 2539–2543.
24. Z. Cheng, B. Chen, G. Liu, H. Zhang, R. Lu, Z. Wang, and X. Yuan, “Memory-efficient network for large-scale video compressive sensing,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2021), pp. 16246–16255.
25. M. Qiao, Z. Meng, J. Ma, and X. Yuan, “Deep learning for video compressive sensing,” *Apl Photonics* **5** (2020).
26. M. Marquez, Y. Lai, X. Liu, C. Jiang, S. Zhang, H. Arguello, and J. Liang, “Deep-learning supervised snapshot compressive imaging enabled by an end-to-end adaptive neural network,” *IEEE J. Sel. Top. Signal Process.* **16**, 688–699 (2022).
27. X. Liu, J. Monteiro, I. Albuquerque, Y. Lai, C. Jiang, S. Zhang, T. H. Falk, and J. Liang, “Single-shot real-time compressed ultrahigh-speed imaging enabled by a snapshot-to-video autoencoder,” *Photonics Res.* **9**, 2464–2474 (2021).
28. E. Vera, F. Guzmán, and N. Díaz, “Shuffled rolling shutter for snapshot temporal imaging,” *Opt. Express* **30**, 887–901 (2022).

29. E. Vera, F. Guzman, and N. Diaz, *Shuffled Rolling Shutter Camera* (Springer International Publishing, Cham, 2024), pp. 499–513.
30. N. Diaz, A. Alvarado, P. Meza, F. Guzmán, and E. Vera, “Multispectral filter array design by optimal sphere packing,” *IEEE Transactions on Image Process.* **32**, 3634–3649 (2023).
31. A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, “D-nerf: Neural radiance fields for dynamic scenes,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), pp. 10318–10327.
32. G. Wu, T. Yi, J. Fang, L. Xie, X. Zhang, W. Wei, W. Liu, Q. Tian, and X. Wang, “4d gaussian splatting for real-time dynamic scene rendering,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2024), pp. 20310–20320.
33. V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, “Implicit neural representations with periodic activation functions,” in *Advances in Neural Information Processing Systems*, vol. 33 H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, eds. (Curran Associates, Inc., 2020), pp. 7462–7473.
34. D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2018), pp. 9446–9454.
35. J. Kepler, *The six-cornered snowflake* (Paul Dry Books, 2010).
36. T. C. Hales, “A proof of the kepler conjecture,” *Annals mathematics* **162**, 1065–1185 (2005).
37. T. Hales, M. Adams, G. Bauer, T. D. Dang, J. Harrison, H. Le Truong, C. Kaliszyk, V. Magron, S. McLaughlin, T. T. Nguyen, and et al., “A formal proof of the kepler conjecture,” *Forum Math. Pi* **5**, e2 (2017).
38. I. Amidror, “Scattered data interpolation methods for electronic imaging systems: a survey,” *J. Electron. Imaging* **11**, 157–176 (2002).
39. Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Process.* **13**, 600–612 (2004).
40. A. Alvarado, N. Díaz, P. Meza, and E. Vera, “A sphere packing approach to design multiplexed multispectral filter arrays,” in *Optica Imaging Congress (3D, COSI, DH, FLatOptics, IS, pcaOP)*, (Optica Publishing Group, 2023), p. CTh2A.6.