Supplementary Information for

Spatiotemporal Topological Combs for Robust High-Dimensional Information Transmission

Authors:

Dawei Liu[1†], Daijun Luo[1†], Huiming Wang[3†], Xingyuan Zhang[2], Zhirong Tao[4,] Dana JiaShaner[1], Zhensheng Tao[5], Qian Cao[1], Xiaoshi Zhang[6], Guangyu Fan[1*], Qiwen Zhan[1,2,7]

Affiliations:

[1]School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, 200093, Shanghai, China.

[2]Zhejiang Key Laboratory of 3D Micro/Nano Fabrication and Characterization, Department of Electronic and Information Engineering, School of Engineering, Westlake University, Hangzhou, Zhejiang 310030, China.

[3]Institute of Photonics, TU Wien, Gusshausstrasse 27/387, Vienna, Austria.

[4]Karlsruhe Institute of Technology, Karlsruhe, Germany.

[5]State Key Laboratory of Surface Physics, Key Laboratory of Micro and Nano Photonic Structures (MOE), Department of Physics, Fudan University, Shanghai 200433, China

[6]Yunnan University, Kunming, Yunnan, 650500, China.

[7]International Institute for Sustainability with Knotted Chiral Meta Matter (WPI-SKCM2), Hiroshima University, Higashihiroshima, Hiroshima, 739-8526, Japan.

†These authors contribute equally to this work.

*E-mail: gfan@usst.edu.cn

Section 1: All-degree-of-freedom spatiotemporal topological modulation

Following Section S1, the vortex pulse train is generated by modulating the source field on a spatial light modulator (SLM) in the $k_x - \omega$ plane. For a general topological pulse sequence, the all-degree-of-freedom (all-DoF) modulation applied to the SLM is:

$$M(x,y) = \sum_n A_n \, exp(-i\ell_n\theta) \, exp(ik\Delta t_n x) \, exp(i\Delta\varphi_n)$$

where (x, y) are the horizontal and vertical coordinates in the $k_x - \omega$ plane; n indexes each sub-pulse; $\theta = \arctan(y/x)$ is the azimuthal angle; k is the carrier wavenumber; and A_n, ℓ_n, Δt_n, and $\Delta\varphi_n$ are, respectively, the amplitude weight, topological charge, temporal delay, and initial phase offset of the n-th sub-pulse.

By grouping several pixels into a superpixel and applying specific phase shifts to each subpixel, a synthetic light field with modulated amplitude and phase can be generated in the far field through diffraction.

The resulting $M(x, y)$ is a complex-amplitude function. To implement it on a phase-only SLM, we adopt the macropixel (double-phase) method: every 2×2 pixel block is treated as a single macropixel. Let the normalized complex modulation be:

$$M = |M|\, exp(\, i\phi)$$

It can be recast as:

$$\begin{aligned} M &= |M|\, exp(\, i\phi) \\ &= cos(\alpha)\, exp(i\phi) \\ &= \left(\frac{exp(i\alpha) + exp(-i\alpha)}{2}\right) exp(i\phi) \\ &= \frac{1}{2}(exp(i\phi + i\alpha) + exp(i\phi - i\alpha)) \\ &= \frac{1}{2} exp\big(i(\phi + \alpha)\big) + \frac{1}{2} exp\big(i(\phi - \alpha)\big) \end{aligned}$$

so that two phase patterns, $\phi \pm \alpha$, encode the desired complex value. Assuming a uniform incident field (no significant input amplitude variation), the phase mask loaded on the SLM is arranged as shown in Fig. 1, the two diagonal pixels of each 2×2 macropixel carry $\phi \pm \alpha$, while the other two carry $\phi - \alpha$. This enables full complex-amplitude modulation using a phase-only device.

Section 2: Star-guided measurement method

The core idea is to fix the first sub-pulse of the ST-Comb as a Gaussian reference and temporally separate it from the following information-bearing pulses. This creates, in the frequency domain, a spectrum analogous to that of a conventional reference arm

interferometer—without requiring an extra optical path. We first generate a reference ST-Comb consisting of two Gaussian sub-pulses to obtain a flat-phase reference spectrum. Then we modulate the actual ST-Comb so that only the first sub-pulse remains Gaussian, while all subsequent sub-pulses carry the encoded information. Figure 2a, b shows the time-domain waveforms and spectra of the reference and test ST-Combs, respectively. The modulated Gaussian reference pulse (first sub-pulse) and the delayed signal pulses enter the spectrometer together and interfere spectrally, producing fringes. The measured spectral intensity is:

$$S(\omega) = S_{\mathrm{ref}}(\omega) + S_{\mathrm{sig}}(\omega) + 2\sqrt{S_{\mathrm{ref}}S_{\mathrm{sig}}}\,cos\left[\phi_{\mathrm{sig}}(\omega) - \phi_{\mathrm{ref}}(\omega) + \omega\tau\right]$$

where τ is the temporal delay between the reference and signal pulses. The cosine term contains both their phase difference and the $\omega\tau$ factor; thus, the average fringe spacing is inversely proportional to τ. Because the first sub-pulse's spectrum is approximately Gaussian with a smooth phase, it serves as a known reference field. The recorded spectral interferogram therefore encodes both the amplitude and phase of each signal pulse.

Applying a Fourier transform to the measured spectrum separates, in the delay domain, the cross-correlation peaks associated with each reference–signal pair (Fig. 3c). Each peak corresponds to the delay of a particular signal sub-pulse relative to the reference, and its complex amplitude carries that pulse's relative phase and intensity. Selecting the peak associated with the desired reference–signal interference term and inverse-transforming it yields the complex spectral envelope (amplitude and phase) of the signal pulse. Dividing this complex spectrum by the known reference spectrum and inverse Fourier transforming gives the complete time-domain electric field of the signal pulse, enabling full reconstruction of its temporal phase and waveform (Fig. 2d).

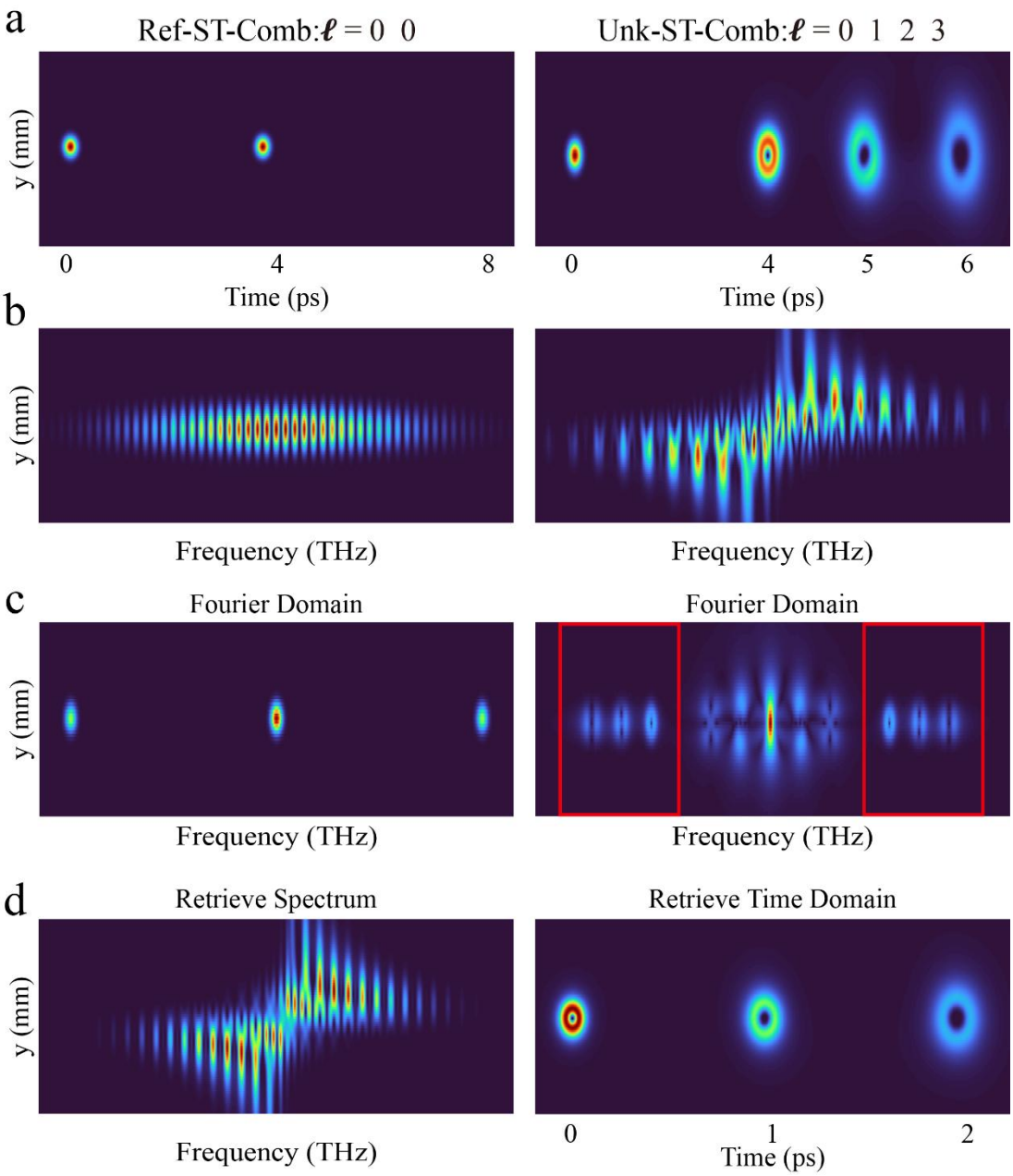

Figure 2: Star-guided measurement workflow. a) The reference ST-Comb comprises two Gaussian sub-pulses separated by a delay ≫ 1 ps; the test ST-Comb is obtained by replacing the second sub-pulse with the information-bearing sequence. b) Spectra of the reference and test pulses. c) Fourier-domain (delay) signal showing cross-correlation terms that encode the unknown information. d) Reconstructed ST-Comb in the frequency and time domains.

Compared with a traditional Gaussian-reference scheme, which typically requires an external reference beam and a stable two-arm interferometer, the present collinear

design is far more robust. Here, the reference and signal pulses share the same optical path, so their relative phase $\Delta\phi(\omega)$ is determined primarily by the known delay and is largely immune to environmental perturbations. Common-path disturbances affect both pulses equally, keeping the relative phase essentially constant and greatly suppressing the phase jitter and noise that plague conventional dual-arm interferometers.

Section 3: Topological invariants

The winding number W is a topological invariant of the optical field that quantifies the helical winding of its phase. For fields such as spatiotemporal optical vortices (STOV), the phase exhibits a helical structure and the intensity typically forms a ring with a central null. The winding number is defined directly through a contour integral:

$$W = \frac{1}{2\pi} \oint_{\Gamma} \nabla\phi \cdot ds$$

where ϕ is the optical phase and Γ is a closed contour encircling the singularity. An optical vortex generally exhibits a phase factor $\exp(i\ell\varphi)$, with φ the azimuthal angle; each photon then carries orbital angular momentum (OAM) $\ell\hbar$, in one-to-one correspondence with W. Because it depends only on the global phase winding, W is a global property[1].

As such, W is conserved during propagation in free space or linear media: diffraction, refraction, or linear interference cannot change the total charge unless new singularities are created or existing ones annihilated[2]. Thus, although the field may broaden or split, its net phase winding remains unchanged.

STOV couples spatial and temporal dimensions. Its field can be written as:

$$E(x, y, t, z) = A(x, y, z, t)\, exp[i\, \ell\varphi st]\, exp[ik(z - vt)]$$

where A is the envelope, k the wavenumber, v the group velocity, and φ_{st} is a spatiotemporal azimuth (in the $x-t$ plane one may define $\varphi_{st} = \tan^{-1}(x/\tau)$, with τ a temporal scale). The intensity forms a closed ring with a near-zero center, and the phase winds helically around it. Unlike conventional vortices whose OAM is parallel to the propagation axis, STOVs carry transverse OAM, the energy flow circulates within a space–time plane, giving photons an angular-momentum component perpendicular to the propagation direction.

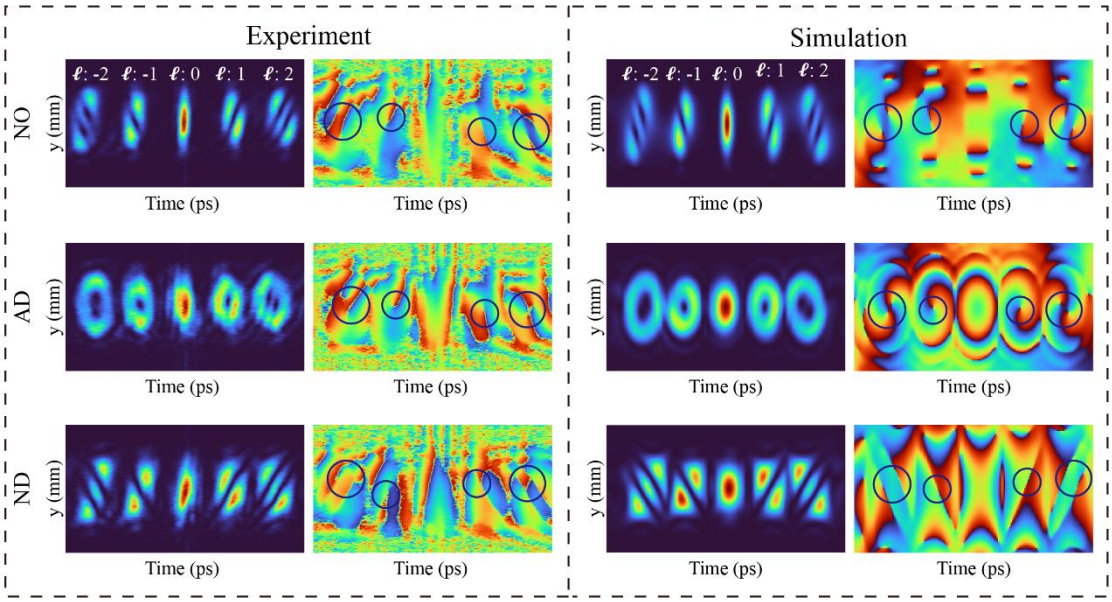

Figure 3: Evolution of the ST-Comb under different dispersion conditions. During normal dispersion propagation, group-velocity dispersion induces intrinsic temporal splitting of the pulse train. Compensation with negative dispersion counteracts this effect, preserving the original temporal envelope, whereas positive dispersion compensation enhances the temporal walk-off and accelerates fragmentation. In all cases, the underlying spatiotemporal topological phase remains invariant, indicating its immunity to dispersion-induced temporal reshaping.

During propagation, high-order STOV ($|\ell| > 1$) in free space often fracture into multiple unit-charge vortices whose charges sum to the original ℓ. This multi-lobe structure arises from imbalance between temporal dispersion and spatial diffraction. As illustrated in Fig. 3, an ST-Comb composed of STOV sub-pulses can split into petal-like pulses. Applying negative dispersion (via the SLM) compensates temporal

broadening and restores a near-circular profile; positive dispersion enhances the fission. Regardless of such shape changes, the total phase winding remains constant, underscoring the protected nature of W.

Section 4: Spatiotemporal optical vortices and the ST-Comb in time and frequency domain

STOV is a vortex structure in the spatiotemporal plane that exhibits circulating phase and energy, with a winding number conserved during propagation. As illustrated in Fig. 4a, the ST-Comb in the time domain can be viewed as the convolution of a Gaussian pulse train (with picosecond spacing) and an STOV waveform. Let a single Gaussian pulse have the temporal envelope:

$$g(t) = A\,exp\left[-\frac{t^2}{2\tau^2}\right]$$

and let the periodic pulse train be:

$$G(t) = \sum_n g(t - nT)$$

where T is the pulse spacing on the order of picoseconds. Denoting the STOV waveform by $V_{\text{STOV}}(t, r)$, which carries a helical phase factor $e^{i\ell\theta}$ with topological charge ℓ, the ST-Comb field is the convolution of these two components:

$$E_{\text{ST-Comb}}(t, r) = [G * V_{\text{STOV}}](t, r) = \int G(\tau)\; V_{\text{STOV}}(t - \tau, r)\, d\tau$$

This convolution model shows that each original pulse is endowed with a STOV state, so every sub-pulse in the train carries a spatiotemporal vortex phase.

Optical frequency combs are renowned for their exceptional coherence and frequency accuracy, a temporal pulse train corresponds in the frequency domain to an array of evenly spaced spectral lines (a comb), often used as a precise frequency ruler. For the Gaussian train above, the spectrum is a set of discrete lines separated by $\omega_R = 2\pi/T$, with linewidth set by the Fourier width of a single pulse. The ST-Comb spectrum is the product of this comb and the STOV spectral envelope (Fig. 2b):

$$\tilde{E}_{ST-Comb}(\omega, k) = \tilde{G}(\omega)\tilde{V}_{STOV}(\omega, k)$$

where $\tilde{G}(\omega)$ is the comb-like spectrum of the Gaussian train, and $\tilde{V}_{STOV}(\omega, k)$ is the vortex envelope (broadband in amplitude and carrying a helical phase $e^{i\ell\theta}$). The product form implies that the amplitude of each comb tooth is modulated by the vortex envelope, while its phase acquires a helical dependence—yielding a richly interleaved structure. Fine comb features (set by the long temporal period T) ride on a broader spectral envelope determined by the short STOV sub-pulses. Interference between different topological phases can be encoded spectrally; for example, pulse-charge sequences [3 2 1] and [2 1 3] exhibit distinguishable fine spectral fingerprints.

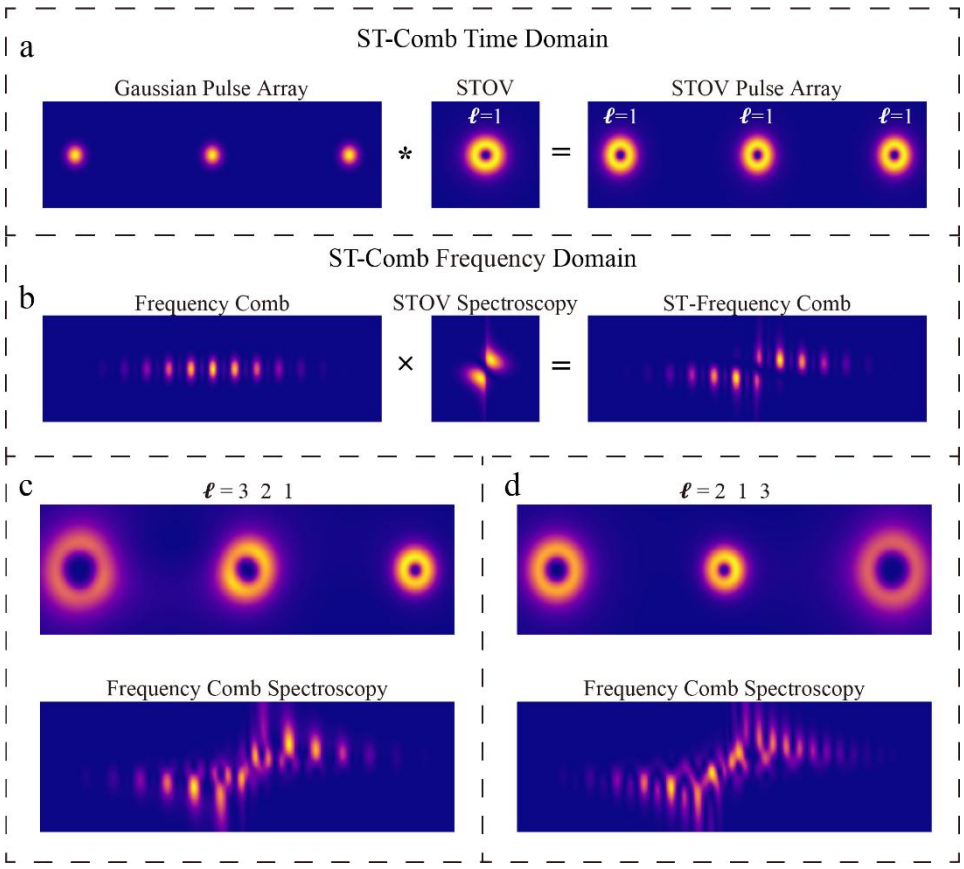

Figure 4: Formation of the ST-Comb in the time and frequency domains. a) In the time domain, the ST-Comb is the convolution of a picosecond-spaced Gaussian pulse train with a STOV waveform; b) In the frequency domain, it appears as a Gaussian optical frequency comb whose teeth are modulated by a spatiotemporal vortex envelope, yielding topological comb lines; c) Comb structures for topological-charge sequences $\ell = [3\ 2\ 1]$ and $\ell = [2\ 2\ 3]$, exhibiting distinct interwoven fine features.

Section 5: Linear and nonlinear ℓ–t modulation of the ST-Comb

To systematically analyze the ST-Comb in both time and topological domains, we introduce an $\ell - t$ map as a pseudo–spatiotemporal representation. Using time t and topological charge ℓ as axes, each sub-pulse is plotted at its temporal position with a vertical coordinate equal to its charge, directly visualizing how the topology evolves across the pulse train (analogous to a time–frequency Wigner distribution). In this framework, the field takes the form of Eq. (1), where the peak amplitude of the nnn-th sub-pulse is C_n and its normalized spatial–temporal envelope is $A_n(x, y, \mu_n)$, with μ_n denoting the topological DoF. The brightness of a point in the $\ell - t$ plane encodes $|C_n|$, while its vertical coordinate gives ℓ. By tuning the spatiotemporal topology, we can track how spectral lines shift linearly in the $\ell - t$ plane and thereby assess their controllable evolution.

As shown in Fig. 5, sweeping all sub-pulse charges from [1,2,3,4,5] to [−5,−4,−3,−2,−1] produces a mirror-symmetric transformation of the ST-Comb spectrum about the central frequency. This symmetry arises from phase conjugation when the sign of ℓ\ellℓ is inverted: the helical phase factors $e^{i\ell\varphi}$ and $e^{-i\ell\varphi}$ are complex conjugates, and their 2D Fourier images are axis-symmetric in the frequency domain. Throughout this operation the pulse spacing τ and amplitudes remain unchanged; only the topology flips, leaving the $\ell - t$ grid equally spaced but reversed in direction. This behavior underscores the utility of the $\ell - t$ representation for revealing the ST-Comb's high-dimensional spectral structure.

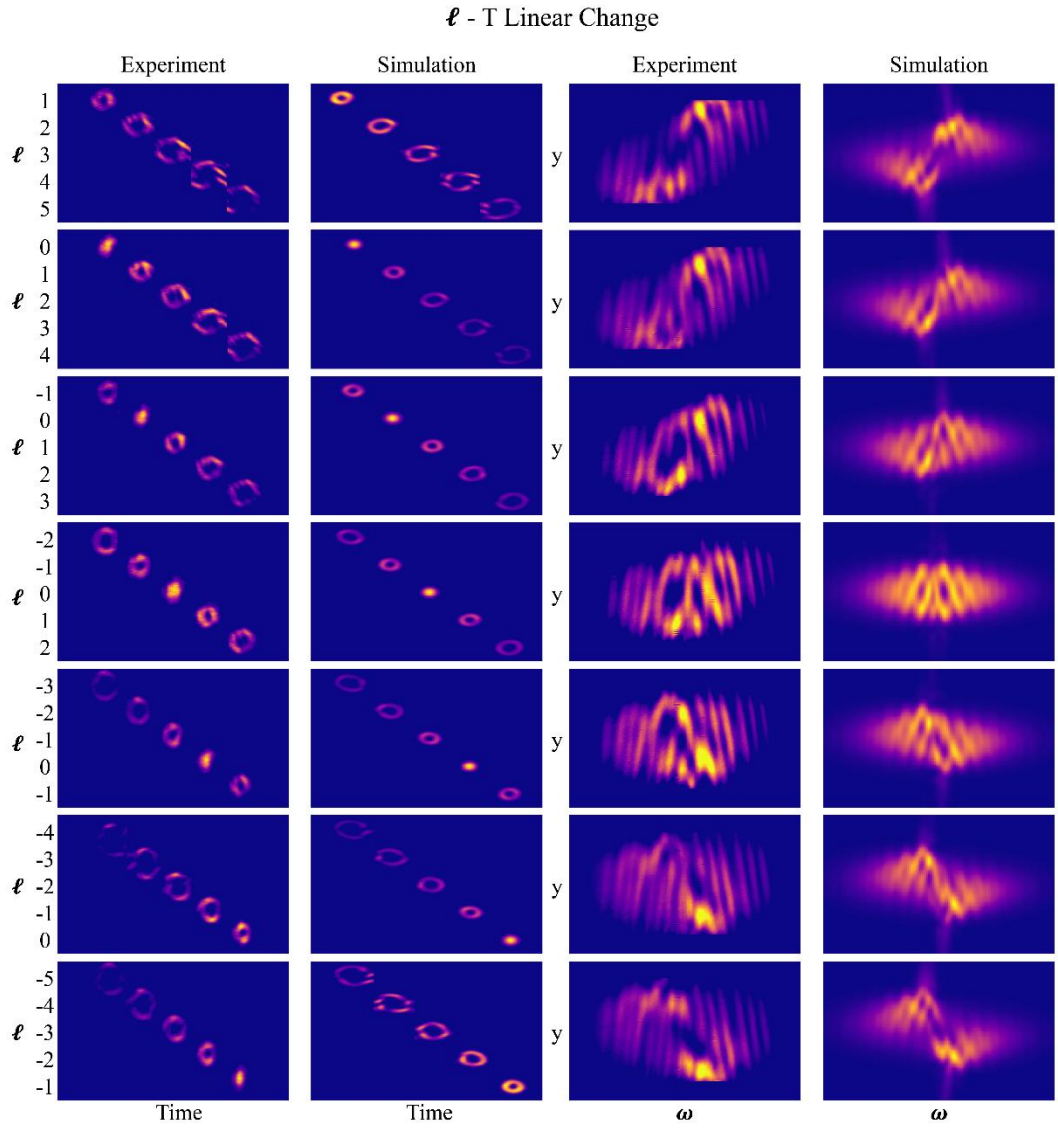

Figure 5: Control of the linear ℓ–t relation in the ST-Comb. Temporal and spectral evolution as the topological-charge sequence is swept linearly from [1,2,3,4,5] to [−5,−4,−3,−2,−1].

Beyond linear topology–time sweeps, we demonstrate arbitrary modulation patterns (Fig. 6), including plateau-like, biphasic counter-trending, and apex-rise trajectories. Plateau-like modulation keeps the topology constant ($\ell = \ell_0$) over a time segment, then abruptly switches to ℓ_1 at $t = T$, yielding a step in the $\ell - t$ plot. Mathematically, $\mu_n = \ell_0$ for $t < T$ and $\mu_n = \ell_1$ for T$t \geq T$. Holding the phase map fixed and toggling the vortex phase at T realizes this scheme. The plateau provides a stable topological state for redundant encoding and enhanced noise immunity, while the jump

acts as a temporal marker aiding boundary detection and frame synchronization. Different plateau lengths and ℓ_0/ℓ_1 pairs can carry multiple bits, boosting information density.

Biphasic counter-trending modulation consists of two consecutive segments with opposite trends, ℓ first increases (linearly from ℓ_a to ℓ_b), then decreases (ℓ_b to ℓ_c), producing a triangular or trapezoidal trace. The pronounced trend reversal is easy for neural networks to recognize and serves as a robust fingerprint for high-speed decoding. Because the net change can be near zero, this pattern offers inherent error self-checking and common-mode noise suppression. Information can be embedded independently in the two slopes, the peak ℓ_b, and the segment durations.

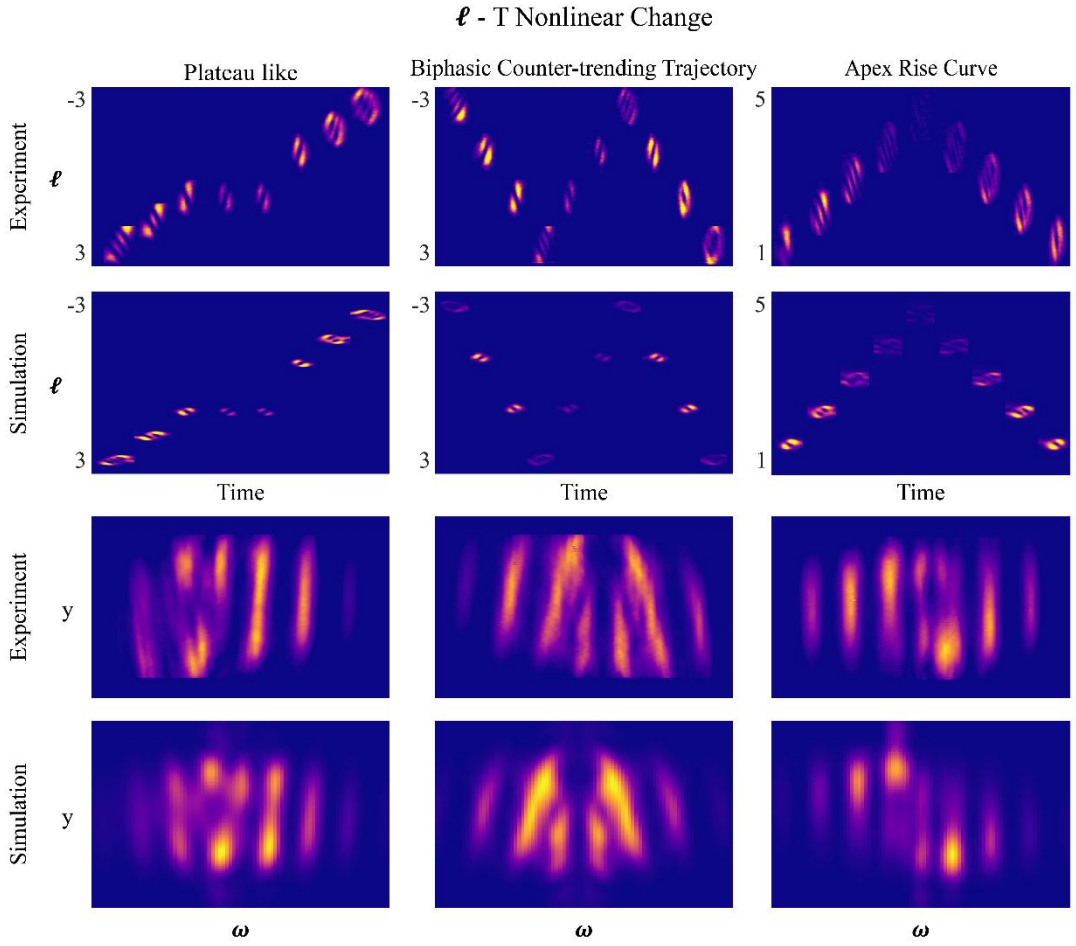

Figure 6: Nonlinear ℓ–t modulation of the ST-Comb. Examples of plateau-like, biphasic counter-trending, and apex-rise modulation schemes, illustrating their respective topology–time trajectories and corresponding ST-Comb characteristics.

Apex-rise modulation smoothly ramps ℓ up to a peak and then back down, forming a bell- or parabola-like curve. Unlike the piecewise-linear biphasic case, this is a continuous second-order function with an extremum at $t = T_{\text{peak}}$ and zero derivative there. The smooth evolution avoids phase discontinuities that could distort the temporal waveform, preserving shaping fidelity. The brief residence at ℓ_{max} reduces the need to sustain high-order topology, mitigating loss and diffraction in practice. Owing to its chirp-like evolution in topology, the pattern can also be identified in a mixed frequency–topology domain, opening an additional multiplexing dimension.

Section 6: Radial ST-Comb Modulation

In Section 2, we presented the fundamental complex amplitude modulation function for topological pulse sequences. Building on this, more complex spatiotemporal information can be encoded into the topological pulse sequence. Using radial modes and superposition states as an example, we demonstrate additional possibilities for the pulse comb design proposed in this work. For radial modes, when each position in the topological pulse sequence is assigned a radial mode, the formula from Section 2 changes to:

$$M(x,y) = \sum_n A_n \, exp(-i\ell_n\theta) \, exp(ik\Delta t_n x) \, exp(i\Delta\varphi_n) L_{p_n}^{|\ell_n|}\left(\frac{2r^2}{w^2}\right)$$

Where $L_{p_n}^{|l_n|}(2r2w2)$ is the generalized Laguerre polynomial, and P is the radial mode number. The specific form of the Laguerre polynomial is:

$$L_p^{|\ell|}(x) = \sum_{k=0}^{p} \frac{(-1)^k}{k!} \binom{p+|\ell|}{p-k} x^k \ell$$

This encoding strategy allows each sub-pulse to carry, in addition to the angular momentum degree of freedom ℓ, an independent radial mode P, enabling radial-azimuthal coupling in spatiotemporal multiplexing. The spatiotemporal pulses constructed in this way not only expand the control parameter space of each wavepacket

but also introduce complex multi-layered interference structures in the frequency domain, formed by the superposition of radial and azimuthal modes, resulting in more intricate comb teeth.

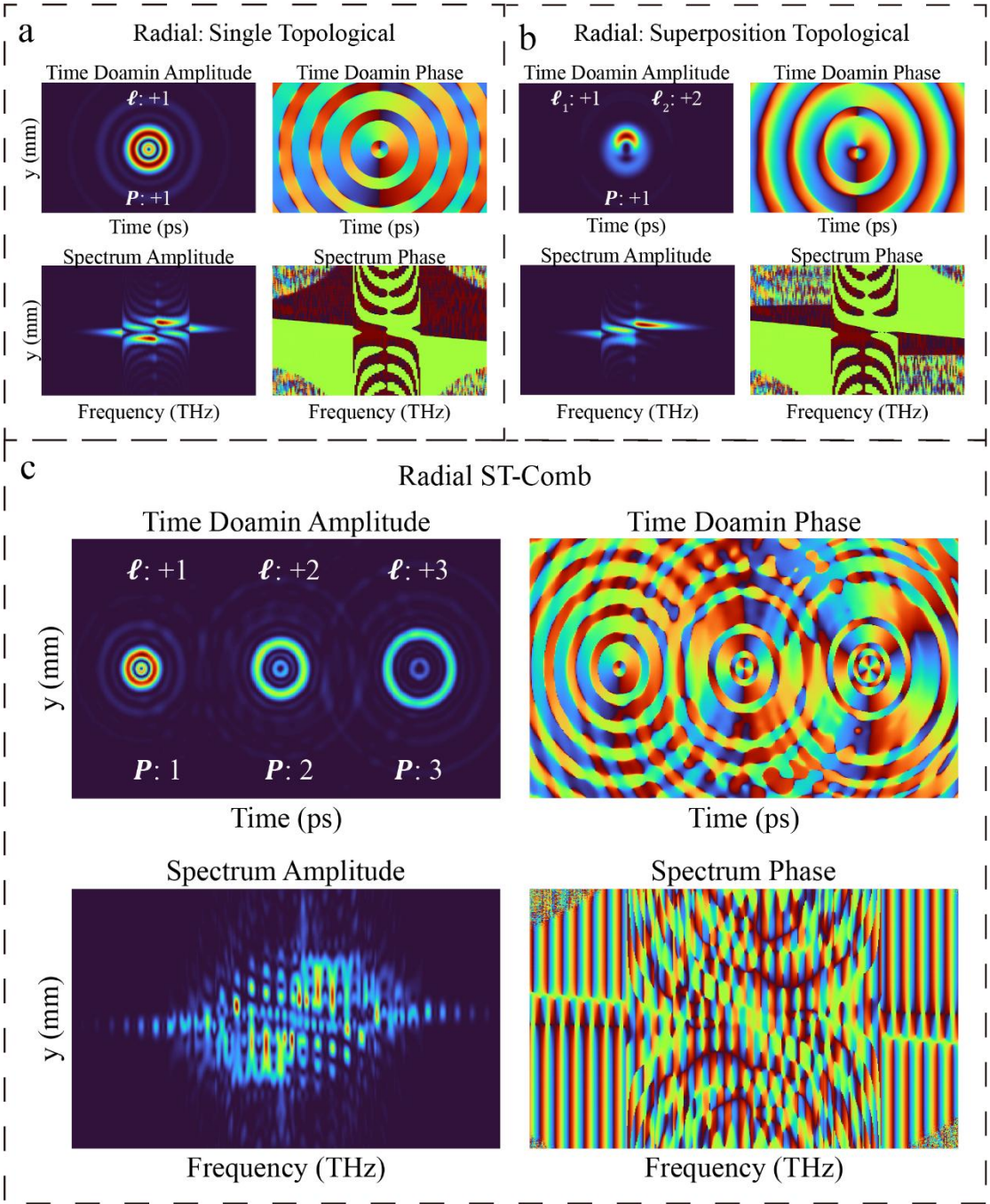

Figure 7: Generation and modulation of the radial ST-Comb. a) Superposition of a single radial and angular topological mode. b) Time and frequency domains of the mixed state formed by the superposition of radial and angular topologies. c) Radial ST-Comb generated by three sub-pulses, showing the spectral comb structure clearly reflecting the encoded topological information.

Since the time coordinates of each topological pulse can be individually controlled via t_n, when we set the time delays of several topological signals to be the same, these pulses naturally form a superposed signal in time. Therefore, each position in the topological pulse sequence can encode a new information dimension independently, while also providing greater flexibility for further information storage. As shown in Fig. 7, panel a display the spatiotemporal wave packet after radial mode and spatiotemporal topology multiplexing, with both radial and angular momentum carried simultaneously. The spectral form exhibits the complex interweaving of these two topologies. Panel b shows the result of multiplexing radial modes and topological superpositions, demonstrating the versatility and flexibility of our encoding scheme. Panel c illustrates the ST-Comb generated by multiplexing radial modes, where the comb structure in the spectral domain clearly reflects the complex encoded topological information.

Section 7: Global initial phase offsets in topological superposition states

In an ST-Comb where each sub-pulse carries a topological superposition state (a coherent sum of modes with different charges), the constituent modes generally possess a global relative initial phase[4]. This offset originates from differences in the reference azimuth φ_0 chosen for each helical mode. Thus, even with a fixed topological charge ℓ, mismatched phase zeros among beams or modes introduce an inherent phase bias. Physically, this appears as a rigid rotation of the entire helical wavefront with respect to the reference frame. While the offset does not alter the charge of any component, it reshapes the interference pattern of the composite field—for example, rotating the fringe orientation as the relative phase shifts. Corresponding fine features in the spectrum likewise depend on these phase combinations. Consequently, the global phase offset is an intrinsic attribute of topological superpositions which, if uncontrolled, can markedly affect both spatial interference structures and spectral signatures[5].

The initial phase offset between superposed components is not merely a manifestation of their interference physics; it is a tunable degree of freedom for information encoding. As shown in Fig. 8, the constructed ST-Comb consists of three sub-pulses, each being

a superposition of topological states with $\ell = 1,2,3$. We introduce different initial phase offsets (0.5π and π) among the components of only the first sub-pulse. The resulting modulation markedly alters the composite waveform in time, yielding distinguishable temporal profiles and spectral features despite an unchanged distribution of topological charges. This demonstrates that by controlling the relative initial phases of the superposed modes, one can generate multiple, clearly discriminable states within a fixed topological-charge set—effectively adding a "phase bit" as an extra encoding dimension. Consequently, phase-offset control enables multilevel information encoding within the same ℓ-subspace and can substantially increase the bit capacity of each sub-pulse.

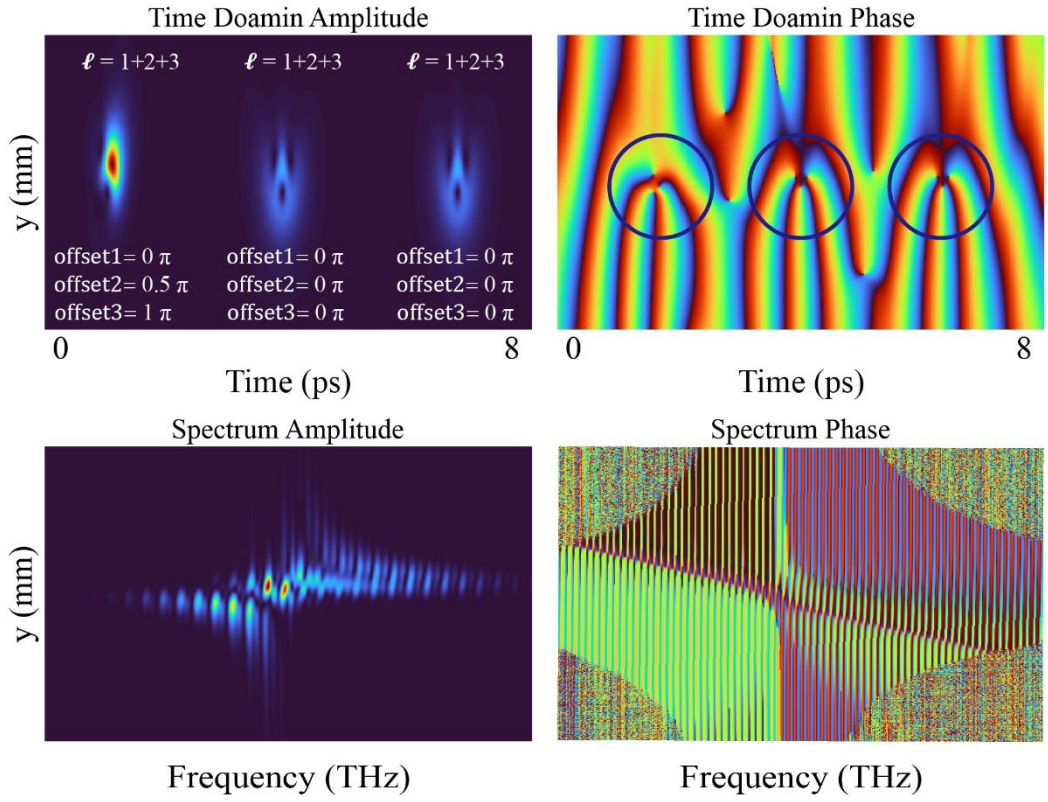

Figure 8: ST-Comb with offset modulation between superposed modes. The comb consists of three sub-pulses, each a superposition of $\ell = 1,2,3$. The relative initial phase between the superposed components is modulated only in the first sub-pulse, yielding a rotated/interferometrically distinct pattern while the others remain unchanged.

Section 8: ST-Comb with 40 sub-pulses

We constructed and verified an ST-Comb pulse train comprising 40 sub-pulses and demonstrated the associated information encoding. As shown in Fig. 9 (the central sub-pulse is highlighted by a red box), we began at the center and expanded the code symmetrically: first “0, 1”, then “0, 0, 1”, then “0, 0, 0, 1”, and so forth—realizing a scalable, structured mapping. Because the detection window is finite, we employed a high-precision delay line and scanned the signal in 10-ps steps. Four independent scans covered the entire train, enabling full time-domain reconstruction of all 40 sub-pulses.

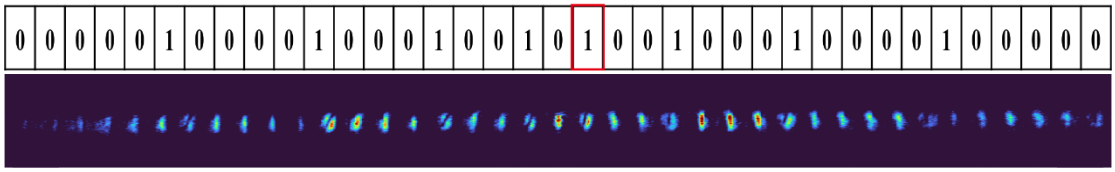

Figure 9: Experimentally generated ST-Comb with 40 sub-pulses.

Beyond tuning each sub-pulse’s topological state, we also treat its amplitude and the inter-pulse delay as independently adjustable degrees of freedom. We therefore propose a three-dimensional joint-modulation scheme—amplitude–time–topology—that markedly enlarges the accessible state space while remaining experimentally practical. The number of attainable states grows exponentially, furnishing a rich physical basis for multi-channel encoding, sparse modulation, and structured reconstruction. This framework lays both theoretical and experimental groundwork for future free-space optical communication systems demanding high capacity, robustness, and flexibility.

Section 9: Communication framework of the ST-Comb

To achieve high-capacity, robust information transmission, we propose a multi-dimensional topological superposition encoding method based on the spatiotemporal optical frequency comb (ST-Comb). This method encodes information from images, audio, video, and other data into binary form, utilizing the highly tunable time and topological charge degrees of freedom of multiple sub-pulses within the ST-Comb. This enables high-density mapping and stable transmission of information within a single optical pulse train. Below, we systematically explain the complete

communication process of this scheme.

As shown in Fig. 10, consider an image as an example. A common approach is to quantize the image's pixels, such as encoding each pixel of a grayscale image with 8 bits (or each RGB component of a color image with 8 bits), and then linearly concatenate them in rows or blocks to form the raw bit stream B_{raw}. To adapt this to the multi-dimensional control structure of the ST-Comb, B_{raw} is divided into bit segments $\{G_1, G_2, \dots, G_M\}$, where each segment G_j corresponds to one sub-pulse in the ST-Comb. For example, using 11 sub-pulses, each segment G_j controls the topological state superposition contained within a single sub-pulse. The predefined topological charge discrete values are $\ell \in \{-7, -6, \dots, +7\}$, resulting in 14 OAM modes, and a fixed mapping is established between bit position k and topological state ℓ_k. If the k-th bit of G_j is 1, the j-th sub-pulse will include the corresponding topological mode φ_{l_k}; otherwise, it will be excluded. For example, the sequence "00000001110000" encodes sub-pulses as a superposition state with $\ell = +1, +2, +3$.

On the receiving end, the ST-Comb signal is recovered using coherent decoding, reconstructing the full waveform and phase sequence in the time domain. Once the time-domain waveform is obtained, it is input into a pre-trained machine learning classifier, ST-Net, to recognize the waveform shape. Since different topological charge combinations correspond to distinct temporal pulse modes, the classifier, after training, can accurately identify the topological state of each sub-pulse. The classification result provides the topological charge information for each sub-pulse. Using the predefined mapping rule, this information is converted back to the original bit segments. Finally, all bit segments are concatenated in sequence to reconstruct the original binary bit stream, recovering the transmitted image data.

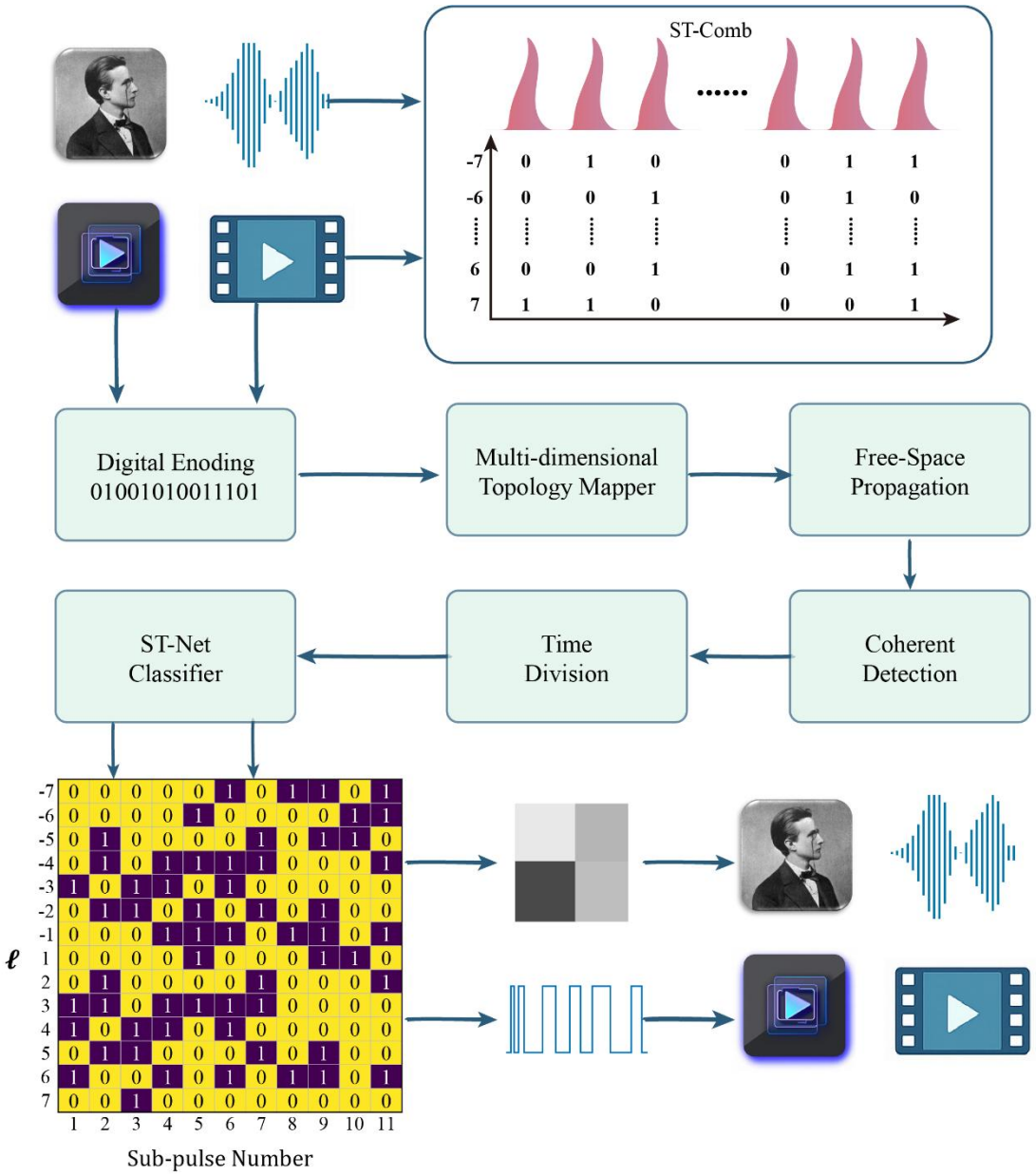

Figure 10: ST-Comb information transmission framework. mages, audio, and video are first transcoded into binary digital information, which is then mapped to the corresponding topological information in the ST-Comb for encoding. After transmission, the encoded ST-Comb signal is sent to a coherent decoding device to reconstruct the time-domain information. The ST-Comb is then temporally divided into multiple sub-pulses, with each sub-pulse sequentially entering ST-Net for topological information decomposition, ultimately recovering the binary digital information and reconstructing the original transmitted data.

Section 10: ST-Net Machine Learning Topology Recognition Model

Orbital angular momentum (OAM) of light offers an infinite set of orthogonal states, theoretically enabling high-capacity information encoding. In our scheme, the ST-Comb signal is reconstructed to obtain the time-domain waveform, and classifying these waveforms allows decoding the transmitted information. As the amount of information encoded in a single ST-Comb increase, the number of temporal modes that need to be distinguished also rises dramatically. Without superposition, there are 7 simple time-domain shapes in the range $\ell = -3$ to $\ell = +3$, which can be easily classified with a standard convolutional neural network (CNN) after minimal training, achieving 100% accuracy. With the introduction of superpositions, the complexity increases exponentially, as the number of time-domain shapes for the range $\ell = -3$ to $\ell = +3$ rises to 64. Traditional physics-based matching methods are inefficient for handling such diverse patterns, and manual matching becomes infeasible. In fact, despite the 64 different waveforms in the time domain, simple CNN models can still achieve 100% accuracy using techniques like data augmentation and transfer learning, but the training becomes more difficult and requires dataset and strategy optimization. Further expanding the topological charge range to $\ell = -7$ to $\ell = +7$ results in 16,384 possible modes, dramatically increasing the number of categories.

To address this high-complexity problem, we designed the specialized ST-Net network architecture. All experiments were conducted under the same system conditions to ensure reliable and comparable results. In each transmission experiment, the ST-Comb signal was processed using the same retrieval algorithm, producing a 1448×1448 pixel grayscale image. This image preserves the full spatiotemporal information of the signal and reflects the interference and attenuation during transmission. For subsequent analysis, the image was normalized and named according to experimental parameters . The images were then pre-processed (denoising, contrast adjustment) to enhance quality and reduce noise interference. Finally, the image size was compressed to 224×224 (suitable for deep learning input), and randomly split into training and validation sets for model training and performance evaluation. Specifically, during the data preparation phase, a stratified random sampling strategy was employed to maintain

class balance, with the following data distribution:

Dataset	Number of Classes	Total Samples	Training Set	Validation Set
No Superposition $\ell = -3$ to $\ell = +3$	7	32561	349	63
Superposition $\ell = -3$ to $\ell = +3$	64	25234	2294	576
Superposition $\ell = -7$ to $\ell = +7$	16384	626845	430464	327680

As the transmitted information and number of categories increase, the demands on the neural network's classification ability and computational efficiency also rise. Particularly for the 16,384-class ST-Comb dataset, the model must distinguish a large number of similar patterns while maintaining training efficiency and generalization. This presents a significant challenge for model design.

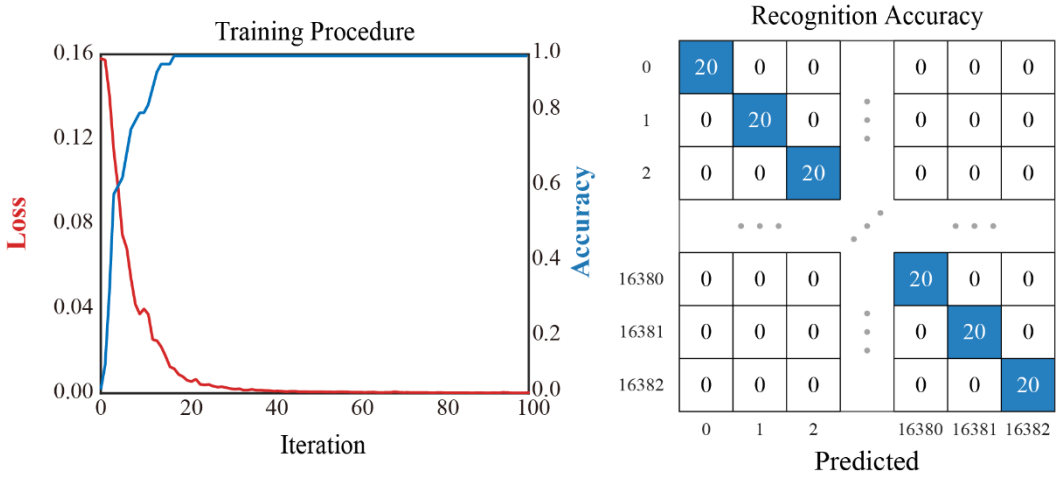

Figure 11: Training process and 100% recognition results for 16,384 superposed states. The figure illustrates the training procedure for the ST-Net model on a dataset containing 16,384 superposed states. The model achieved 100% recognition accuracy across all classes.

In ST-Net, we retained the 16 pre-trained convolutional layers of the VGG19 model and replaced the original fully connected layers. The new network head includes a 512-dimensional dropout layer (dropout rate: 0.5) and a SoftMax output layer with the

number of classes (7 in this case). The model training followed a two-phase fine-tuning strategy. Phase 1 (fully connected layers only), freeze all convolutional layers, use the Adam optimizer (initial learning rate 1×10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$) to train the fully connected layers for 20 epochs. Phase 2 (full network fine-tuning), unfreeze the last four convolution blocks and perform full network fine-tuning with cosine-annealed learning rate scheduling. After fine-tuning, the model achieved 100% classification accuracy on 32561 test samples, with a Kappa coefficient of 1000. Precision, recall, and F1 scores for all classes were ideal. The accuracy and loss curves, as well as the confusion matrix, are shown in Fig. 11.

To evaluate the advantages of ST-Net, we also trained a simple CNN and a standard VGG19 model on the 64-class dataset. Experimental results show that the simple CNN failed to reach 100% accuracy. Although the VGG19 model also converged to 100% accuracy, it required nearly twice the training time of ST-Net. This comparison further demonstrates that ST-Net offers faster training convergence and superior classification performance for handling large ST-Comb time-domain shape classification tasks.

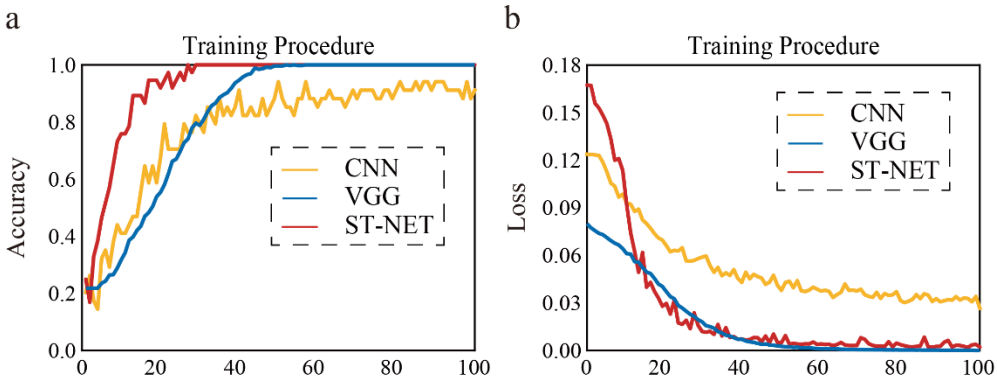

Figure 12: Training performance comparison between ST-Net and baseline models. ST-Net converges markedly faster and attains higher final accuracy than the two reference models (simple CNN and VGG19), demonstrating clear advantages in both training speed and classification fidelity.

Section 11: Resilience of the ST-Comb to Disturbance

In practical applications, the ST-Comb proposed in this work offers significant

advantages in interference resistance over traditional spatial OAM optical communication. When light pulses are transmitted in free space, atmospheric turbulence introduces severe impacts, such as beam jitter, scintillation, beam spread, and image jitter[6]. While information encoded in the spatiotemporal domain is significantly disrupted by turbulence-induced phase distortions, the information in the spectral domain, after Fourier transform, maintains much higher fidelity compared to conventional OAM spatial beams[7]. This phenomenon has been validated through both simulations and experiments.

To simulate the effects of atmospheric turbulence on pulse transmission, we employ the phase screen method. Several thin phase screens model the additional noise phase distribution caused by long-distance atmospheric turbulence. These turbulence phase screens can be viewed as a 2-D random field, generated from its power spectral density (PSD) via an inverse Fourier transform.

The phase screen is generated as follows:

$$\phi(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Psi\left(f_x, f_y\right) e^{i2\pi\left(f_x x + f_y y\right)} df_x df_y$$

where $\Psi\left(f_x, f_y\right)$ is the continuous phase screen spatial frequency spectrum, f_x, f_y represent the spatial frequencies. In the discrete case, the expression becomes:

$$\phi(x,y) = \sum_{n-\infty}^{\infty}\sum_{n-\infty}^{\infty} c_{n,m} \exp\left[i2\pi\left(f_{x_n} x + f_{y_n} y\right)\right]$$

Here, $c_{n,m}$ are the discrete Fourier coefficients, which are obtained using Gaussian spectral coefficients, symmetric in the radial direction:

$$c\left(f_x, f_y\right) \sim \mathcal{CN}\left(0, \Phi_\phi(f)\Delta f^2\right)$$

where $\Phi_\phi\left(f_x, f_y\right)$ is the simplified von Kármán phase power spectrum, and r_0 is the coherence diameter, calculated as:

$$r_0 = (0.423 \cdot k^2 \cdot C \cdot Z)^{-\frac{3}{5}}$$

where C is the atmospheric refractive index structure constant, and Z is the propagation distance. The von Kármán phase power spectrum is expressed as:

$$\Phi_{\varphi}(f_x, f_y) = 0.023 \cdot r_0^{-\frac{5}{3}} \cdot exp(-\frac{f^2}{f_m^2})(f_0^2 + f^2)^{-\frac{11}{6}}$$

where $f_m = \frac{5.92}{2\pi l_0}$ is the inner-scale frequency, $f_0 = \frac{1}{L_0}$ is the outer-scale frequency, and l_0, L_0 are the inner and outer scale sizes.

Traditional FFT methods cannot simulate low-frequency components with periods larger than the aperture. Therefore, the formula above generates only the high-frequency components of the phase screen:

$$\varphi_{\text{hi}} = \phi_{\text{hi}}(x, y) = \Re\left\{\sum_{k,l} c\left(f_{x,k}, f_{y,l}\right) e^{i2\pi(f_{x,k}x + f_{y,l}y)}\right\}$$

To include low-frequency components, we add the low-frequency portion of the phase:

$$\phi_{lo}(x, y) = \sum_{p=1}^{P} \sum_{m,n=-1}^{1} c_{m,n}^{(p)} e^{i2\pi\left(f_{x,m}^{(p)}x + f_{y,n}^{(p)}y\right)}$$

where p represents different subharmonic orders, and the frequency steps Δf_p are defined as:

$$\Delta f_p = \frac{1}{3^p D}, \quad c_{m,n}^{(p)} \sim \mathcal{CN}\left(0, \Phi_{\phi}\left(f_p\right)\Delta f_p^2\right)$$

where $p = 1,2,3$ corresponds to the first three subharmonic orders. The total phase screen is:

$$\phi(x, y) = \phi_{hi}(x, y) + \alpha \cdot \phi_{lo}(x, y)$$

where α is the weight factor for the low-frequency components. We used a single-phase screen to simulate the effects of atmospheric turbulence on transverse OAM (STOV) and longitudinal OAM beams with topological charge $\ell = 1$ under various conditions, as shown in the figures. The primary evaluation metric is the intensity distribution at the receiver and its similarity to the original beam.

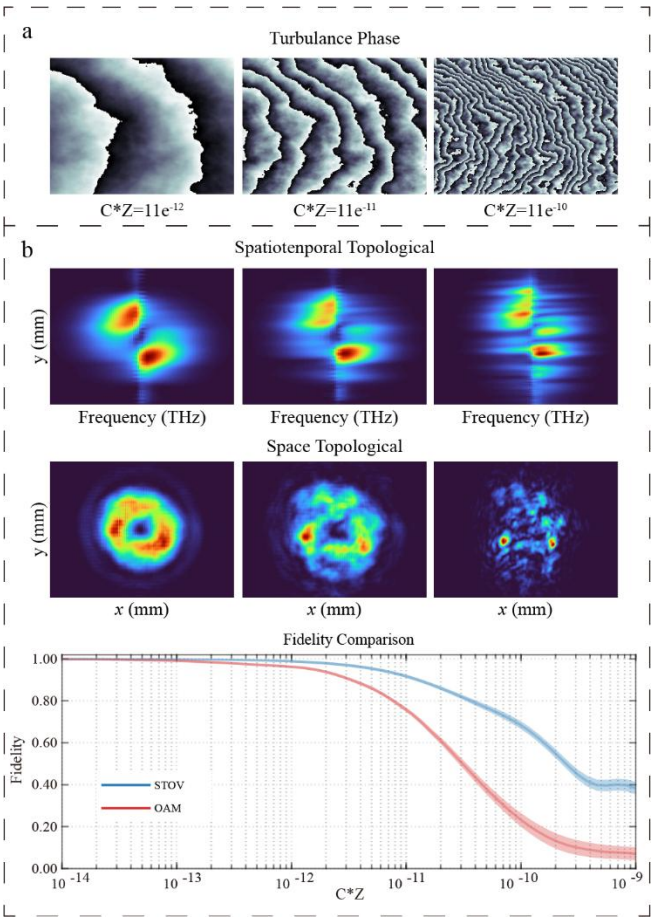

Figure 13: Comparison of spectral and spatial information of STOV and spatial OAM under three different turbulences.

It should be noted that, in our model, the effects of the atmospheric refractive index structure constant C and the transmission distance Z are similar. Therefore, we treat the product of C and C as a physical quantity representing the total turbulence strength. In ideal conditions with weak turbulence, the longitudinal OAM beam shows a clear annular intensity profile in the far-field. However, after introducing turbulence, the simulation results clearly show that the intensity ring of the longitudinal OAM beam quickly breaks up into irregular speckles, and its frequency spectrum also blurs, with energy spreading from the characteristic ring to a broader frequency range, causing the

original spectral features to almost disappear.

In contrast, the STOV exhibits significantly better interference resilience. Despite the distortion and jitter in its intensity distribution under turbulence, the core energy structure of the beam remains much more consistent. In the spatial frequency spectrum, the primary topological features also experience some broadening, fragmentation, and noise base elevation, but their relative positions and overall profile are maintained. We used the spectral correlation coefficient as a quantitative measure and calculated the cross-correlation between the disturbed and ideal spectra. The results show that the spectral fidelity of the transverse OAM beam remains significantly higher than that of the longitudinal OAM beam under the same turbulence conditions[8]. This indicates that the information carried by the transverse OAM beam is better protected in the frequency domain.

We further studied the impact of turbulence intensity C_n. The simulation results indicate that turbulence is a cumulative effect: as the transmission distance or turbulence intensity increases, the total amount of phase distortion the beam undergoes also increases, leading to further degradation in both types of beams. For the longitudinal OAM beam, the ring structure can still be faintly discerned under weak turbulence. However, as the turbulence intensity increases, its intensity and spectral features are completely overwhelmed by noise, making them unrecognizable. The advantage of the transverse OAM beam becomes particularly evident during high turbulence intensity transmission. Although its spectral fidelity also decreases with increasing turbulence intensity, the rate of decrease is much slower than that of the longitudinal OAM beam. Even at transmission distances where the longitudinal OAM mode has completely failed, the spectral features of the transverse OAM beam remain significantly identifiable. This resilience arises from its ability to carry information in the frequency domain, allowing it to better resist phase distortions accumulated over long-distance transmission across multiple scales.

Section 12: High-Speed Modulation Communication Scheme for ST-Comb

In this work, we have demonstrated the communication potential of a single ST-Comb pulse carrying 154 bits of information. If each pulse from a laser with a frequency on the order of kilohertz (KHz) or even megahertz (MHz) could be modulated, the communication capacity per second would be immense. However, most commercially available spatial light modulators (SLM) can only achieve frame rates up to 60 Hz, making high-speed modulation infeasible with current technology.

With the development of metasurfaces and waveguides, Panuski introduced a programmable high-speed spatial light modulator (DoF-SLM) based on a two-dimensional photonic crystal microcavity array[9]. They demonstrated an 8×8 (64-pixel) microcavity array that can switch wavelength-scale modes at nanosecond scale ($\tau \approx 1.1\,\mathrm{ns}$) and achieve femtosecond-level energy consumption. For example, the modulation bandwidth $\omega_s \approx 2\pi \times 135\,\mathrm{MHz}$ was obtained through transmission function measurements driven by optical pulses. In comparison, conventional liquid crystal or MEMS SLMs typically operate at frame rates of 60–200 Hz, making the DoF-SLM a monumental leap in modulation speed. Additionally, this system supports programmable pixel-level control and on-chip integration, meeting the space and time bandwidth limit conditions (i.e., achieving nearly full aperture mode coverage and modulation bandwidth approaching the delay limit), opening up a new era of fully adjustable multimode light fields.

Combining the ST-Comb structure with the capabilities of this high-speed SLM elevates pulse shaping and topological state control from static to real-time dynamic modulation. The megahertz modulation bandwidth of the DoF-SLM means that each femtosecond-to-picosecond pulse can be programmed individually, instead of using fixed patterns, significantly enhancing the accuracy and flexibility of single-pulse control. For example, the OAM state of the output pulse can be reconfigured at any time, eliminating the stability and refresh rate limitations of slower SLMs. This allows precise control of the complex superposition state structure in each pulse, improving the controllability of time-domain modes. Furthermore, with a modulation bandwidth reaching the MHz range, the ST-Comb system's pulse repetition rate can be drastically increased, leading to an exponential expansion in communication capacity.

In other words, by enabling real-time topological state encoding for each pulse at megahertz rates, the system's transmission rate could theoretically be pushed to the terabit-per-second (Tbps) level (modulating millions of topological state symbols per second). Additionally, the on-chip integration and all-optical control logic of the DoF-SLM are highly compatible with the ST-Comb system's architecture. The CMOS-programmable micro-LEDs and micro-cavity trap arrays offer flexible digital control interfaces, enabling high-density topological state operations (similar to OAM mode switching) within integrated optical circuits. Overall, this high-speed spatial light modulator not only overcomes the bandwidth bottleneck of traditional SLMs but also enables the potential for Tbps-level high-speed communication in ST-Comb systems under megahertz pulse control.

As an alternative to large-scale DoF-SLM deployment, we propose a scheme using high-speed electro-optic modulators to achieve sub-pulse superposition topological state control. First, a single incident broadband pulse is split into multiple independent channels using a beam splitter. Then, each channel passes through an electro-optic phase modulator pre-set with different topological phases. Finally, the beams from the different channels are coherently combined in a combiner, forming an ST-Comb output carrying multimodal composite topological states.

Modern electro-optic modulators, such as silicon-based PN junction modulators and thin-film lithium niobate MZM (Mach-Zehnder Modulators), now support bandwidths in the tens of GHz and even over 100 GHz. This scheme vastly outperforms traditional SLMs in terms of speed. For example, silicon-based devices have been reported to achieve bandwidths around 90 GHz[10]. Meanwhile thin-film lithium niobate modulators have surpassed a 3 dB bandwidth of 100 GHz[11]. Based on this, the OAM phase modulation for each sub-channel can be completed on the nanosecond to picosecond timescale, allowing the topological state of the entire composite pulse to independently vary during each modulation cycle. This enables real-time control at GHz rates.

Furthermore, integrated optical components such as multi-mode interferometric (MMI) splitters and arrayed waveguide gratings can effectively reduce volume and losses. While the number of channels increases with system complexity, modern photonic

integration technologies (CMOS-compatible processes) are rapidly advancing, and multi-channel interconnection has become quite mature[12]. The sub-pulse superposition modulation scheme based on high-speed electro-optic switches offers strong practicality. It leverages existing GHz-scale modulator devices, allowing the ST-Comb system to expand into real-time multimode superposition without waiting for DoF-SLM commercialization. This scheme can serve as a transitional path for achieving high-speed communication with ST-Comb and provides new ideas for topological state multiplexing in integrated optical circuits. It also lays the foundation for on-chip ultra-high-capacity optical communication systems.

Supplementary References

1. Wan, C., Chong, A. & Zhan, Q. Optical spatiotemporal vortices. *eLight* **3**, 11 (2023).
2. Jhajj, N. *et al.* Spatiotemporal Optical Vortices. *Phys. Rev. X* **6**, 031037 (2016).
3. Gui, G., Brooks, N. J., Kapteyn, H. C., Murnane, M. M. & Liao, C.-T. Second-harmonic generation and the conservation of spatiotemporal orbital angular momentum of light. *Nat. Photonics* **15**, 608–613 (2021).
4. Guo, X. *et al.* Intense vortical-field generation using coherent superposition of multiple vortex beams. *Sci. Rep.* **13**, 1104 (2023).
5. Lu, L., Joannopoulos, J. D. & Soljačić, M. Topological photonics. *Nat. Photonics* **8**, 821–829 (2014).
6. Peters, C., Cocotos, V. & Forbes, A. Structured light in atmospheric turbulence—a guide to its digital implementation: tutorial. *Adv. Opt. Photonics* **17**, 113–184 (2025).
7. Zhu, Z. *et al.* Compensation-free high-dimensional free-space optical communication using turbulence-resilient vector beams. *Nat. Commun.* **12**, 1666 (2021).
8. Zhang, R. *et al.* Turbulence-resilient pilot-assisted self-coherent free-space optical communications using automatic optoelectronic mixing of many modes. *Nat. Photonics* **15**, 743–750 (2021).
9. Panuski, C. L. *et al.* A full degree-of-freedom spatiotemporal light modulator. *Nat. Photonics* **16**, 834–842 (2022).
10. Han, C. *et al.* Exploring 400 Gbps/λ and beyond with AI-accelerated silicon photonic slow-light technology. *Nat. Commun.* **16**, 6547 (2025).
11. Zhang, M., Wang, C., Kharel, P., Zhu, D. & Lončar, M. Integrated lithium niobate

electro-optic modulators: when performance meets scalability. *Optica* **8**, 652–667 (2021).

12. He, C., Shen, Y. & Forbes, A. Towards higher-dimensional structured light. *Light Sci. Appl.* **11**, 205 (2022).