Supplemental Material

Deep-learning-assisted inverse design of dual-spin/frequency metasurface for quad-channel off-axis vortices multiplexing

Kai Qu, Ke Chen*, Qi Hu, Junming Zhao, Tian Jiang, and Yijun Feng*
Nanjing University, School of Electronic Science and Engineering, Nanjing, China

E-mail: ke.chen@nju.edu.cn, yjfeng@nju.edu.cn

S1. Complexity of the Design of Dual-frequency Spin-Decoupled Metasurface Element

Full 360° phase coverage under illumination of circularly polarized (CP) wave can be achieved by only varying the rotation angle of the element, which is one of the reasons for Pancharatnam–Berry phase (PB phase, or called geometric phase) being widely used in CP wave modulation. Thanks to this principle, it will take no more stages to construct a PB phase-based Metasurface working at single frequency (shown in Fig. S1). Therefore, to design a wave functionality based on PB metasurface, we only need to perform two-stage work: first, we need to design an element that meets the condition of PB phase (the relative phase difference between two orthogonal linearly polarized (LP) wave is π but with identical high amplitude responses); second, we calculate the distribution of rotation angle according to target wavefront. As for propagation phase, the metasurface will take three stages, because in addition to designing the corresponding structure, we also need to calculate the distribution of *x*-LP and *y*-LP, respectively. For spin-decoupled metasurface design, it combines the above two kinds of phases, so there are at least four stages: PB phase structure design and three distributions calculation (orientation angle and propagation phase for *x*-LP and *y*-LP).

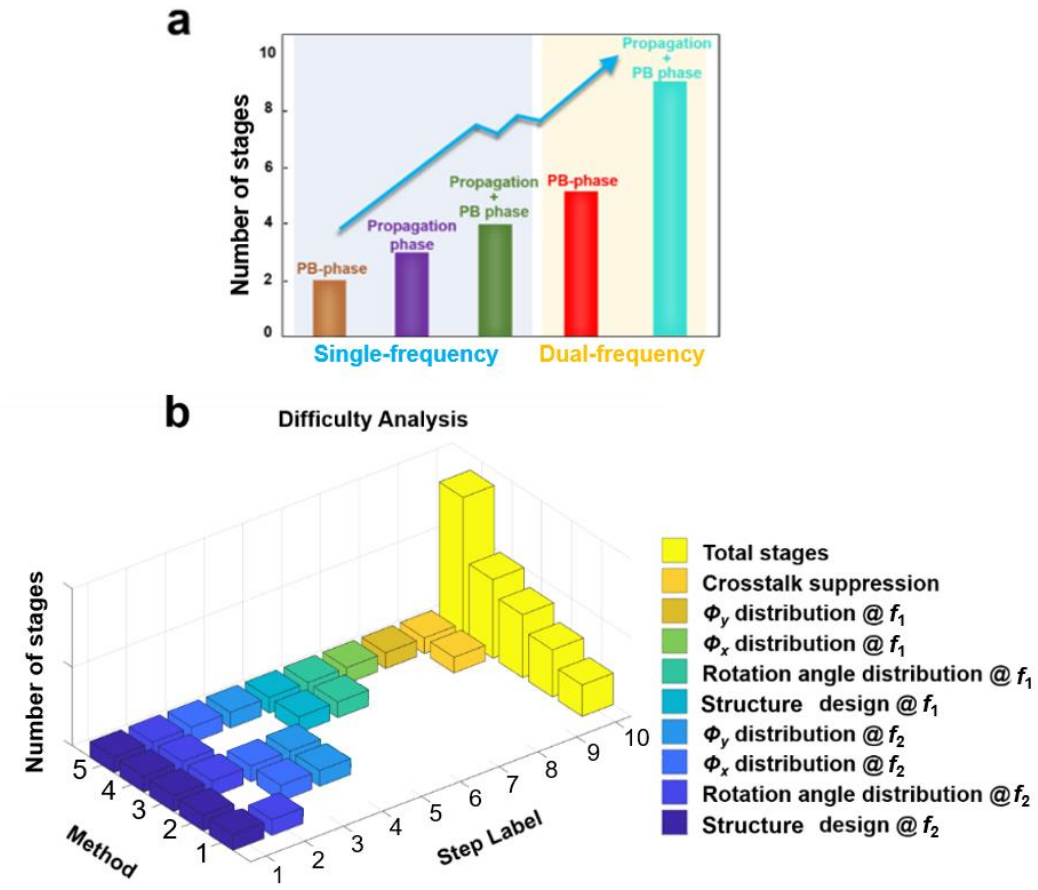

Fig. S1 (a) Complexity of designing metasurface based on different modulation principle at single and dual frequency. (b) The specific stages required by several different methodologies. Method 1-5 correspond to the five ones from left to right shown in (a). Φ_x and Φ_y are the propagation phase for *x*-LP and *y*-LP waves, respectively. λ_1 and λ_2 are the wavelengths for the two operation frequencies, respectively.

Frequency will offer more possibilities for increasing channel capacity and device compactness. Nevertheless, multi-band/wavelength metasurface also puts forward higher requirements for design and calculation, because crosstalk between different frequency bands has been a limitation for frequency-multiplexing meta-device. Figure S1 shows the stages for design dual-band metasurface. Obviously, they are more than twice that required for single frequency. In fact, designing structures corresponds to response at dual frequency and sparing no efforts to suppress crosstalk are responsible for these stages. As can be seen, applying and combining both the PB phase and propagation phase at dual frequencies as well as the suppression of crosstalk is the most complicated compared to the previous ones. Although there

are a few more stages, it takes much more computational and time cost, which is nonlinear with the number of added stages, due to the increase of variables and the unpredictability of crosstalk.

S2. Effect of the Top and Middle Layers of the Meta-atom on the High Frequency Response

The structures in the first and middle layers are identical, and together they provide a strong response at the high frequency of 17.8 GHz. The purpose of this two-layer configuration is to increase the high-frequency propagation phase coverage (co-polarized phase under the linear-polarized incidence) when the length of high-frequency resonators varies. As a verification, we perform full-wave simulations for the structure without the top layer and the structure without the middle layer, respectively. Besides, we compare the propagation phase coverage in these two cases with that of the structure proposed in the main text. As shown in Fig. S2(a), for the configuration applied in the main text, a wide phase coverage near 360° can be obtained when l_{x2} is changed from 3.3 mm to 4.5 mm. However, the propagation phase coverage drops to less than 100° either the top-layer structure or the middle-layer one is removed, as clearly shown in Figs. S2(b)-(c).

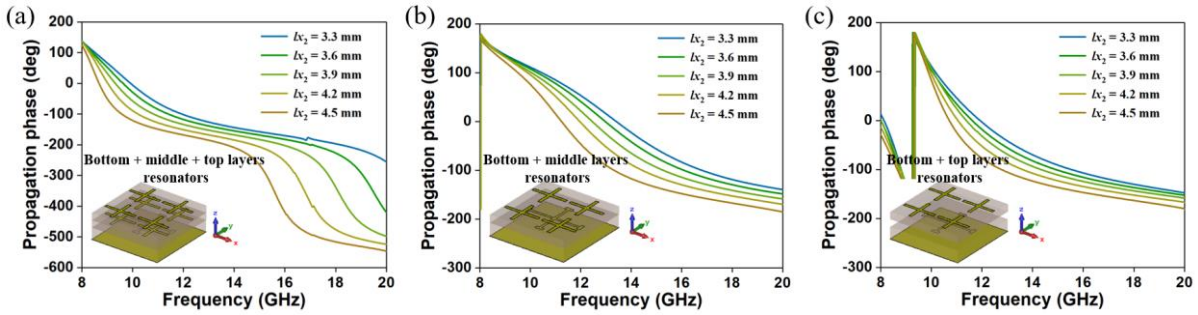

Fig. S2 Under the illumination of *x*-polarized plane wave, the co-polarization spectra of the meta-atoms with different configurations. (a) Configuration in the main text, (b) configuration without top layer, and (c) configuration without middle layer.

S3. Surface Current Distributions of the Proposed Element

The dual-band working principle of the element is shown in Fig. S3 and Fig. S4. When under illumination of plane wave (*x*-LP, *y*-LP, RCP, and LCP) at two frequencies, different metal cross resonators respond distinctly to the incidence wave. Resonant frequency is related to the size

of resonators. Specifically, larger structures will resonant at low frequency, while the smaller ones at high frequency. Firstly, we analyze the working principle of propagation phase at these two frequencies according to electric field distribution. Herein, we take the element with the relative phase modulation 0°/0°/0°/0° in dual CP channels at two frequencies as an illustrative example. The element is designed with the parameters $\{l_{x1}, l_{y1}, \alpha, l_{x2}, l_{y2}, \beta\}$ being 8.50 mm, 5.03 mm, 0°, 4.50 mm, 3.91 mm, and 0°, respectively. α and β are the rotation angles of the low-frequency resonator and high-frequency resonators, respectively. l_{x1} and l_{y1} are the lengths of mental arms of the low-frequency resonator in x-direction and y-direction, respectively. l_{x2} and l_{y2} are the lengths of mental arms of the high-frequency resonators in x-direction and y-direction, respectively.

As shown in Fig. S3(a) and S3(b), when y-LP (x-LP) plane wave at 9.8 GHz is incident on the element, there is obvious current concentrating on the metal arms in y (x) direction of all the resonators. Moreover, the main current distribution is still on the large resonator, and the electric field on small cross resonators will cause an uncertain crosstalk on the low-frequency propagation phase. When the frequency is adjusted to 17.8 GHz, there will be similar current distribution on all the small resonators (Fig. S3(c) and S3(d)). On the contrary, there is almost no surface current on the large resonator, which means that the operating mode of high-frequency propagation phase is hardly affected by crosstalk from the large resonator.

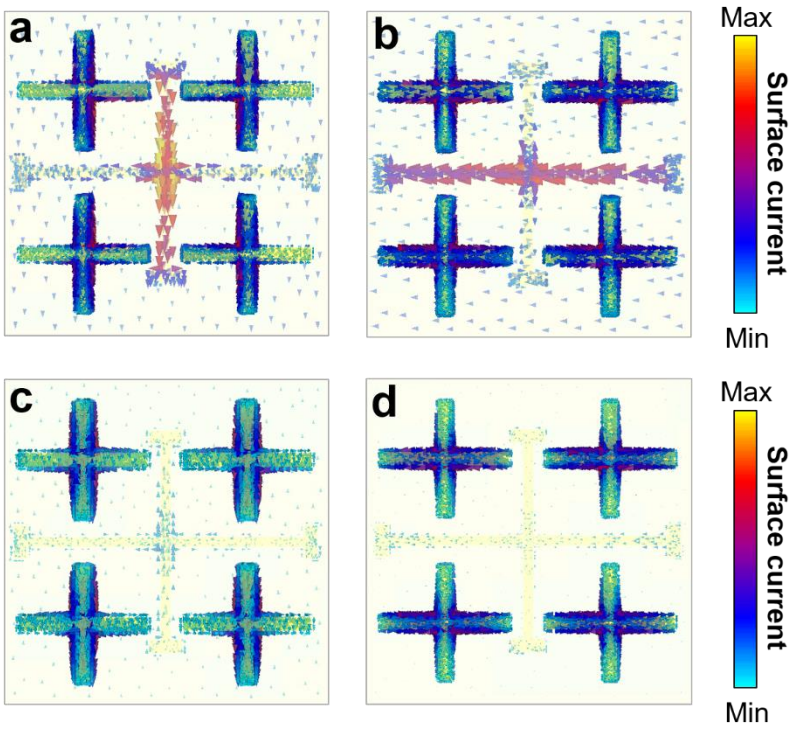

Fig. S3 Top view of simulated surface current distributions of the element under the illumination of plane waves with different frequencies and linear polarizations. (a) *y*-LP incidence at 9.8 GHz. (b) *x*-LP incidence at 9.8 GHz. (c) *y*-LP incidence at 17.8 GHz. (d) *x*-LP incidence at 17.8 GHz.

The above analysis is also applicable to the incidence of CP EM waves, which is shown in Fig. S4. Aiming at proving its universality, we set all the resonators to rotate at a certain angle with some overlap. Figures S4(a) and S4(c) indicate that the all the resonators resonate at 9.8 GHz, in which large currents are observed on the largest resonator. However, it doesn't resonant at 17.8 GHz (shown in Fig. S4(b) and S4(d)), greatly avoids crosstalk to high frequency channels.

These phenomena clearly illustrate that the crosstalk between two frequency bands is "unidirectional". In other words, there only exists the crosstalk from the high-frequency resonators to low-frequency channel.

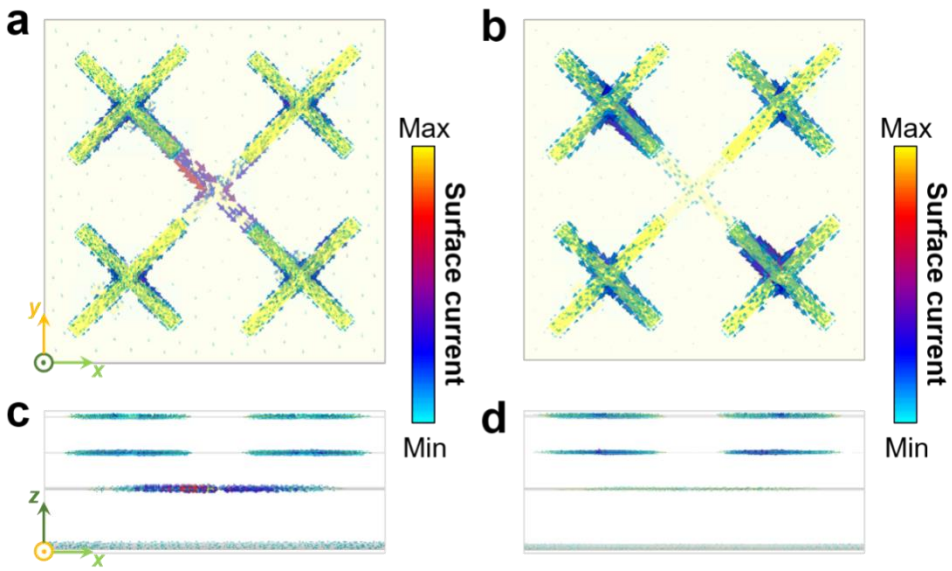

Fig. S4 Simulated surface current distribution of the element with the incidence of CP waves. (a) Top view at 9.8 GHz. (b) Top view at 17.8 GHz. (c) Side view at 9.8 GHz. (d) Side view at 17.8 GHz.

S4. Analysis of the Crosstalk Between the Two Frequency Bands

By adjusting the rotation angles (α and β for low frequency and high frequency, respectively) and lengths (l_{x1} (l_{y1}) and l_{x2} (l_{y2}) for low frequency and high frequency, respectively) of the resonators in the proposed element, the element can be mapped to geometric phase and propagation phase in dual band, further attaining the independent control of two orthogonal CP channels in two frequency bands. Specifically, the frequency- and polarization- crosstalk mainly comes from two parts: the propagation phase crosstalk and geometric phase crosstalk.

The modulation of propagation phase requires length parameters (l_{x1}, l_{y1}, l_{x2} and l_{y2}) to be changed, which may bring the crosstalk of propagation phase in the two bands due to the correlation between the resonant frequency and the length of the resonator. As shown in Figs. S5(a) and S5(b), when l_{x1} is kept as 6 mm, changing the value of l_{x2} from 2.5 to 8.5 mm will bring propagation phase crosstalk of about 100° and slight amplitude crosstalk to EM response at low frequency. On the contrary, the propagation phase crosstalk from the variation of l_{x1} to high frequency will be limited to no more than 1° when changing l_{x1} (shown in Fig. S5(d)). Besides, the amplitude crosstalk to high frequency is also slight and allowable (Fig. S5(c)).

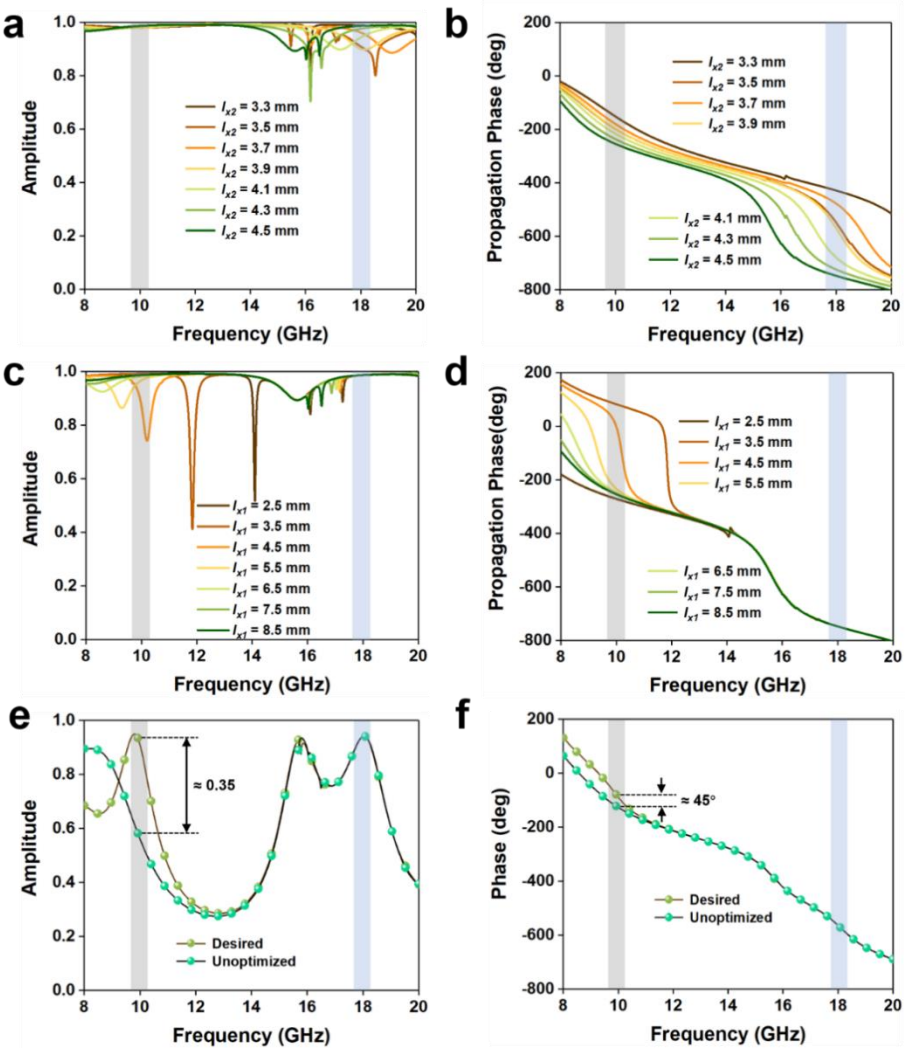

Fig. S5 Influence of propagation phase crosstalk on LP/CP reflection spectra in the two frequency bands. The (a) amplitude and (b) phase of the *x*-LP co-polarized reflection coefficient when l_{x1} is fixed at 8.5mm and l_{x2} varies. The (a) amplitude and (b) phase of the *x*-LP co-polarized reflection coefficient when l_{x2} is fixed at 4.3 mm and l_{x1} varies. $\{ly_1, \alpha, ly_2, \beta\}$ is fixed as {5 mm, 0°, 3.9 mm, 0°}. The influence of the propagation phase crosstalk on (e) amplitude and (f) phase in LCP-LCP/RCP-RCP channel.

In order to obtain high efficiency in the co-polarized channel, the phase difference of two LP propagation phase should be fixed at 180°. However, the abovementioned crosstalk of propagation phase will make it challenging to effectively control the phase difference in element design. Herein, we take the element in Fig. S3 as an example. When l_{x2} varies from 4.5 mm to 4.0 mm, the amplitude and phase response at low frequency will deviate from the results we desire once the low-frequency physical parameters are not optimized (Figs. S5(e) and S5(f)). In general, both the amplitude error and phase error will occur at low frequency due to the

propagation phase crosstalk from the high-frequency physical parameters. Actually, the amplitude error and phase error are not limited in the range of 0.35 and 45°, respectively. It will be larger when more high-frequency physical parameters are changed as the general case.

Besides the amplitude/phase error due to the propagation phase crosstalk, the crosstalk of geometric phase can also cause another part of phase error. We still take the element shown in Fig. S3 as the example. As shown in Fig. S6, when α and β varies independently, only the low frequency channel is subject to crosstalk from the variation of β.

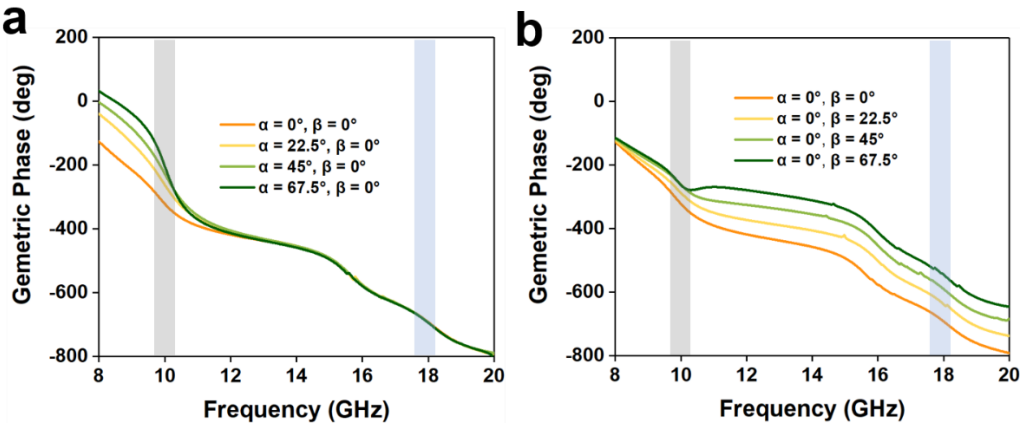

Fig. S6 Phase response of reflected CP wave when only (a) α or (b) β varies.

S5. Explanation of the Degree of Influence in the Radar Chart

In order to intuitively clarify the relationship between the reflected CP phase at two frequencies and the physical parameters (l_{x1}, l_{y1}, α, l_{x2}, l_{y2} and β), we propose and define a criterion to describe the influence of physical parameters vary onto the phase variation. The circularly polarized channels here refer to the reflected co-polarized channels, unless otherwise specified. Based on the analysis of the reflection spectrum of the meta-atoms, four levels are defined, as shown in Table S1. φ' is the maximum phase shift imposed by the variation of one of the physical parameters. For example, a PB phase shift coverage of 360° can be imposed with the variation of β for the high-frequency circularly polarized channels, for which the effect level of β on the high-frequency channels is "Fundamentally affect". Particularly, due to the mirror relationship between LCP and RCP, the effect levels of the six parameters on their phases can

be considered equivalent. Therefore, the phase channels at f_1 and f_2 in Fig. 2(g) represent the RCP (or LCP) channels at the low frequency and high frequency.

Table S1. The definition of the impact levels

Level	Description	Impact on Phase (φ')
1	Hardly affect	$\varphi' < 10°$
2	Mildly affect	$10° \le \varphi' < 90°$
3	Severely affect	$90° \le \varphi' < 180°$
4	Fundamentally affect	$\varphi' \ge 180°$

S6. The Prior-Knowledge-Guided Adaptively Optimization of the Physical Parameters

The combination of propagation phase and geometric phase can independently tailor the wavefront of two orthogonal CP waves. For the proposed meta-atom, if there is no crosstalk from the high-frequency structure, the theoretical propagation phase and rotation angle α at low frequency can be easily obtained by,[24, 28]

$$\varphi_{x1} = \frac{\varphi_{rcp1\_target} + \varphi_{lcp1\_target}}{2}, \tag{S1}$$

$$\varphi_{y1} = \varphi_{x1} - \pi, \tag{S2}$$

$$\alpha = \frac{\varphi_{rcp\_target} - \varphi_{lcp\_target}}{4}. \tag{S3}$$

Where φ_{x1} and φ_{y1} are the propagation phase for x-LP and y-LP reflected waves at low frequency, respectively. φ_{rcp1_target} and φ_{lcp1_target} represent the desired phase shift for RCP incidence and LCP incidence at low frequency, respectively.

Due to the existence of propagation phase and geometric phase crosstalk, there will be an uncertain error between the actual low-frequency phase modulation of RCP and LCP and the target ones. Moreover, the CP reflection efficiency will also be easily reduced by the crosstalk effect. Therefore, we have summarized the experience based on the mechanism of the spin-decoupled metasurface, and obtained the prior knowledge for optimizing part of the physical parameters of the proposed element.

In order to gradually narrow the gap between the actual CP phase shift and the target

values in every iteration process, the current low-frequency parameters $\{\alpha, \varphi_{x1}, \varphi_{y1}\}$ should be replaced with the $\{\alpha', \varphi_{x1}', \varphi_{y1}'\}$ according to each simulated result for the element. The current simulated phase shifts for RCP and LCP at low frequency are denoted as φ_{rcp1} and φ_{lcp1}, respectively. The adaptive adjustment principle based on prior knowledge (Equations S1-S3) can be expressed by

$$\Delta\varphi_{x1} = \frac{(\varphi_{rcp1\_target} - \varphi_{rcp1}) + (\varphi_{lcp1\_target} - \varphi_{lcp1})}{2}, \tag{S4}$$

$$\Delta\alpha = \frac{(\varphi_{rcp1\_target} - \varphi_{rcp1}) - (\varphi_{lcp1\_target} - \varphi_{lcp1})}{4}, \tag{S5}$$

$$\alpha' = \alpha + k \cdot \Delta\alpha, \tag{S6}$$

$$\varphi_{x1}' = \varphi_{x1} + k \cdot \Delta\varphi_{x1}, \tag{S7}$$

$$\varphi_{y1}' = \varphi_{x1}' - \pi. \tag{S8}$$

Where $\Delta\varphi_{x1}$ is the average phase error in RCP channel and LCP channel at low frequency. The average error should be subtracted by φ_{x1} and φ_{y1}. $\Delta\alpha$ is the part that should be tuned for the current α. The phase shifts of two orthogonal CP waves brought by $\Delta\varphi_{x1}$ and $\Delta\alpha$ can be divided into two parts: the propagation phase shift ($\Delta\varphi_{x1}$) and the geometric phase shift ($\pm 2\Delta\alpha$). Importantly, the purpose of Equation S5 is to make the phase shift of RCP ($\Delta\varphi_{x1} + 2\Delta\alpha$)[24] equal to the phase difference between the current phase and the target ones ($\varphi_{rcp1\_target} - \varphi_{rcp1}$). For LCP, its phase shift of $\Delta\varphi_{x1} - 2\Delta\alpha$ can also offset the phase difference of $\varphi_{lcp1\_target} - \varphi_{lcp1}$. k is a coefficient set between 0-1 to avoid getting stuck in the locally optimal solution, which is similar to the learning rate in machine learning. In short, if we continuously replace the current three variables $\{\alpha, \varphi_{x1}, \varphi_{y1}\}$ with $\{\alpha', \varphi_{x1}', \varphi_{y1}'\}$, we can gradually get the final desired meta-atom with accurate phase modulation.

S7. Classifier Based on Neural Network and Multi-Layer Perceptron (MLP) model

Neural network, whose full name is artificial neural network (ANN), can use computers or

physically achievable devices to simulate certain structures and functions of neural networks in organisms, and obtain information processing mechanisms inspired by biological nervous systems. A single neuron has a simple structure and limited functions, while a large number of neurons can form a highly complex nonlinear adaptive system. The application and research of neural networks covers a wide range of fields, such as pattern recognition, computer vision, and language recognition, etc. In addition, it can also solve many problems that would be difficult or even impossible by conventional methodology. Generally speaking, there are two typical problems solved by neural networks, namely Classification and Regression. The essential difference between the two functions is the type of data output by the neural network. Specifically, output discrete category labels or specific continuous values correspond to classification and regression respectively.[38]

The basic principle of neutral network is a mathematical model, which is obtained by nesting many functions into each other. A single neuron will receive signals from other neurons (X_1-X_n, n is the number of input data), and these signals will be weighted (W_{1j}-W_{nj}, j is the sequence number of the neuron in this layer) at the same time. The input total superimposed value will be compared with the threshold of this neuron (θ_j), and then the neuron will output a value of y_j (shown in Fig. S7), which depends on the activation function.[39]

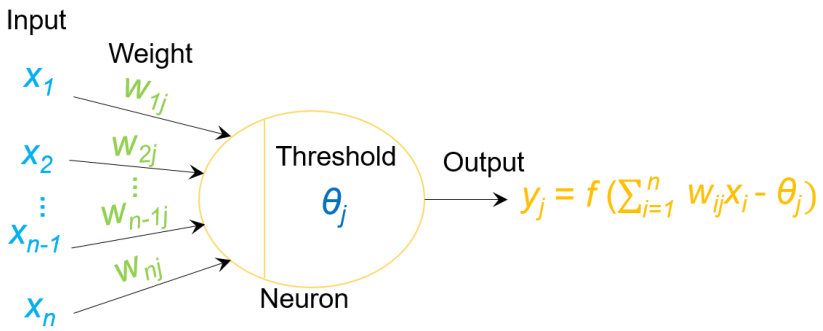

Fig. S7 A typical neuron model.

Similarly, the output of this neuron also operates the input of the neuron in the next layer.

In supervised learning, the neural network adjusts the weight of the network according to the difference between output at each time and the expected output result, so as to make the final output infinitely close to the expected value. In this paper, we apply the classification and regression in the neural network to the optimal design of the metasurface element and to solve the crosstalk problems mentioned above.

In the proposed inverse design model, we use a classifier to quickly divide all kinds of propagation phase-length correspondences (the number of species is denoted as N) with different crosstalk from the high-frequency structure into specified categories (the number of categories is denoted as T), so that the accurate propagation phase and length can be searched and input into the adaptive optimization in time. The classifier shown in Fig. S8(a) is based on back propagation (BP) neural network with 3 hidden layers (30 neurons per layer).

For a 2-bit quad-channel coding metasurface, N and T are 32 and 12, respectively. The determination of N and T will be discussed in Sec. S10. The input of the classifier is two phase coding in high frequency and a label of polarization (1 for x-LP and 0 for y-LP), and their detailed implication is shown in Fig. S10.

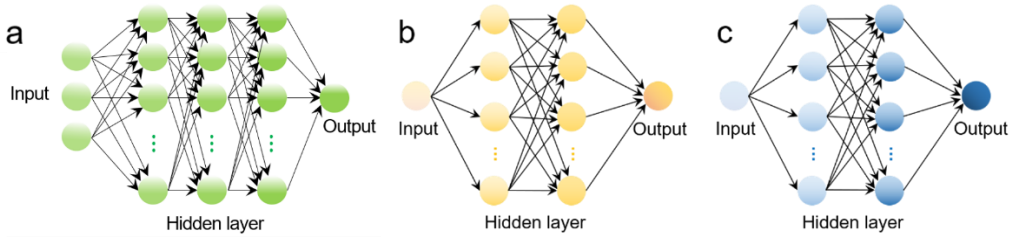

Fig. S8 Three structures of the neural network model used in this paper. (a) A classifier based on BP neural network. (b) A multi-layer perceptron with 10 neurons in each hidden layer, in which the input and output are the length and propagation phase, respectively. (c) The same neural network structure as (b), but the propagation phase and length are used as input and output respectively.

Neural network models shown in Figs. S8(b) and (c) are ANN models with the same number of hidden layer and neuron, both of them have 2 hidden layers with 20 neurons per layer. Both of them are typical Multilayer Perceptron (MLP) models. The purpose of the MLP

model in Fig. S8(b) is to predict the propagation phase with different crosstalk in milliseconds based on the input of length (l_{x1} or l_{y1}). and quickly predict the length based on the propagation phase. The model in Fig. S8(c) is trained by exchanging the input and output of the neural network in Fig. S8(b). The purpose of training these two kinds of MLP models is to achieve the transformation between length and propagation phase (at low frequency) under various crosstalk conditions on the millisecond scale during the adaptive optimization process.

S8. More Training Details of the Inverse Design Model

For more training details of the inverse design model, we present the details of each part in the following table.

Table S2. Learning rate and optimizer of the inverse design model

Name	Learning rate	Optimizer
Classifier	0.01	Adam
MLP	0.001	Adam
Iterative algorithm	0.5	Prior-knowledge-guided

Unlike the learning rate in the general case of neural network models (0.001-0.01), the learning rate in the iterative algorithm is calculated as the ratio between the actual parameter change and the theoretical parameter change for each iteration. It is set to 0.5 because too large value will affect convergence, while too small value will lead to an increase in the number of iterations and thus increase the time cost. Besides, the "optimizer" in this part means that our parameter iteration process is calculated based on the current error and electromagnetic prior-knowledge.

The training loss, accuracy and error histogram distribution of the neural networks are specifically presented in Fig. S9.

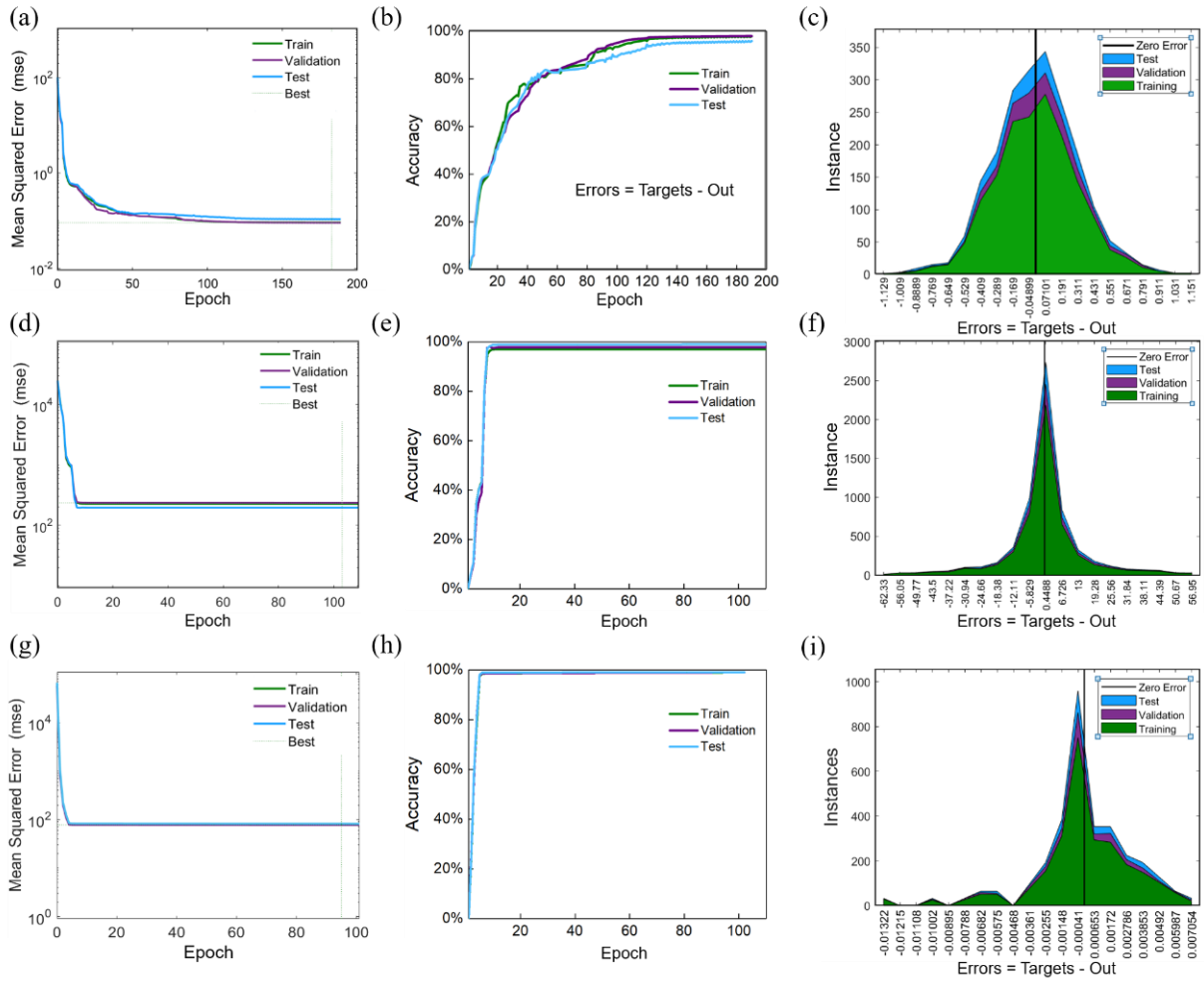

Fig. S9. Details of the trained neural network models. (a) Training loss, (b) accuracy and (c) error histogram distribution of the classifier. (d) Training loss, (e) accuracy and (f) error histogram distribution of the MLP model for propagation phase prediction. (g) Training loss, (h) accuracy and (i) error histogram distribution of the MLP model for l_{x2} (l_{y2}) prediction.

Other details about the time cost, model size, and result precision are shown in the Table S3 below.

Table S3. Time cost, model size and result precision of the inverse design model

Prediction time	Time cost	Model size	Result precision
Time for classifier to output	0.036 s	12 KB	97%
Time for MLP to output	0.022 s	20 KB	98%
Adaptive Iteration	2~10 min	18 KB	92%

S9. The Specific Implementation Process of the Adaptive Optimization Assisted by the Deep Learning

The trained MLP models in Figs. S8(b) and (c) are expressed as the following forms to intuitively show that the mapping relationship between the length and the propagation phase at low frequency:

$$\varphi_{x1} = \text{length2phase}(l_{x1}), \tag{S9}$$

$$l_{x1}' = \text{phase2length}(\varphi_{x1}'). \tag{S10}$$

Where the function "length2phase" means "from length to phase" that the length l_{x1} (l_{x1}', l_{y1}, l_{y1}') is the independent variable while the phase response φ_{x1} (φ_{x1}', φ_{y1}, φ_{y1}') is the dependent variable. Similarly, the function "phase2length" means "from length to phase". In the above equations, we take *x*-LP coefficient as an example, but the they are also suitable for *y*-LP coefficient. Specifically, the two kinds of nonlinear mapping (Equations S9-S10) are applied to realize the conversion between the EM response (Equations S4-S8) and the physical parameters.

As explained in Sec. S5, in prior knowledge model, the variables $\{\alpha, \varphi_{x1}, \varphi_{y1}\}$ need to be continuously replaced with $\{\alpha', \varphi_{x1}', \varphi_{y1}'\}$ to gradually approach the phase shifts for RCP and LCP at low frequency. Therefore, how to choose suitable and accurate propagation phase-length functions is fundamentally important, which is also the key task of the following.

As shown in Fig. S10(a), the high-frequency structures can be directly designed based on the required phases of RCP (φ_{rcp2}) and LCP (φ_{lcp2}) at high frequency. To illustrate the design process, we take three examples of high-frequency structures ($\varphi_{rcp2}/\varphi_{lcp2}$ = 0°/0°, 180°/180° and 180°/270°, respectively) in the inset of the bottom panels in Fig. S10(a). The crosstalk in low-frequency channel caused by the different high-frequency structures is also different. As mentioned in the main text, the large Jerusalem-cross-shaped resonator is separated from the four small resonators by a layer of dielectric and it operates as the low-frequency structure. Due to the anisotropy of the high-frequency structures in the *x*- and *y*-directions, the propagation phase-length functions at low frequency also needs to be discussed separately in these two orthogonal directions. Figure S10(b) illustrates the process of picking up the suitable low-

frequency propagation phase-length functions. For example, if the current high-frequency phase requirements are $\varphi_{rcp2}/\varphi_{lcp2}$ = 180°/270°, and we want to select suitable propagation phase-length functions for the low-frequency *x*-LP wave, then a 1×3 array of [2 3 1] should be input to the neural-network-based classifier. The first two codes in the array are the 2-bit phase coding for φ_{rcp2} and φ_{lcp2}, and the third one is the label for the polarization (1 for *x*-LP and 0 for *y*-LP). And if the suitable functions for the low-frequency *y*-LP wave are required, the input 1×3 array should be tuned to [2 3 0]. It is because that the first two numbers in the 1×3 array are the phase codes for high-frequency RCP and LCP, and the third number is the label of low-frequency LP wave (mentioned in Sec. 6). Once the array is input to the classifier, a classification result (one of the *T* types) will be output and each type corresponds to a pair of length2phase and phase2length functions, which are pretrained MLP models and support fast implementation of nonlinear mapping between length and propagation phase. In other words, the predictions from $\{\varphi_{x1}', \varphi_{y1}'\}$ to $\{l_{x1}', l_{y1}'\}$ or the prediction from $\{l_{x1}', l_{y1}'\}$ to $\{\varphi_{x1}', \varphi_{y1}'\}$ under the crosstalk caused by the current high-frequency structures can be quickly obtained in the order of milliseconds.

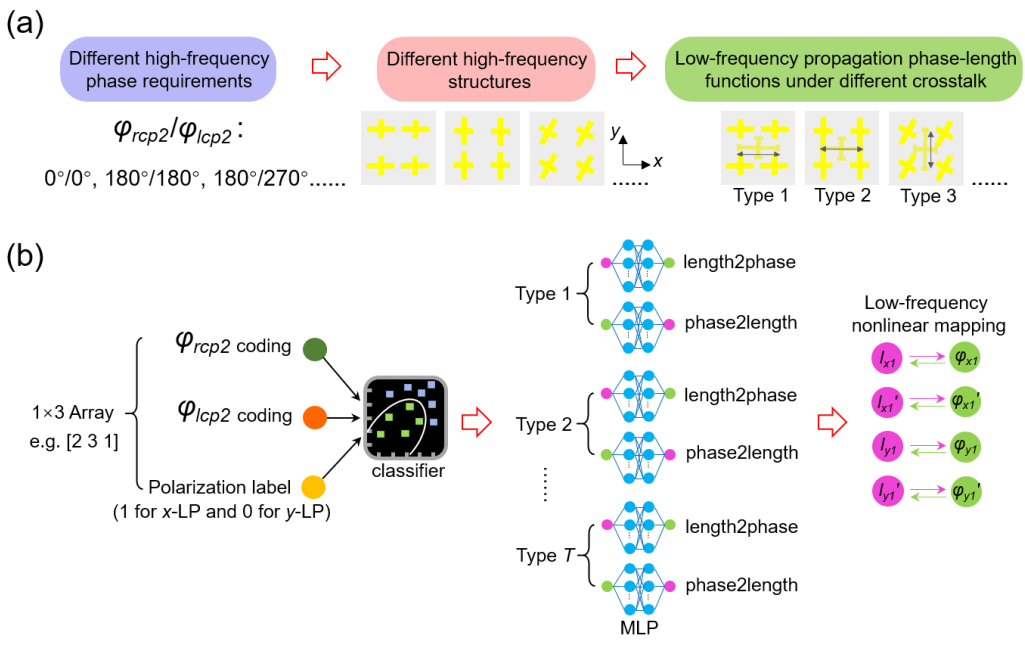

Fig. S10 (a) The schematic of the crosstalk in low-frequency channel caused by the different high-

frequency structures. (b) Process of the selecting suitable low-frequency MLP models based on the current high-frequency phase requirements.

Based on the above process, the three low-frequency parameters $\{\alpha, l_{x1}, l_{y1}\}$ will be updated in the next iteration calculation to generate new three parameters $\{\alpha', l_{x1}', l_{y1}'\}$ according to the current phase error, so as to get closer to the target until the modulation phase requirements are met. In order to hand over the entire task of inverse design, we use commercial software MATLAB and CST MWS to perform parameter replacement and the simulation of phase modulation, respectively. The specific flow chart is shown in Fig. S11.

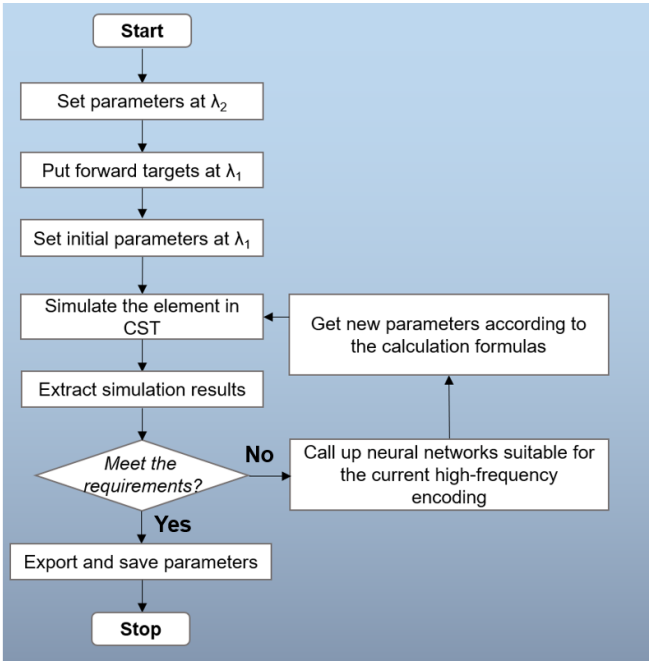

Fig. S11 Overall flow chart of the deep-learning-assisted inverse design method.

S10. 12 Kinds of Propagation Phase-Length Mapping

As mentioned in S7, there are T types propagation phase-length mapping relationships for a 2-bit quad-channel coding metasurface. In this section, the conclusion of $T = 12$ is discussed in detail.

In the main text, the high-frequency structures are defined by three physical parameters: l_{x2}, l_{y2} and β. l_{x2} and l_{y2} are the lengths of the high-frequency metallic arms in x and y directions, respectively. β is the rotation angle of the high-frequency resonators. For a metasurface with 2-

bit discrete phase modulations, there are 16 kinds of high-frequency structures, as shown in Fig. S12(a). Obviously, the crosstalk caused by these different high-frequency structures is also different for the low-frequency propagation phase-length functions. As shown in Fig. S12(b), "$C_1/C_2/C_3/C_4$" represent the 2-bit phase coding corresponding to four phase shifts, which are RCP reflection phase at f_2 (or high frequency), LCP reflection phase at f_2, RCP reflection phase at f_1 (or low frequency), and LCP reflection phase at f_1. The large Jerusalem-cross-shaped resonator is separated from the four small resonators by a layer of dielectric and it operates as the low-frequency structure. When the length of the its metallic arm (in x or y direction) varies, the change of the propagation phase (x-LP or y-LP waves) in the corresponding direction is different for these sixteen elements. Moreover, for each element in Fig. S12(b), the propagation phase shifts for x-LP or y-LP waves are also different when the variations of l_{x1} and l_{y1} are the same. Specifically, the anisotropy is reflected not only in the high-frequency structures, but also in their crosstalk to the low frequency. Based on this analysis, the propagation phase-length functions for x-LP and y-LP should be characterized separately. We take the element encoded with $C_1/C_2 = 0/1$ as an example, there is an obvious difference between its propagation phase-length functions for x-LP and y-LP waves (Figs. S12(c) and (d)). In conclusion, the propagation phase-length functions of x-LP and y-LP waves for the sixteen elements in Fig. S12(b) should be characterized separately, which means that there are $16\times2 = 32$ kinds of functions (corresponding to $N = 32$ mentioned in Sec. 7).

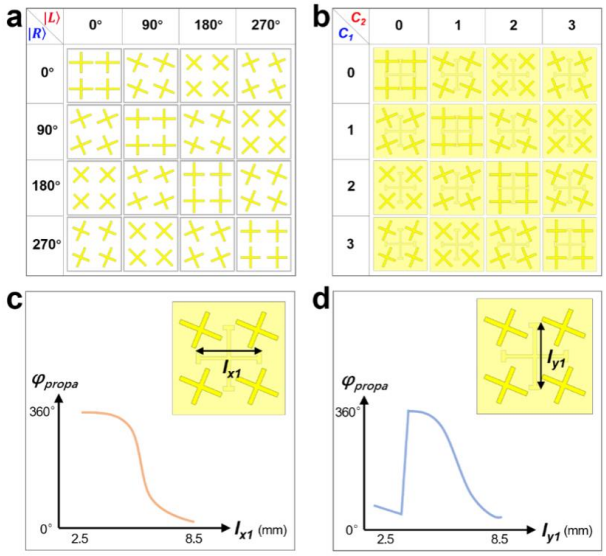

Fig. S12 (a) 16 kinds of designed high-frequency structures for 2-bit coding metasurface with independent phase modulation for two CP waves. (b) Mapping relationship between the coding "C_1/C_2" of high frequency phase modulation and structures of elements, where the low-frequency resonators have not been stretched and rotated. For the element encoded with $C_1/C_2 = 0/1$, propagation phase-length functions for (c) *x*-LP and (d) *y*-LP waves, respectively, where φ_{propa} is the propagation phase for *x*-LP or *y*-LP waves.

Due to the existence of structural symmetry, some of these 32 functions are the same. Actually, it can be simplified to 12 different functions. In other words, it is only necessary to classify these 32 cases into 12 types (corresponding to $T = 12$ mentioned in Sec. 6), and each type uses its own propagation phase-length function. The results of classifying these 32 cases into 12 types are listed in Fig. S13, where the coding marked in each element consists of the high-frequency phase coding (the first two numbers) and the low-frequency polarization (the third character). The coding, for example, "23x", denoted on the element represents that when the high-frequency structures are set as $\varphi_{rcp2} = 180°$, $\varphi_{lcp2} = 270°$ (2-bit phase coding "2/3"), the variation of the arm length is designed to realize different low-frequency propagation phases for *x*-LP incidence. For the classification, we take the type 1 as an example. The second element of type 1 can be obtained by rotating the first one with an angle of 90° if the arms (denoted by double arrows) are with same value. Thus, we just need to evaluate the EM performances of the first element and then through a rotating operation to obtain the results for the second one. This

can be verified by the propagation phases of the two elements with coding of "00x" and "22y", as seen in Figs. S14(a) and S14(b), where the two functions between the propagation phase and length for them are the same. Similar, for type 3, the first element can be mirror-flipped into the second one, and propagation phase-length functions for these two elements are also the same (verified in Figs. S14(c) and S14(d)). For elements that have both the rotation and mirror relationships, they can also be classified into the same type based on the above analysis, such as the first and fourth elements in types. With this classification, the calculations could be drastically reduced.

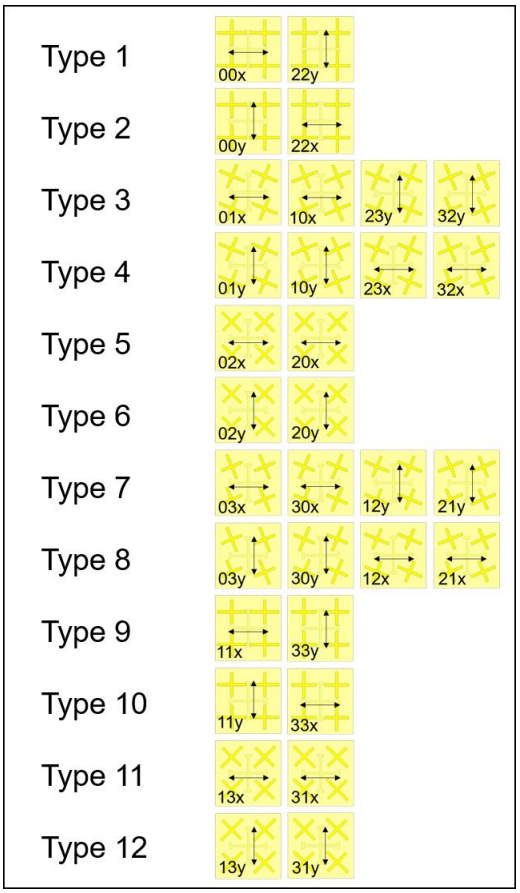

Fig. S13 32 kinds (16 for *x*-LP and 16 for *y*-LP) of propagation phase-length correspondences can be divided into 12 types according to the geometric symmetry.

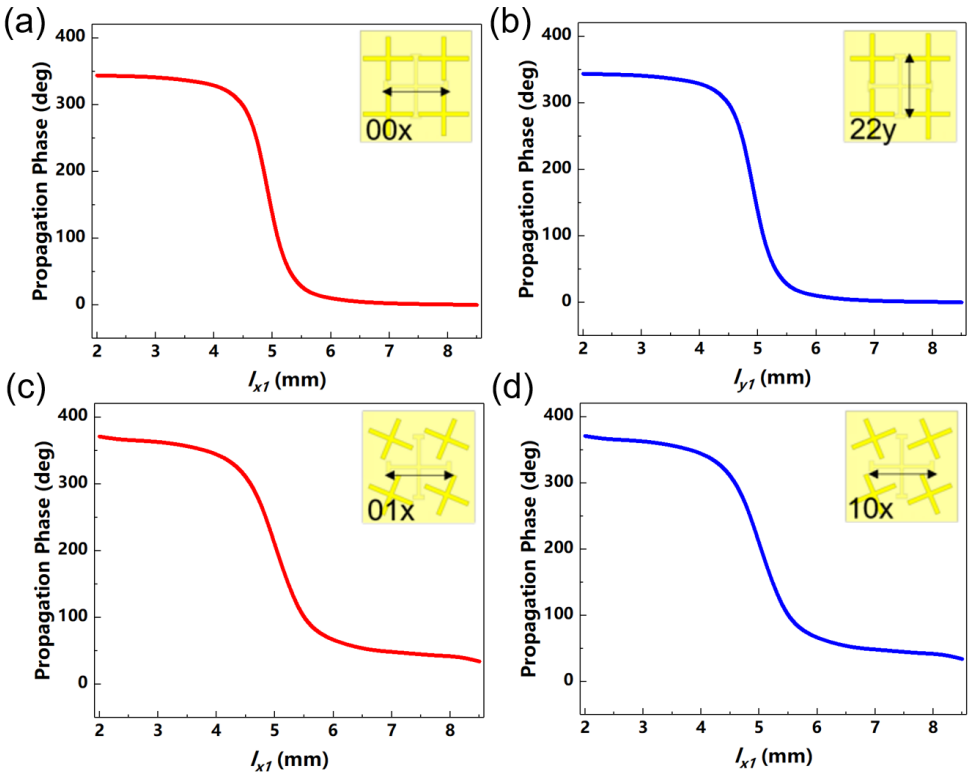

Fig. S14 Low-frequency propagation phase versus length of metallic arm in the corresponding direction. (a) For element encoded with "00x". (a) For element encoded with "22y". (a) For element encoded with "01x". (a) For element encoded with "10x".

The twelve types of propagation phase-length functions are shown in the Fig. S15. In the process of the inverse design, the specific task of classification and characterizing the propagation phase-length functions are implemented in the millisecond scale by the pretrained classifier and MLPs (mentioned in Sec. S6), respectively.

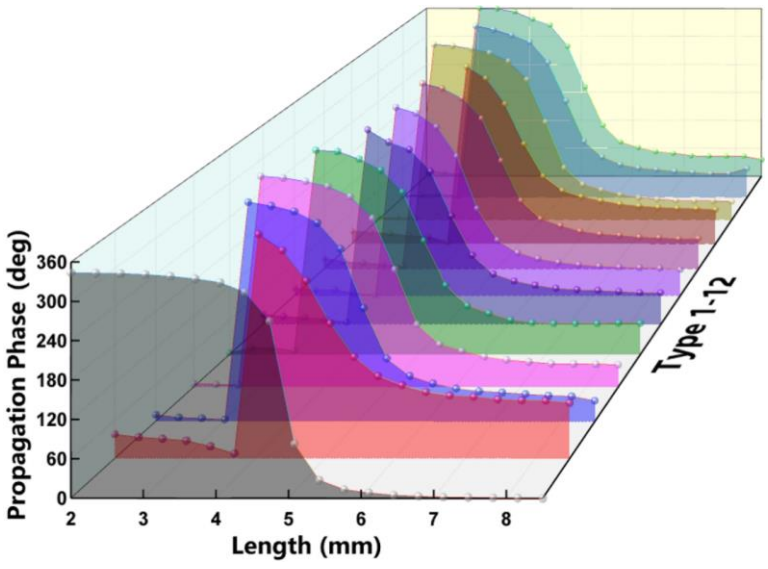

Fig. S15 Propagation phase-length functions at low frequency of the 12 types, where the length refers to l_{x1} or l_{y1}.

S11. Amplitude Analysis of Optimized 256 (4^4) Types of Elements

The reflection amplitude is important for the metasurface design. Herein, we offer the amplitude of total optimized 256 elements in four CP channels. These elements are used for designing the metasurfaces in the main text. The high-frequency (f_2) performance of the elements are all above 0.95 (shown in Figs. S16(a) and S16(b)). For low frequency operation, 90% of the elements have reflection amplitude higher than 0.9 (Figs. S16(c) and S16(d)).

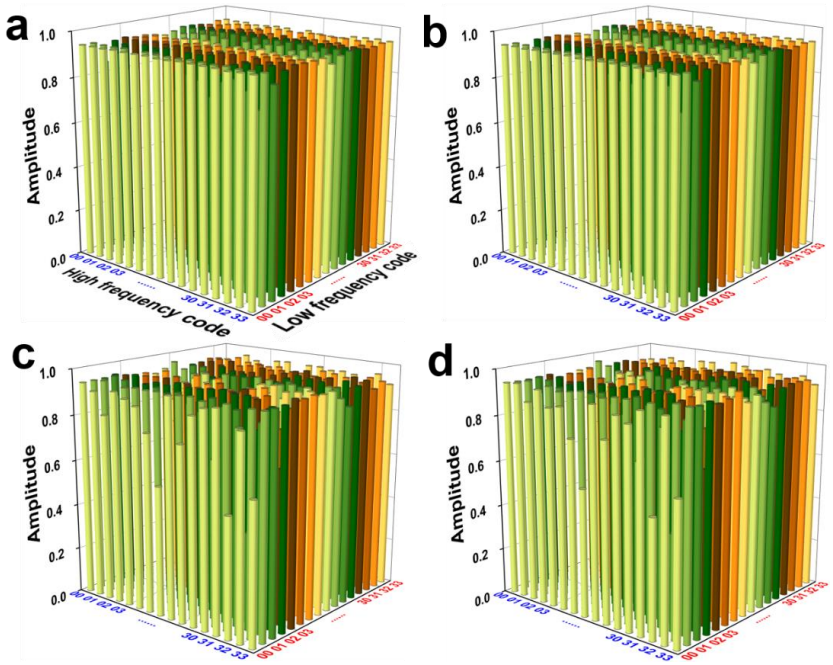

Fig. S16 Amplitude performance of the 256 kinds of elements used for designing metasurfaces. (a) LCP channel at f_2 or high frequency. (b) RCP channel at f_2 or high frequency, (c) LCP at f_1 and (d) RCP-RCP at f_1 (low frequency). The high frequency code and low frequency code correspond to C_1/C_2 and C_3/C_4, respectively.

S12. The Design Process of the Off-Axis Vortex Beams

Figure S17 illustrates the detailed design process of the phase distributions for the off-axis vortex beams in Fig. 3. Firstly, the spiral phase for orbital angular momentum (OAM) generation with different topological charges are calculated and discretized to 2-bit states. Secondly, one-dimensional phase gradient in different directions are calculated according to the

required deflection angle and generalized Snell's law.[25] And then, the discretized spiral phase and phase gradient are added together to obtain the final phase distributions.

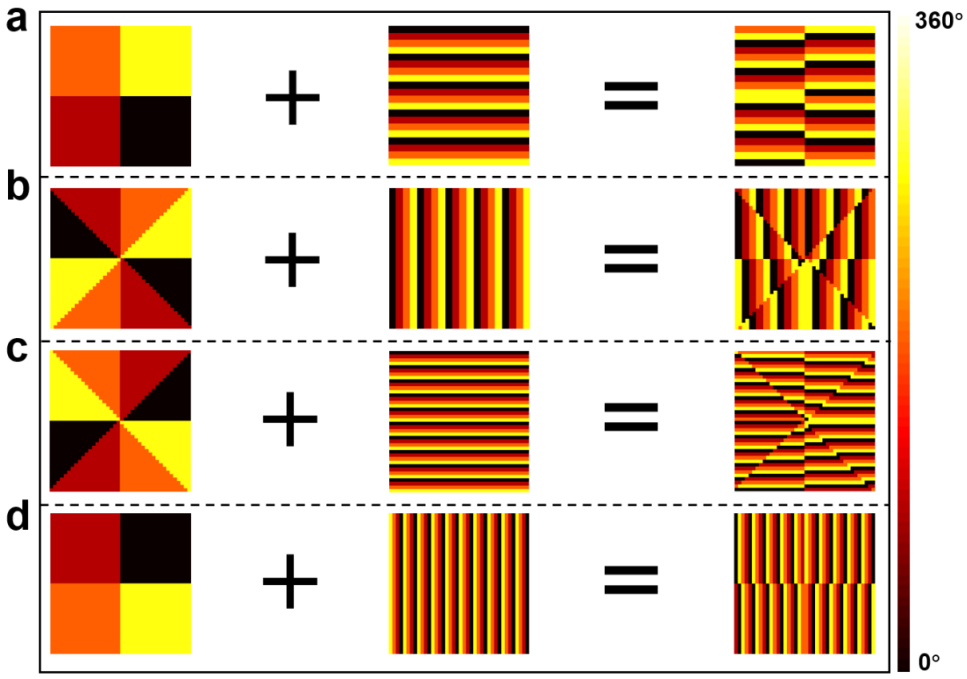

Fig. S17 Design process of the off-axis vortex beams in different deflection angles for channel 1 (a), channel 2 (b), channel 3 (c) and channel 4 (d). The sign of "+" represents the operation of direct addition.

S13. Effect of the Discretized Phase Profile on the Intensity Distribution

For the imperfect donut-shaped intensity distribution of OAM ±1, it is due to the discretization of the continuous spiral phase profiles. As shown in Fig. S18, continuous and discrete phase profiles theoretically produce circular and square intensity distributions, respectively.

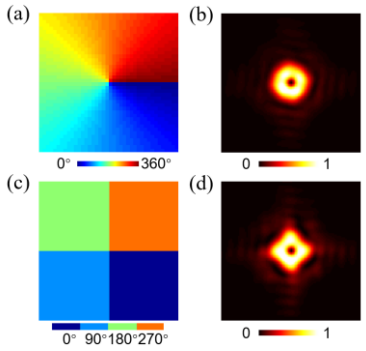

Fig. S18 Theoretical analysis about the intensity distribution of OAM beams generated by different spiral phase profiles. (a) Continuous spiral phase profile and (b) the generated intensity distribution. (c) Discretized spiral phase profile and (d) the generated intensity distribution.

S14. Advantages of Spin-Decoupled Element for Vector Vortex Beams Generation

Vector vortex beams (VVBs) consist of azimuthal and radial polarized cylindrical vector beams. Different from the classical optical vortices, it can hardly describe the wavefront of the VVBs by a single state of polarization (SOP). Because VVBs are the superpositions of OAM with the opposite spin states. When both RCP and LCP waves have typical spiral phase distribution at the same time, the phase difference that changes with spatial distribution will cause the two CP waves to synthesize different polarization states. Poincaré sphere is used to describe polarization, which is one of the most important features of light, but it is limited when face the beams with inhomogeneous SOPs. Fortunately, the emergence of higher-order Poincaré sphere (HOPS) overcome this restriction.[1] In $HOPS_{m,n}$, the north pole and south pole represent RCP and LCP carrying OAM modes of m and n, respectively. PB phase based metasurface produces opposite OAM modes ($m = -n$) under the illumination of RCP and LCP. Herein, we use reflective metasurface as an example. As shown in Fig. S19(a), if output vortex beams are along the z axis, then only VVB with the wave function of $HOPS_{l,-l}$ can be generated. Where l is the topological charge, ϕ is the azimuth angle. Additionally, the number of lobes ($|m - n| = |2l|$) in the power pattern must be an even number. If OAM beams generated by RCP and LCP are not along z axis, the characteristic of axis symmetry will make them unable to coherently superimpose. $m \neq n$ also can be obtained by adding two axisymmetric deflection angles[70] to the two output OAM beams in RCP (LCP) channel (shown in Fig. S19(b), the deflection angle is denoted as θ, A and B represent the amplitude coefficients, γ_i represents the phase of the beams). Obviously, there is also a coupling relationship between the output VVBs, which can be described as that $HOPS_{l1,l2}$ and $HOPS_{-l1,-l2}$ are generated in pairs and symmetric about the z axis.[73] Therefore, PB phase metasurface only produces off-axis VVBs *in pairs and symmetric* about the z axis.

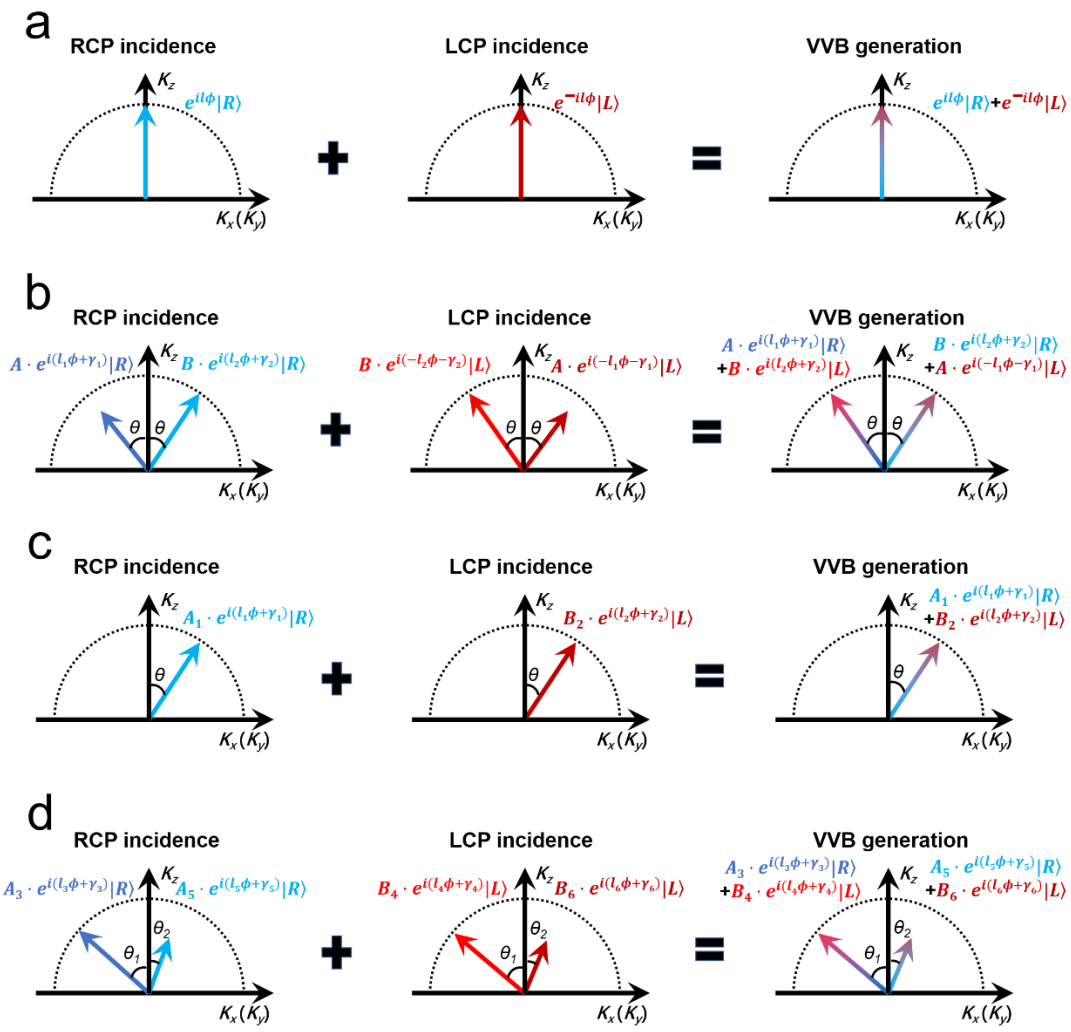

Fig. S19 Schematics illustrating the principle of (a) single along-axis VVB, (b) multi off-axis (angle symmetry) VVBs by PB phase metasurface. (c) Single off-axis VVB, (d) multi off-axis (angle asymmetry) VVBs by spin-decoupled metasurface.

The combination of PB phase and propagation phase make it possible to independently control the phase of two CP channels.[29] Similarly, we still discuss the single beam and multi beams separately. When the element satisfies the spin-decoupled condition,[76] single VVB can be designed towards an arbitrary direction(shown in Fig. S19(c)), not limited to be along z axis as that conventionally done by PB phase metasurface, because RCP and LCP channels are able to generate OAM in the same off-axis direction. For multi beams generation, $m \neq n$ and an arbitrary deflection angle ($\theta_1 \neq \theta_2$) can be met at the same time (Fig. S19(d)). More importantly, unlike PB phase metasurface (the ratio of the energy occupied by two VVBs in a

pair must be 1: 1), it is more flexible to modulate energy allocation of the multi-VVBs by the proposed method. In a conclusion, the generation of VVBs by spin-decoupled metasurface will offer relatively more freedom in direction, mode, and energy allocation than PB phase metasurafce.

S15. Discussion about the Imperfection of the Measured Results

The three main reasons for the unevenness of partial measured results are as follows:

(1) Complexity of near-field measurement in microwave region. The near-field measurement in the microwave region requires point-by-point detection, which takes about few hours to stably map a picture of the near field distribution. Besides, it's even more difficult for reflection measurement. Because we need to subtract the incident field from the directly measured field data. If there is the slight deviation in the spatial position of each point during the reflected field measurement and incident field measurement, the final measured results will be seriously affected.

(2) The imperfect detection probes. A series of home-made (not commercially available) helical antennas are used in the near-field detection. The unavoidable fabrication errors of these probes also have impacts on the final electric distributions.

(3) The fabrication error of the sample. Limited by the tolerances in the PCB technique, the final measured results are slightly affected.

S16. Large-Capacity Vortex beams Generation

Frequency can provide new degree of freedom to further increase the channel capacity for metasurface. Herein, we further design a large channel capacity method based on the proposed metasurface and the design method. We designed a 16-channel off-axis vortices multiplexer according to these parameter distributions (shown in Fig. S20). There are four vortex beams carrying OAM are generated in each CP channel. Since the 16 vortex beams are designed in 8

different directions, they are superimposed coherently into 8 HOPS beams, denoted as #1 to #8.

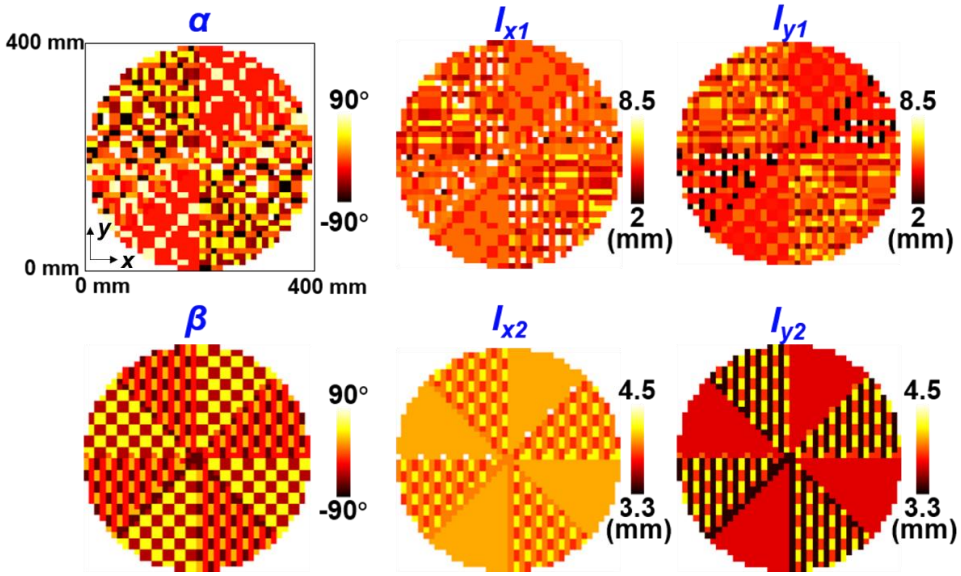

Fig. S20 Distributions of six physical parameters of the vortex meta-multiplexer.

Figure S21 shows the details about the polarization states, power patterns, and radiation directions for the generated vortex beams at low frequency and high frequency. When the vortex beams multiplexer is illuminated by *x*-LP plane wave at 9.8 GHz, $HOPS_{1,1}$ (#1 and #2, $m = n$) and $HOPS_{1,-1}$ (#3 and #4, $m \neq n$) are produced in different directions. $HOPS_{2,2}$ (#5 and #7, $m = n$) and $HOPS_{-2,2}$ (#6 and #8, $m \neq n$) are the output vortex beams when the frequency of the incident wave is tuned to 17.8 GHz. The inhomogeneous states of polarization can be clearly distinguished from Fig. S21(a) for $m \neq n$, as well as the uniform state for $m = n$.

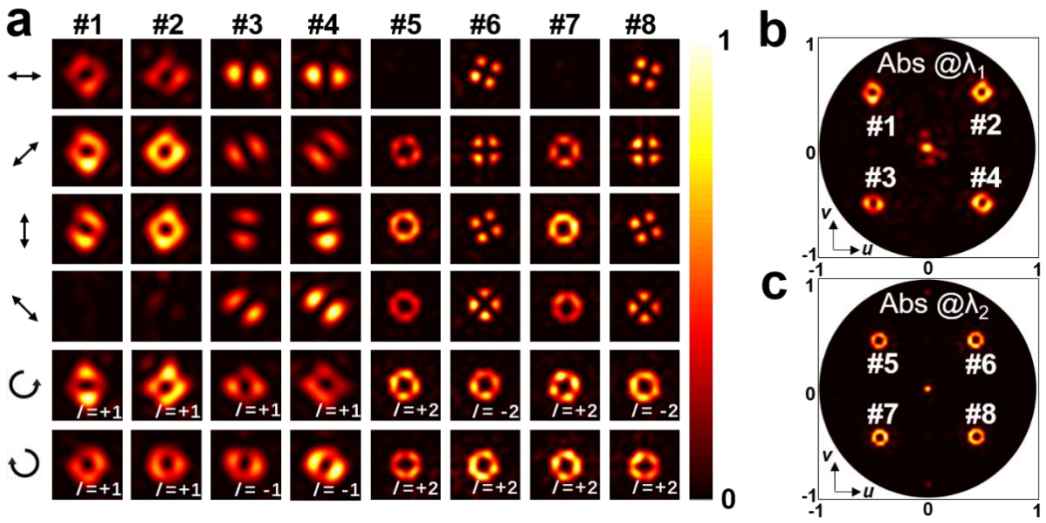

Fig. S21 Large-capacity vortex generation with the illumination of *x*-LP plane waves at 9.8 GHz and 17.8 GHz. (a) Simulated normalized power pattern of 6 polarization components of these 8 HOPS beams. (b) Normalized total power pattern of the vortices in momentum space.

S17. Another example of Arbitrary Spin-to-Orbit Conversion

In the main text, we have shown the simulated and measured results of the prototype with l = +1, +2, -1, and -2. We present here another design example (shown in Fig. S22) and the corresponding simulated results (shown in Fig. S22) of spin-to-orbit meta-convertor. And the transformation it performs can be described as

$$\left|R, f_1\right\rangle \rightarrow e^{+1\cdot i\phi}\left|R, f_1\right\rangle, \tag{S11}$$

$$\left|L, f_1\right\rangle \rightarrow e^{-2\cdot i\phi}\left|L, f_1\right\rangle, \tag{S12}$$

$$\left|R, f_2\right\rangle \rightarrow e^{+3\cdot i\phi}\left|R, f_2\right\rangle, \tag{S13}$$

$$\left|L, f_2\right\rangle \rightarrow e^{-4\cdot i\phi}\left|L, f_2\right\rangle. \tag{S14}$$

Where R and L are the polarization states of the incident and reflected waves (R for RCP and L for LCP).

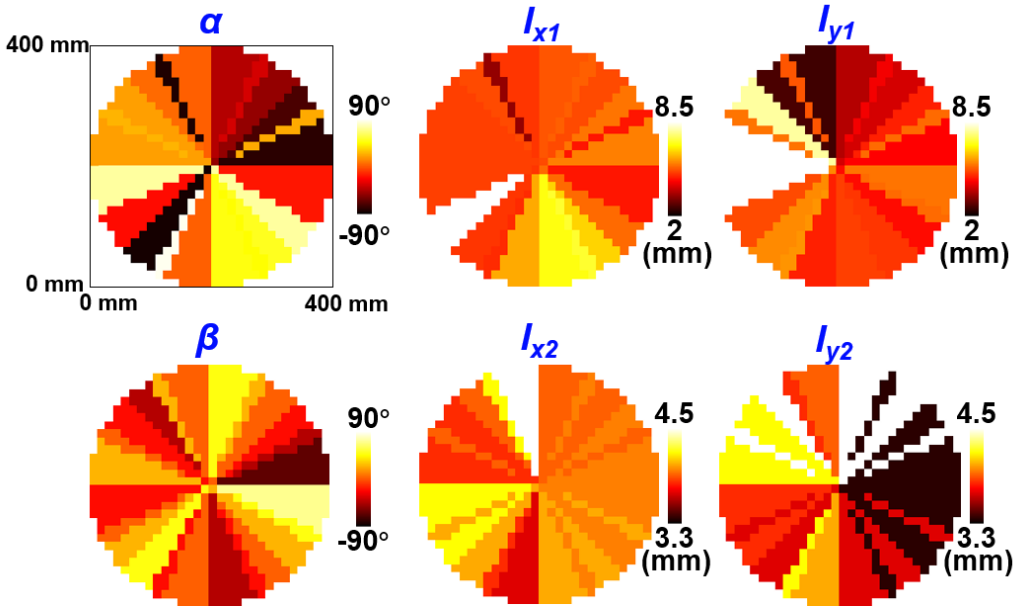

Fig. S22 Distributions of six physical parameters of the spin-to-orbit meta-convertor for l = 1, -2, 3, and -4.

According to the parameter distributions shown in Fig. S23, we simulated a 400 mm×

400 mm metasurface sample, and obtained the OAM modes with the specified topological charges generated under the incidence of RCP and LCP waves at f_1 and f_2.

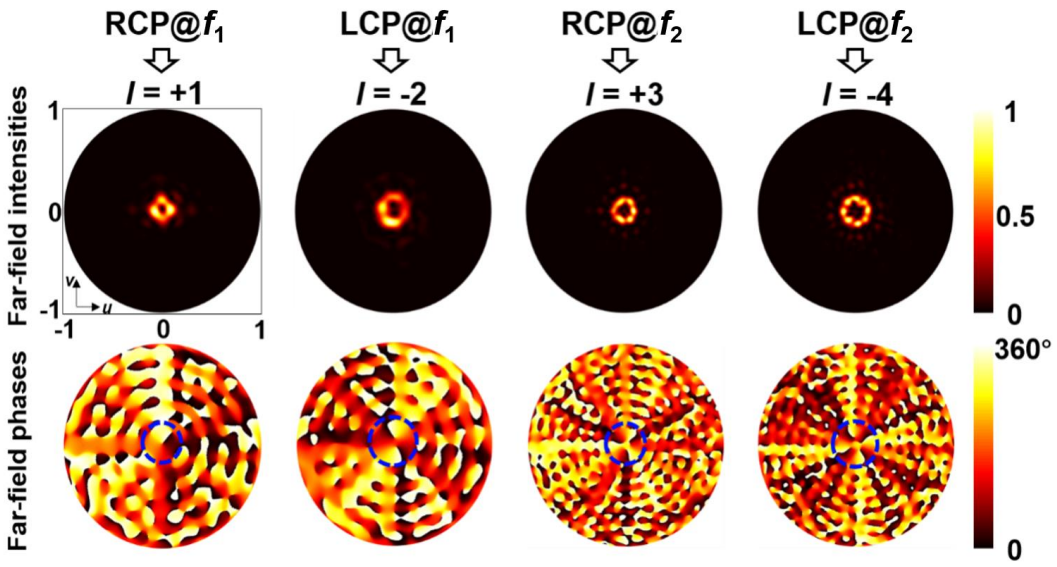

Fig. S23 Normalized simulated far-field intensity patterns and phases for the four reflected CP channels at 9.8 GHz and 17.8 GHz.

S18. Vortex Beam Generation for Oblique Incidence

A phase gradient is formed when the plane wave is incident obliquely to the plane where the metasurface is located. As the design example, we consider the case of 24° oblique incidence. The theoretical phase gradient due to the oblique incidence is shown in Fig. S24. In fact, by pre-applying another opposite phase gradient in each phase channel, the uneven initial phase distribution caused by obliquely incident EM waves can be compensated to planar one.

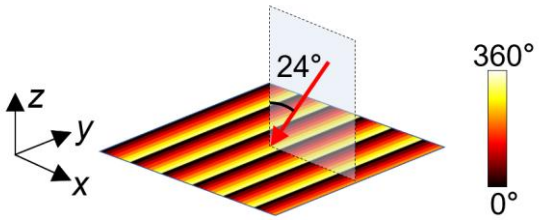

Fig. S24 Phase gradient caused by the obliquely incident plane wave.

The first row in Fig. S25 shows the schematic diagram of the working principle of the vortex multiplexer described in the main text, that is, in the case of normal incidence. The schematic shown in the second row is our working mode under oblique incidence, generating four vortex beams with the same orientation and topological charge as the first row. For each circular

polarization channel, we can obtain the required phase distribution for oblique incidence operating mode by adding the theoretical phase distribution in the main text and the opposite of the phase gradient shown in Figure S24. The four phase distributions shown in the third row are calculated in this way, and further we perform the full-wave simulation of the quad-channel metasurface. The simulated far-field scatter patterns (shown in the fourth row) in these four channels are in good agreement with the results shown in Figure 3, demonstrating the feasibility of generating these vortex beams at oblique incidence.

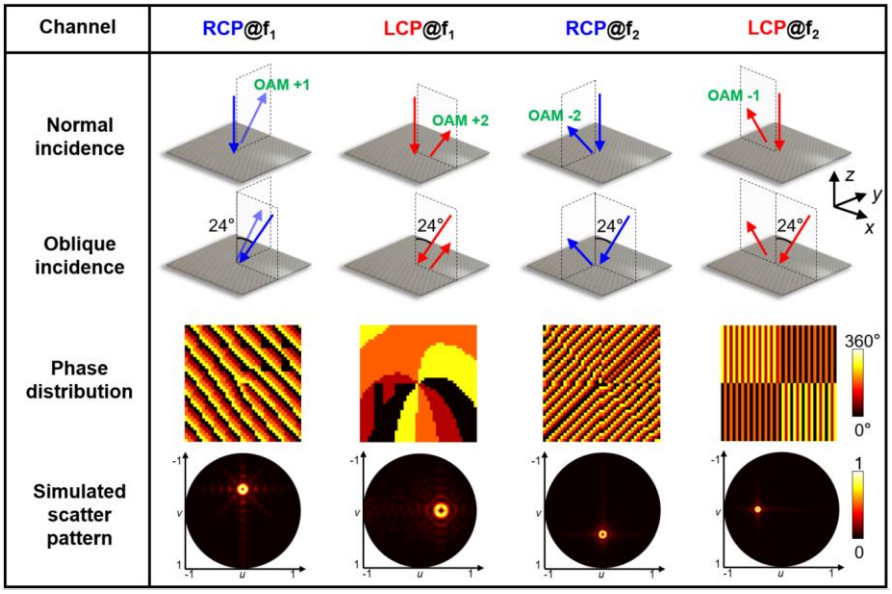

Fig. S25 Design of the vortex beam multiplexer operating in the case of oblique incidence.

S19. Comparison of the Design Method and Traditional Optimization Algorithms

The conventional genetic algorithm (GA), particle swarm optimization (PSO), etc. can also realize the on-demand design of metasurface, but they will have higher time cost than the approach proposed in the main text.

These conventional methods often require dozens or hundreds of parameter iterations and simulations, they are usually suitable for metasurfaces with periodic property to find optimal EM performance. For example, in the design of broadband absorbers, only one meta-atom structure needs to be optimized and then it is periodically extended to complete the optimal design of the overall metasurface. But these methods are not applicable to the design of a series meta-atoms with different phase responses. If we want to optimize the quad-channel metasurface with 2-bit phase coding (mentioned in the main text), 256 different meta-atoms

need to be optimized via these methods. It will take too much time to execute 256 times of GA or PSO. In particular, each execution process also includes dozens or even hundreds of full-wave simulations, which means that the total time cost can be days or weeks.

S20. Near-Field Measurement Setup for the Prototypes

As shown in Fig. S26, the prototype to be measured is surrounded by the wave-absorbing materials. Fig. S26(a) shows the photograph of the CP transmitting antenna to produce a qusi-plane wave at the sample surface. We use the CP antenna for illumination when measuring the spin-to-orbit meta-convertor. When it comes to the measurement of VVBs meta-multiplexer, the transmitting antenna is replaced with a LP antenna which is not shown in the photograph.

The distance between scanning plane and the prototype is 500 mm. Theoretically, the cone angle of the vortex beam at high frequency will be smaller than the that at low frequency. However, when the distance between the scanning plane and the prototype is fixed, the equivalent distance calculated as operating frequency becomes larger compared to that of low frequency band. Hence, the beams in high frequency spread out more obvious. As a result, there is slight difference between the radius of the donut distribution at 9.8 GHz and 17.8 GHz measured in the scanning plane. In fact, different from transmitted field, the near-field scanning measurement of the reflected field is more likely to suffer from interference from the emitting antenna. In order to solve the problem, on the one hand, we wrap the wave-absorbing material on the left side of the transmitting antenna without affecting the work of the emitting antenna. On the other hand, after directly measuring the reflected field, we then subtract the incident field from it. This method can alleviate the ripple-like interference in the measured results.

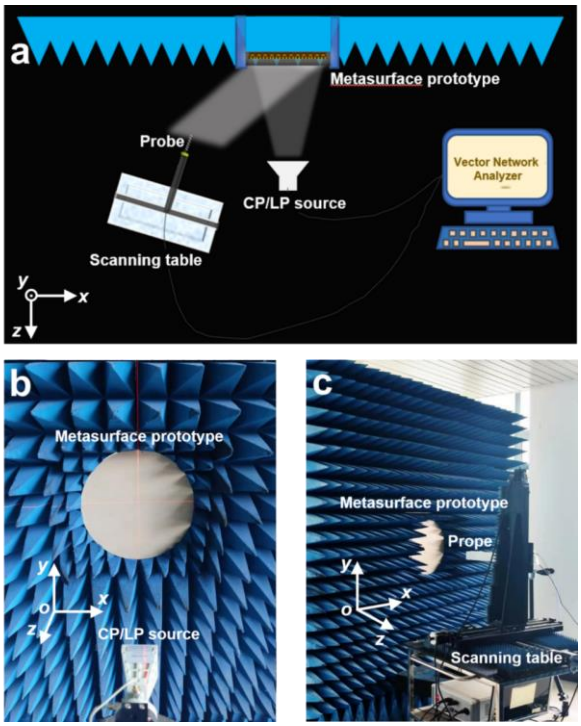

Fig. S26 (a) Near-field measurement setup of the metasurface prototypes. The photograph of (b) EM wave excitation and (c) reflection field measurement.