A Review of Metasurfaces: Principles and Applications

Jeffrey Wei[1, *]

[1] *Department of Physics, Yale University*

(Dated: December 17, 2025)

The miniaturization of optical systems is currently limited by the volumetric nature of refractive lenses. Metasurfaces overcome this barrier by encoding optical functionality into ultrathin, subwavelength arrays [1, 2]. This review establishes the theoretical foundations of flat optics, contrasting classical phase accumulation with the generalized laws of refraction and scalar diffraction theory. We then detail a hierarchical design framework, moving from the physics of meta-atom scattering to system-level optimization. Key applications are surveyed, including compact imaging, holographic displays, and polarization sensing. Finally, we discuss critical bottlenecks in efficiency and bandwidth, highlighting the emerging role of data-driven design. Ultimately, these advances position metasurfaces as a scalable, industrial-grade technology set to revolutionaize consumer products and the optical research frontier.

I. INTRODUCTION

For centuries, the quest to image, manipulate, and control the behavior of light has driven advances in optics, from resolving the smallest structures in nature to engineering complex photonic systems. Yet the reliance on bulky refractive elements has imposed fundamental limits on how precisely and how compactly light can be shaped. Metasurfaces offer a striking change from this paradigm [1, 2]. By encoding optical functionality into a single, nanostructured interface, they provide local control over phase, amplitude, and polarization with subwavelength resolution [1, 3]. This flat-optics framework enables unprecedented control of wavefronts in ultracompact geometries and is rapidly reshaping imaging, sensing, and on-chip photonics [2, 4].

Evolution of Metamaterials

The development of metasurfaces builds on decades of progress in artificial electromagnetic materials. Early work on artificial dielectrics showed that embedding subwavelength inclusions inside a host medium could modify the effective permittivity and permeability, much like atoms determine the response of natural crystals. Veselago's prediction of materials with simultaneously negative permittivity and permeability inspired the field of metamaterials, which became experimentally realized in the late 1990s through three-dimensional structures such as metallic wires and split-ring resonators [5]. These bulk metamaterials enabled exotic effects, including negative refraction and cloaking (invisibility), but their practical use was limited by fabrication complexity, high losses, and the need for volumetric unit cells much smaller than the wavelength [5, 6]. The shift to two-dimensional metasurfaces emerged as a natural evolution, retaining the wavefront-engineering capabilities of metamaterials while eliminating the bulk and reducing losses [1, 2].

Scope and Aim

This review provides a structured introduction to the physics, design strategies, and applications of optical metasurfaces. We begin by establishing the theoretical foundations in Section II that describe how phase discontinuities and scalar diffraction theory diffraction govern metasurface wavefront shaping. We then discuss practical design methodologies in Section III, including phase implementation mechanisms, meta-atom engineering, and system-level considerations. Building on this framework, we survey key applications such as flat lenses, imaging systems, polarization control in Section IV. Finally, we outline current challenges and emerging directions that will shape the next generation of metasurface technologies in Section V.

II. THEORETICAL FOUNDATIONS

This section establishes the physical principles governing metasurfaces. We begin by contrasting the phase mechanisms of classical refractive optics with the abrupt phase discontinuities of metasurfaces (II.A, II.B). We then discuss the wave theory describing how we can extract coherent information from these abrupt phase discontinuities (II.C).

A. Classical Optics and Phase Accumulation

To understand the innovation of metasurfaces, we must first consider how conventional optical elements work. Classical components, such as glass lenses and prisms, shape light by relying on a gradual accumulation of phase during propagation. As a monochromatic field travels a distance $L(x, y)$ through a material with refractive index n, it accumulates a position-dependent phase shift $\Delta\phi$

given by

$$\Delta\phi(x, y) = k_0 n L(x, y) \tag{1}$$

where $k_0 = 2\pi/\lambda$ is the free-space wavenumber and λ is the wavelength. By deliberately varying the thickness profile $L(x, y)$, classical lenses implement a spatially varying optical path length that reshapes the outgoing wavefront.

A lens that focuses an incident plane wave must impose the phase of a converging spherical wave. If the focal point lies at $(0, 0, f)$, the required phase delay at transverse coordinate (x, y) is determined by equalizing the optical path from every point on the surface to the focus.

Consider two waves in the wavefront, going through $(0, 0)$ and (x, y) respectively. We want the two waves to accumulate the same total phase. However, the distance from the origin to the focal point is simply $d = f$, whereas the distance from point (x, y) is $d = \sqrt{x^2 + y^2 + z^2}$. The condition for constructive interference at the focus is that the total phase accumulation along any path must be equal (modulo 2π). Comparing a path through (x, y) to the axial path:

$$\Phi_{\text{lens}}(x, y) + \phi_{\text{air}}(x, y) = \Phi_{\text{lens}}(0, 0) + k_0 f$$

Assuming $\Phi_{\text{lens}}(0, 0) = 0$ as a reference point, we can solve for the ideal phase profile Φ_{lens},

$$\Phi_{\text{lens}}(x, y) = -k_0 \left(\sqrt{x^2 + y^2 + f^2} - f \right) \tag{2}$$

Since the focal length is much longer than the distance from (x, y) to the optical axis, we can use the paraxial approximation $x^2 + y^2 \ll f^2$, which simplies the ideal phase profile to

$$\Phi_{\text{lens}}(x, y) \approx -\frac{k_0}{2f}(x^2 + y^2). \tag{3}$$

From Equation 3, the closer to the origin, the smaller the ideal phase. To align the wavefront, we then need the largest phase delay at the center, which quadratically decreases further away from the center. Thus, as shown in Figure 1, the converging lens is thickest in the center and gradually becomes thinner towards the edges.

However, this mechanism imposes a fundamental limitation. Since physical thickness is the only "degree of freedom" for controlling phase, achieving a full 2π modulation typically requires a propagation distance on he order of $\lambda/(n - 1)$. For visible wavelengths and common optical glasses, this corresponds to multi-micron to millimeter-scale thickness, far larger than the wavelength. Crucially, in refractive optics, the diameter and thickness are coupled, such that increasing the numerical aperture or lens diameter requires a physically thicker device. This coupling prevents the realization of large-aperture, short-focal-length lenses in a compact form factor.

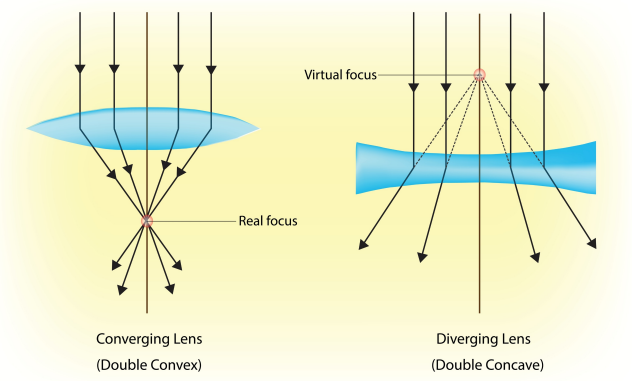

FIG. 1: Left: Converging lens. Right: Diverging lens. Demonstrates the effect of the ideal phase profile for refractive optics.[7]

Beyond geometric thickness, classical refractive optics also face limitations in flexibility. Aberrations such as spherical aberration, coma, and astigmatism arise naturally because only a few macroscopic surfaces are available to approximate the ideal phase profile. Correcting these errors typically requires stacking several lens elements, increasing system size, alignment complexity, and weight.

For centuries, we assumed we need a thick piece of curved glass to change the phase of light. But physically, it does not matter how we implement this phase change, as long as there is a spatial phase profile in place. As the next section will show, metasurfaces break from this paradigm by implementing the phase profile, or even more complex ones, using abrupt phase discontinuities at a single, optically thin surface. This shift from volumetric to surface-based phase engineering is the key enabling principle of flat optical metasurface systems.

B. Generalized Laws of Reflection and Refraction

From the previous section, we saw that classical refractive optics manipulate light through gradual phase accumulation over a propagation distance $L(x, y)$. In contrast, metasurfaces operate by engineering abrupt phase discontinuities at a single interface.

To describe the trajectory of light across a metasurface, we cannot rely on the conventional Snell's Law. Consider an interface at $y = 0$ separating two media of refractive indices n_i and n_t. A plane wave is incident at an angle θ_i. In the conventional case, the phase must remain continuous along the boundary, which enforces conservation of tangential momentum and yields the standard Snell's law

$$n_i \sin\theta_i = n_t \sin\theta_t. \tag{4}$$

A metasurface, however, imposes a spatially varying

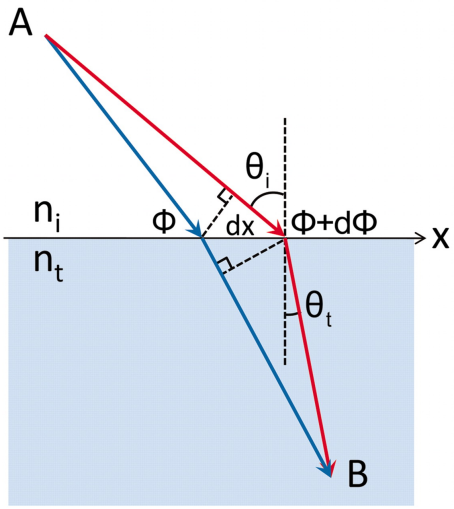

FIG. 2: Schematic of the generalized Snell's law derivation. Two parallel rays separated by dx acquire a phase difference due to the metasurface gradient $\Phi(x)$, steering the beam to an anomalous angle [8].

phase shift $\Phi(x)$ at the interface. This additional phase contribution modifies the stationary-phase condition governing the trajectory of the refracted beam.

To derive the resulting law, we apply Fermat's Principle. Consider two rays incident on neighboring points separated by dx along the interface. For the transmitted rays to form a coherent wavefront at angle θ_t, the total phase difference between these two paths must vanish. The total accumulated phase consists of (i) the phase acquired in the incident medium, (ii) the abrupt phase jump from the metasurface, and (iii) the phase accumulated in the transmitted medium. Therefore,

$$k_0 n_i \sin(\theta_i)\, dx + \frac{d\Phi}{dx}\, dx = k_0 n_t \sin(\theta_t)\, dx. \tag{5}$$

Dividing by dx and rearranging yields the *Generalized Snell's Law of Refraction* [8]:

$$n_t \sin(\theta_t) - n_i \sin(\theta_i) = \frac{1}{k_0}\frac{d\Phi}{dx}, \tag{6}$$

where $k_0 = 2\pi/\lambda$ is the free-space wavenumber. The phase gradient thus imparts an additional tangential momentum

$$k_{\parallel}^{(\text{meta})} = \frac{d\Phi}{dx},$$

allowing the transmitted beam to emerge at angles otherwise forbidden by classical refraction. In the special case $d\Phi/dx = 0$, Equation 6 reduces to the standard Snell's Law.

Applying the same reasoning to the reflected path yields the *Generalized Law of Reflection* [8]:

$$\sin(\theta_r) - \sin(\theta_i) = \frac{1}{k_0 n_i}\frac{d\Phi}{dx}. \tag{7}$$

This generalized law explains how a flat metasurface can act as a lens, a beam splitter, or even a retroreflector, simply by engineering the local phase gradient along the surface. A simple linear phase profile $\Phi(u) = k \sin\theta u$, leads to anomalous beam deflection by an angle θ. By designing more complex phase maps, metasurfaces can accomplish even more complicated functionalities than convex lenses, all with a flat design.

The generalized laws of reflection and refraction derived above describe how a metasurface redirects light by imparting a spatial phase discontinuity in a ray-optical view. However, many of the most important applications of metasurfaces are inherently wave-optical phenomena. In the next subsection, we therefore move from a ray-based description to a wave-based framework, using scalar diffraction theory to link metasurface phase distributions to the resulting propagated wavefronts.

C. Wavefront Reconstruction & Diffraction Theory

A rigorous description of focusing, imaging, and complex wavefront shaping requires a full wave-optical treatment. The generalized laws specify the *direction* of the outgoing wavefront, but diffraction theory determines how this field evolves in space and interferes

In a homogeneous, linear, isotropic medium, the propagation of monochromatic light is governed by the scalar Helmholtz equation

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = 0, \tag{8}$$

where $\psi(\mathbf{r})$ represents a component of the electric or magnetic field, and $k = nk_0$ is the wavenumber in the medium.

A metasurface can be modeled as a thin phase screen at $z = 0$. When an incident wave ψ_i impinges on the surface, the meta-atoms impose a local amplitude and phase modulation, producing the transmitted source field at $z = 0^+$:

$$\psi(x, y, 0) = A(x, y)\, e^{i\Phi(x,y)} \cdot \psi_i(x, y, 0^-). \tag{9}$$

Here $A(x, y)$ denotes the local transmission amplitudes and $\Phi(x, y)$ is the engineered phase profile.

To determine the field at a distance z beyond the metasurface, we invoke Huygens' Principle, which states that every point on a wavefront radiates a secondary spherical wave. The total field is obtained by coherently summing these contributions. Solving the Helmholtz equation with Green's functions and method of images leads to the Rayleigh–Sommerfeld diffraction integral:

$$\psi(x, y, z) = \frac{1}{i\lambda}\iint_{\Sigma} \psi(x', y', 0)\, \frac{e^{ikr}}{r}\, \cos\theta\, dx' dy', \tag{10}$$

where r is the distance from the source point (x', y') to the observation point (x, y, z), and θ is the angle between the propagation direction and the surface normal.

Under the paraxial approximation, the Rayleigh–Sommerfeld integral simplifies into a form that provides direct physical insight into metasurface focusing. Let the source point be $(x', y', 0)$ and the observation point be (x, y, z). The distance between them is

$$r = \sqrt{(x - x')^2 + (y - y')^2 + z^2}.$$

When $z \gg |x - x'|, |y - y'|$, we expand r to first order:

$$r \approx z + \frac{(x - x')^2 + (y - y')^2}{2z}, \quad \frac{1}{r} \approx \frac{1}{z}, \quad \cos\theta \approx 1 \tag{11}$$

Substituting Eq. (11) into the Rayleigh–Sommerfeld integral and extracting the common phase factor e^{ikz} gives the Fresnel diffraction formula:

$$\psi(x, y, z) \approx \frac{e^{ikz}}{i\lambda z} \iint_{\Sigma} \psi(x', y', 0) \exp\left[i\frac{k}{2z}\left((x - x')^2 + (y - y')^2\right)\right] dx'dy'. \tag{12}$$

To expose the structure of Eq. (12), expand the quadratic term:

$$(x - x')^2 + (y - y')^2 = x^2 + y^2 + x'^2 + y'^2 - 2(xx' + yy').$$

Substitution yields

$$\psi(x, y, z) \approx \frac{e^{ikz}}{i\lambda z} \exp\left[i\frac{k}{2z}(x^2 + y^2)\right] \iint_{\Sigma} \psi(x', y', 0) \exp\left[i\frac{k}{2z}(x'^2 + y'^2)\right] \exp\left[-i\frac{k}{z}(xx' + yy')\right] dx'dy'. \tag{13}$$

The final exponential is a Fourier kernel with spatial frequencies

$$k_x = \frac{kx}{z}, \qquad k_y = \frac{ky}{z}.$$

If the metasurface implements the thin-lens phase

$$\Phi_{\text{lens}}(x', y') = -\frac{k}{2f}(x'^2 + y'^2),$$

then substituting $\psi(x', y', 0) = \psi_i(x', y', 0^-)e^{i\Phi_{\text{lens}}}$ into Eq. (13) and evaluating at $z = f$ cancels the quadratic term. This yields

$$\psi(x, y, f) \propto \iint_{\Sigma} \psi_i(x', y', 0^-) \exp\left[-i\frac{k}{f}(xx' + yy')\right] dx'dy', \tag{14}$$

showing that the field at the focal plane is the two-dimensional Fourier transform of the incident field at the metasurface. This result explains why prescribing $\Phi(x, y)$ at the metasurface directly determines the outgoing wavefront in the focal region.

Because a metasurface samples a continuous phase profile using discrete unit cells, the lattice pitch must be sufficiently small to avoid generating propagating higher-order diffraction modes, motivating why we want nano-sized structures as scatterers. This requirement emerges naturally from the Fourier decomposition of the target wavefront and reflects a fundamental link between metasurface geometry and achievable optical fields. We will expand on this further in section III.C.

Summary. Taken together, the three components of this section establish the theoretical basis for flat-optical wavefront control. In Section II.A, we contrasted classical thickness-driven phase accumulation with the ideal phase profiles required for focusing, highlighting the constraints that motivate planar alternatives. Section II.B showed that abrupt, subwavelength phase discontinuities fundamentally generalize Snell's laws by allowing engineered momentum transfer at an interface. Section II.C connected these phase profiles to full wave propagation through scalar diffraction theory.

Having established the physical principles that govern metasurface behavior, we now turn to their practical implementation. The next section examines how individual meta-atoms generate the required phase, amplitude, and polarization responses, and how these building blocks are assembled into complete, high-performance flat-optical devices.

III. DESIGN PRINCIPLES

Having established how a prescribed phase profile $\Phi(x, y)$ determines the resulting optical field, we now turn to the question of how such a profile is physically realized. In Section III.A, we review the four principal phase implementation mechanisms: geometric, propagation, resonant, and Huygens, each providing complementary routes to realizing the 2π phase coverage. Section III.B describes meta-atom design strategies, and Section III.C expands the design perspective to a system level. Together, these methodologies build from a theoretical level to the foundations for engineering flat metasurfaces.

A. Phase Implementation Mechanisms

The central goal of metasurface design is to realize a prescribed complex transmission function $t(x, y) = A(x, y)e^{i\Phi(x,y)}$ using an array of subwavelength meta-atoms as motivated by Section II. We want $A(x, y) \approx 1$ so that we have 100% transmission and as little reflected light.

In practice, each unit cell must be assigned a specific nanostructure whose electromagnetic scattering response

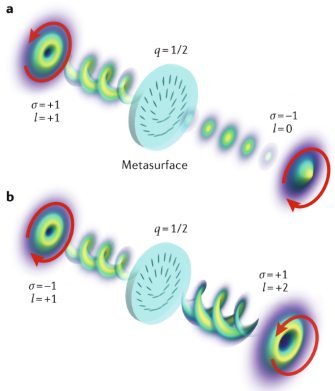

FIG. 3: Image of left handed and right handed circularly polarized light travelling through a PB phase guide [9].

approximates the desired local amplitude, phase, and polarization transformation. Phase implementation mechanisms provide a set of physical principles that link geometry and material parameters to complex transmission coefficients. We review the four most widely used mechanisms below.

Geometric (Pancharatnam-Berry) Phase. The geometric phase (or Pancharatnam-Berry phase) arises purely from the spatial orientation of the meta-atom. This mechanism relies on the polarization evolution of light as it interacts with an anisotropic scatterer, such as a rectangular nanofin, which functions as a localized waveplate.

Consider a meta-atom functioning as a birefringent waveplate with phase delay ϕ. Its Jones matrix in a rotated frame (angle θ) is given by $J(\theta) = R(-\theta)J_0R(\theta)$. When illuminated by left-circularly polarized light ($|L\rangle$), the transmitted field is:

$$J(\theta)|L\rangle = \cos\left(\frac{\phi}{2}\right)|L\rangle + i\sin\left(\frac{\phi}{2}\right)e^{i2\theta}|R\rangle \tag{15}$$

The second term represents the cross-polarized component (spin-flipped from L to R). Crucially, this component acquires a phase shift of 2θ, which depends purely on the orientation angle. By orienting meta-atoms from 0 to π, the transmitted wavefront can be shaped with a full 0 to 2π phase range. The conversion efficiency is maximized when the meta-atom acts as a perfect half-wave plate ($\phi = \pi$), suppressing the co-polarized leakage term.

The geometric phase offers two distinct advantages. First, it is robust against fabrication errors. Small variations in the pillar diameter do not alter the phase, provided the rotation angle is correct. Second, it is inherently broadband, as the phase depends on spatial orientation rather than spectral resonance.

Propagation Phase. The propagation phase mechanism operates on the principle of optical path length modulation. Each meta-atom is modeled as a truncated dielectric waveguide of height H. When an incident plane wave couples into this structure, it excites a fundamental guided mode with a propagation constant β.

The phase accumulated by the wave as it traverses the pillar is given by the product of the propagation constant and the height:

$$\Phi(D) = \beta(D) \cdot H = \frac{2\pi}{\lambda} n_{\text{eff}}(D) \cdot H \tag{16}$$

where n_{eff} is the effective refractive index of the fundamental mode.

The critical design parameter is the effective index $n_{\text{eff}}(D)$, which is a modal property determined by the confinement of the optical field rather than a bulk material constant. For small diameters, the mode is weakly confined and a significant fraction of the field resides in the low-index surroundings, leading to a low n_{eff}. As the diameter increases, the mode is pulled into the high-index core and n_{eff} approaches the bulk index. By sweeping D at fixed height H, one can therefore tune the phase delay continuously and, provided that $k_0 H(n_{\max} - n_{\min}) \gtrsim 2\pi$, achieve full 0–$2\pi$ coverage.

Whereas refractive bulk optics changed the optical path length by increasing thickness, we keep the thickness here constant and tune the effective index. In doing so, our degree of freedom switches to more thin-friendly parameter.

For rotationally symmetric pillars this mechanism is nearly polarization insensitive, which makes propagation-phase designs the standard choice for high-efficiency, polarization-independent metalenses and imaging systems.

Resonant Phase. The resonant phase mechanism fundamentally differs from propagation phase by relying on the temporal storage and re-emission of light energy. A resonant meta-atom is modeled as a driven damped harmonic oscillator. Its optical response is described by the complex polarizability $\alpha(\omega)$, which follows a profile:

$$\alpha(\omega) \propto \frac{1}{\omega_0^2 - \omega^2 - i\gamma\omega} \tag{17}$$

where ω is the incident optical frequency, ω_0 is the natural resonant frequency of the meta-atom, and γ is the damping rate.

The phase of the scattered field is determined by the argument of this complex polarizability:

$$\Phi(\omega) = \arg[\alpha(\omega)] = \arctan\left(\frac{\gamma\omega}{\omega_0^2 - \omega^2}\right) \tag{18}$$

This equation reveals the mechanism for phase control:

- Far below resonance ($\omega \ll \omega_0$), the structure responds in-phase ($\Phi \approx 0$).
- At resonance ($\omega = \omega_0$), the response lags by $\pi/2$.

- Far above resonance ($\omega \gg \omega_0$), the response approaches an out-of-phase condition ($\Phi \approx \pi$).

In a practical design, the operating frequency ω is fixed. Therefore, to spatially modulate the phase $\Phi(x, y)$, one tunes the local resonant frequency $\omega_0(x, y)$ by varying the geometry, such as the length of a plasmonic rod or the width of a dielectric block.

A fundamental limitation of single-mode resonators, like an electric dipole, is that they can only provide a phase shift range of π, which is insufficient for full wavefront control. To achieve the full 2π coverage required for lensing, multiple resonances must be combined—typically an electric dipole and a magnetic dipole—leading to the concept of Huygens' metasurfaces. Furthermore, because the phase varies rapidly near ω_0, resonant metasurfaces are inherently narrowband and dispersive, exhibiting strong chromatic aberration.

Huygens Metasurfaces. To achieve unity transmission with full 2π phase coverage, one must invoke the concept of Huygens' sources, which rely on the interference between electric and magnetic dipole modes.

By doing a multipole expansion, the complex transmission (t) and reflection (r) coefficients of a meta-atom array can be expressed in terms of its effective electric (α_e) and magnetic (α_m) polarizabilities:

$$r = \frac{ik}{2S}(\alpha_e - \alpha_m \eta_0), \quad t = 1 + \frac{ik}{2S}(\alpha_e + \alpha_m \eta_0) \tag{19}$$

where S is the unit cell area and η_0 is the impedance of free space.

Examining the reflection equation reveals the condition for zero backscattering. If the electric and magnetic polarizabilities are balanced such that $\alpha_e = \alpha_m \eta_0$, the reflection coefficient vanishes ($r = 0$). This is known as the *first Kerker condition*. Physically, this corresponds to the destructive interference of the back-scattered fields from the induced electric and magnetic dipoles, resulting in a unidirectional forward scattering pattern. A representative physical implementation of the Huygens metasurface can be found in Figure 4.

While plasmonic structures are dominated by electric dipoles, high-index dielectric cylinders (such as Silicon or TiO_2) naturally support both electric and magnetic Mie resonances.

- By tuning the aspect ratio (height-to-diameter ratio) of the pillar, the spectral positions of the electric and magnetic dipole resonances can be shifted relative to each other.

- When these two resonances are engineered to spectrally overlap at the operating wavelength, the Kerker condition is satisfied.

In summary, these four mechanisms offer complementary routes to wavefront control. However, selecting a phase control mechanism is only the first step. In the following section, we examine the practical workflows for meta-atom design, moving from material selection to the computational strategies used to build unit-cell libraries.

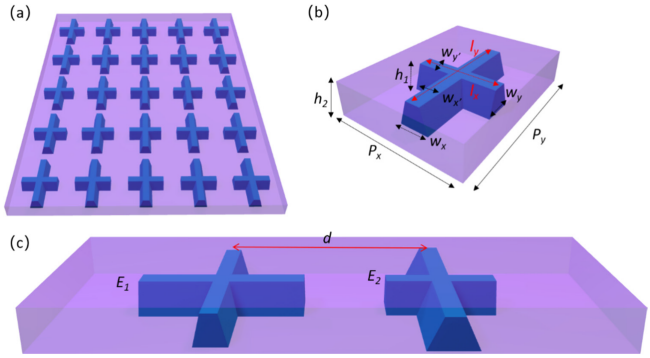

FIG. 4: T shaped Huygen's metasurface implementation, where the horizontal and vertical components are the orthogonal electric and magnetic dipoles [10].

B. Meta-atom Design

The phase implementation mechanisms of Section III.A describe the physical principles by which individual scatterers can impart a desired complex transmission coefficients $t = Ae^{i\Phi}$. In this subsection, we outline the key considerations in constructing meta-atoms, including material selection, modal properties, geometric parameterization, lattice effects, and lookup-table generation for full-device design.

Material Platforms. The choice of material strongly influences the optical performance of a meta-atom, as it determines the achievable refractive index contrast, absorption losses, dispersion, and fabrication compatibility. Early optical metasurfaces often employed plasmonic metals such as gold, silver, or aluminum, which support strong field confinement and enable ultrathin resonant structures but suffer from significant Ohmic losses, limiting their efficiency and bandwidth.

For most visible and near-infrared applications, high-index dielectrics such as TiO_2, SiN, GaN, and amorphous Si have become the dominant platform because they support low-loss Mie-type electric and magnetic resonances, allowing for Huygen's sources, and provide large phase shifts with high transmission.

Geometry and Parameterization. Once a material platform is selected, the primary design degrees of freedom arise from the geometry of the meta-atom, which determines how incident light couples into guided modes, resonant modes, or polarization-converting channels. Typical meta-atom shapes include cylindrical and rectangular pillars, elliptical posts, nanofins, and slot-based structures, each offering a specific set of tunable

parameters such as diameter, width, length, aspect ratio, height, and in-plane rotation. These geometric parameters directly control quantities such as effective index, modal confinement, resonance frequency, scattering cross section, and polarization response, thereby shaping the complex transmission function $t = Ae^{i\Phi}$.

Symmetry plays a fundamental role: structures with rotational symmetry (e.g., circular pillars) produce polarization-insensitive responses, whereas anisotropic shapes support birefringence and enable geometric-phase control through in-plane rotation. The achievable design space is also constrained by fabrication limits, including minimum feature size, sidewall angle, and allowable aspect ratios, which can restrict the range of realizable phase shifts or reduce efficiency if the geometry deviates from its ideal profile. Together, these geometric considerations define the map from physical shape to optical function and form the basis for constructing unit-cell libraries used in full-aperture metasurface design.

Coupling and Lattice Effects. Although meta-atoms are often designed and simulated under the assumption of local periodicity, their optical response in a real metasurface can deviate from isolated unit-cell predictions due to interactions with neighboring elements and the periodic lattice. These coupling effects arise from near-field interactions, where evanescent fields overlap between closely spaced scatterers, and from far-field diffraction channels that depend on the lattice pitch. When the lattice period is deeply subwavelength, the structure supports only the zeroth diffraction order, and the local periodicity approximation (LPA) provides an accurate estimate of the transmission and phase.

As the pitch approaches the wavelength or as the phase gradient increases, the LPA breaks down. These interactions can either degrade performance through parasitic coupling or be deliberately exploited to create nonlocal metasurfaces with engineered spatial dispersion.

C. System-Level Design

The meta-atom design strategies discussed above determine the local scattering response of individual unit cells. However, the performance of a complete metasurface emerges only when these responses are assembled across a finite aperture. System-level design addresses how local phase and amplitude assignments translate into global wavefront control, accounting for discretization, dispersion, and collective optimization.

Forward Design and Unit-Cell Assignment. Most metasurface devices are constructed using a forward design workflow. A target phase profile $\Phi(x, y)$ is first specified based on the desired optical function, such as focusing or beam steering. A precomputed unit-cell library is then used to assign, at each position (x, y), the meta-atom whose transmission response best approximates the required local phase and amplitude. This approach is computationally efficient and provides clear physical intuition, making it the standard method for metalenses and beam shapers.

However, forward design relies on the assumption that each meta-atom behaves independently and that its response is well approximated by local periodic simulations. As device complexity increases or when strong coupling, broadband operation, or multifunctionality is required, this approximation becomes less accurate. These limitations motivate the development of global optimization techniques, which are discussed later in this section.

Aperture Sampling and Phase Discretization. Implementing a continuous phase profile using a discrete array of meta-atoms is fundamentally constrained by sampling theory. To avoid spatial aliasing, the unit-cell pitch P must resolve the highest spatial frequency in the phase profile, which for a metalens of numerical aperture NA requires

$$P < \frac{\lambda}{2\,\mathrm{NA}}. \tag{20}$$

This constraint becomes increasingly restrictive at high NA, where steep phase gradients occur near the aperture edge. In practice, the pitch must also remain subwavelength to suppress higher-order diffraction modes, further tightening design margins.

In addition to spatial sampling, metasurfaces exhibit phase discretization because only a finite number of distinct geometries can be fabricated. Early binary designs suffered large efficiency losses, but modern libraries with 8–16 phase levels provide sufficiently fine discretization for high-quality wavefront reconstruction. When combined with adequate spatial sampling, phase quantization errors can be reduced to a minor contribution compared to fabrication and material losses.

Chromatic Dispersion. Chromatic aberration remains a central challenge for metasurface optics. Because metasurfaces are inherently diffractive, the focal length of a conventional metalens scales approximately as $f(\lambda) \propto \lambda$, leading to strong wavelength-dependent defocus. Correcting this behavior requires meta-atoms that provide not only a prescribed phase $\Phi(\omega_0)$, but also a tailored group delay $\partial\Phi/\partial\omega$.

Physical constraints limit how much group delay a single meta-atom can supply, linking dispersion correction to structure height, resonance lifetime, and refractive index contrast. As a result, achromatic metasurfaces are currently restricted to small apertures or narrow bandwidths. System-level strategies to address dispersion include resonant phase engineering, multi-layer metasurfaces, and hybrid refractive–diffractive architectures, each trading bandwidth, efficiency, and fabrication complexity.

Summary. Together, the phase implementation mechanisms of Section III.A, the meta-atom design prin-

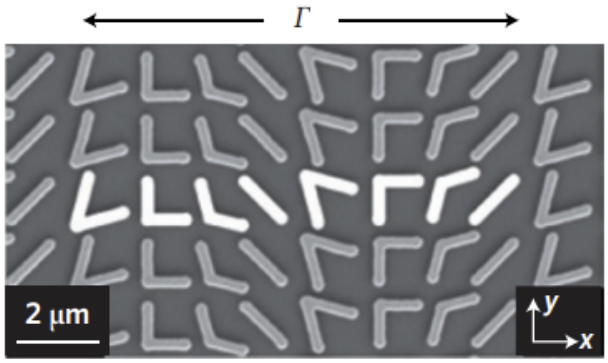

FIG. 5: V shaped gold optical antennas on silicon base. An example of individual meta-atom design and system-level design [1].

ciples of Section III.B, and the system-level considerations presented here establish a complete framework for translating abstract wavefront requirements into realizable flat-optical devices. These layers of design determine not only whether a metasurface functions in principle, but whether it performs efficiently, robustly, and over useful bandwidths. With this design hierarchy in place, we now turn to the applications enabled by metasurfaces, examining how these principles are realized in imaging, display, and sensing systems.

IV. APPLICATIONS

Metasurfaces translate the phase, amplitude, and polarization control developed in Sections II–III into a broad range of practical optical functions. Their ability to engineer arbitrary wavefronts within a single planar interface enables devices that are thinner, lighter, and more multifunctional than those achievable with refractive or diffractive optics alone. In this section, we survey key application domains—from metalenses for imaging systems and beam shaping (Section IV.A) to holography and display (Section IV.B), sensing and polarization detection (Section IV.C).

A. Metalenses for Imaging and Beam Shaping

Metasurfaces offer an alternative route to conventional optics by encoding the desired phase, amplitude, and polarization response directly onto a single patterned interface with subwavelength spatial resolution. A metalens can implement the full hyperboloidal phase profile required for aberration-free focusing without increasing physical thickness, and can simultaneously correct aberrations or engineer extended depth of focus by appropriate choice of local phase $\Phi(x, y)$. The same design principles enable beam-shaping elements that generate Bessel, vortex, or other structured beams, with the corresponding phase profiles directly realized through geometric, propagation, resonant, or Huygens-type phase mechanisms described in Section III. Because these devices are planar and compatible with micro- and nanofabrication, they can be monolithically integrated with image sensors, folded optical paths, and chip-scale photonic circuits, enabling compact cameras and beam-delivery systems that are difficult or impossible with bulk optics.

A broad set of demonstrations has established the viability and versatility of metasurface-based imaging and beam shaping. Phase-gradient metasurfaces were first used to realize generalized reflection and refraction, providing the conceptual framework for flat lenses and anomalous beam deflection [1, 8]. Dielectric metalenses have since achieved high NA, high efficiency, and broadband performance across the visible and near-infrared, as surveyed in [2, 3]. In Figure 6, we see this culminating in an effectively similar NA level optical system as the iPhone, but order of magnitudes thinner. At the system level, metasurface lenses have been integrated into ultrathin and folded camera architectures [11], while co-designed optics–neural pipelines such as Nano-3D use a single metasurface to encode depth cues for downstream 3D reconstruction [12]. These results indicate that metalenses and beam-shaping metasurfaces are transitioning from laboratory demonstrations to practical optical components, with ongoing work focused on improving efficiency, bandwidth, and manufacturability at scale.

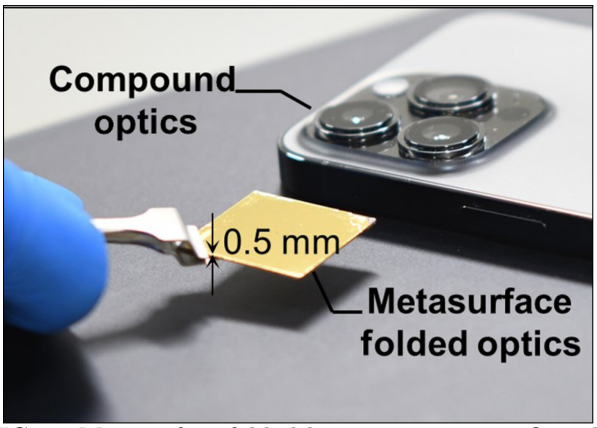

FIG. 6: Metasurface folded lens system is significantly thinner than iPhone camera lens [11].

B. Holography and Display

Conventional holography and display platforms typically rely on spatial light modulators (SLMs), digital micromirror devices, or bulk diffractive elements to control phase and amplitude across an aperture. These architectures enable dynamic holographic projection and near-eye display, but they are constrained by pixel pitch, fill

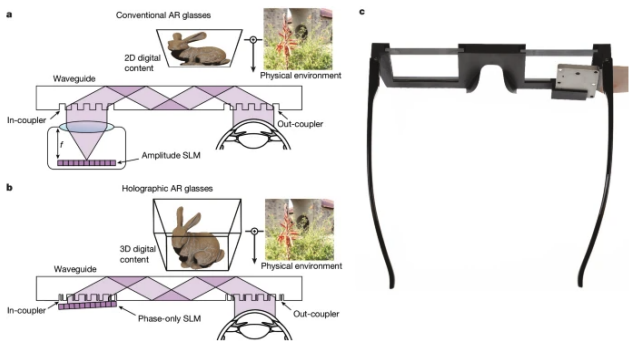

FIG. 7: A) Conventional AR waveguides. B) Holographic AR waveguides using metasurface SLMs. C) Holographic AR glasses in compact form [13].

factor, system thickness, and limited joint control over polarization and spectral response. In augmented- and mixed-reality, these constraints become more acute because wide field of view and large eyebox must be delivered in a lightweight, glasses-scale form factor.

Metasurfaces provide a compact route to holographic wavefront synthesis by implementing computer-generated phase masks with subwavelength sampling in a single patterned layer [1, 2]. Phase-only holograms can be realized using geometric, propagation, or resonant phase mechanisms, allowing high-efficiency reconstruction of designed 2D or 3D fields without the thickness of volume holograms [1, 2]. When integrated with planar waveguides, metasurfaces can act as in-couplers, out-couplers, and pupil expanders that shape the wavefront delivered to the eye within a thin optical stack [2, 13]. Multiplexed designs (wavelength, polarization, angle, or depth) further enable full-color operation and multi-plane image delivery within the same footprint [2, 4].

State-of-the-art demonstrations have realized full-colour 3D holographic AR displays using metasurface waveguides that jointly address coupling, beam expansion, and image formation in a glasses-like geometry [13], see Figure 7. More broadly, foundational flat-optics work established metasurface holography as a general framework for arbitrary wavefront synthesis via designer phase discontinuities [1]. Recent surveys emphasize rapid progress in efficiency, multiplexing, and waveguide integration, positioning metasurface holography as a leading candidate for next-generation AR/VR combiners and compact projection [2]. In parallel, inverse-design methods are increasingly used to co-optimize the holographic field, device constraints, and fabrication limits, improving fidelity in regimes where local phase assignment becomes inadequate [4].

C. Metasurface Sensors and Polarization Detection

Optical sensing and polarization analysis have traditionally been implemented using bulk dispersive optics (prisms and gratings), interference filters, and cascaded polarization optics (polarizers and waveplates) placed ahead of a detector. These instruments can achieve excellent sensitivity and dynamic range, but they require nontrivial propagation distances and multi-element optical trains that constrain size, alignment tolerance, and cost. Even compact polarization cameras based on micropolarizer arrays trade spatial resolution for polarimetric capability and often face channel cross-talk and limited spectral flexibility.

Metasurfaces enable sensing functions to be integrated directly at the aperture or detector plane by engineering spatially varying amplitude, phase, and polarization transfer functions at the unit-cell level [1, 2]. Anisotropic meta-atoms can encode polarization-dependent phase delays, allowing polarization states (or Stokes parameters) to be mapped into distinct intensity patterns that can be read out in a single shot [1, 14]. Resonant and Huygens-type metasurfaces can also generate sharp spectral features and strong near-field enhancements that increase sensitivity to refractive-index changes or molecular binding events near the surface [2, 5, 6]. More generally, spatial multiplexing across the surface allows a metasurface to implement a structured measurement operator that is inverted computationally, compressing spectral, angular, and polarization information into fewer measurements [2, 4, 12].

Current work spans both nanophotonic and computational directions, including compact spectrometers and dispersive sensors that replace bulk gratings with folded or planar metasurface architectures [2, 15]. On the polarimetry side, dielectric metasurface polarimeters have demonstrated full-Stokes capability in compact optical stacks, pointing toward pixel-scale polarimetric cameras without external polarization optics [2, 14]. Roadmap and review efforts emphasize that performance is increasingly determined by co-design across optics, materials, and readout algorithms rather than by the meta-atom response alone [2, 5, 6]. This convergence with inverse design and learned reconstruction is now enabling metasurface sensors that function as end-to-end measurement systems, rather than miniaturized versions of classical instruments [4, 12].

Summary. Taken together, the three application domains reviewed in this section illustrate the breadth of functionality enabled by metasurfaces. In the next section, we examine these limitations in detail and discuss emerging strategies that aim to close the gap between metasurface research and robust, deployable optical systems.

V. LIMITATIONS & FUTURE CHALLENGES

The applications highlighted in the previous section demonstrate the transformative potential of metasurfaces to replace bulky refractive optics and enable novel wavefront functionalities. However, a significant gap remains between high-performance laboratory prototypes and robust, commercially viable components ready for widespread consumer use and industrial adoption. In this section, we examine the critical barriers currently limiting the field. First, we consider the manufacturing challenges associated with scaling nanostructure manufacturing (Section V.A), and shifting from forward design unit-cell libraries to computational inverse design and topology optimization frameworks (Section V.B).

A. Fabrication and Scalability

The transition of metasurfaces from laboratory curiosities to commercial products hinges entirely on the scalability of fabrication.

The Limits of Electron-Beam Lithography. The vast majority of high-performance metasurfaces reported in the literature are fabricated using Electron-Beam Lithography (EBL). EBL offers unparalleled resolution (< 10 nm) and design flexibility, allowing researchers to define arbitrary, aperiodic meta-atom patterns without the need for physical masks. However, EBL is inherently a sequential writing process. The beam physically etch every pixel of the pattern. Consequently, writing a single centimeter-scale metalens can take many hours or even days, making EBL fundamentally unviable for mass production [2, 16]. This "throughput gap" creates a dichotomy where the highest-performing devices cannot be economically scaled.

DUV Lithography. To achieve industrial throughput, fabrication must migrate to Deep-Ultraviolet (DUV) photolithography, the standard for semiconductor manufacturing [16, 17]. While DUV enables the parallel patterning of millions of devices, it imposes strict design rules. Unlike EBL, DUV resolution is diffraction-limited. This restricts the minimum gap size between pillars and the minimum pillar diameter, potentially excluding the most efficient high-aspect-ratio meta-atom designs found in EBL libraries. Thus, despite high through-put, DUV limits the flexibility of meta-atom design, a key feature to meta-atom functionality.

Nanoimprint Lithography (NIL). An alternative route to scalability is Nanoimprint Lithography (NIL) [16, 18, 19], where a master mold is physically pressed into a polymer resist to replicate the pattern. NIL is highly attractive for its low cost and ability to replicate sub-10 nm features over large areas without complex optics. However, NIL faces challenges with defect density and mold durability. The high-aspect-ratio pillars required for visible-light metasurfaces are fragile. The demolding process can cause pillar collapse or ripping, particularly for dense, high-NA designs.

High-Aspect-Ratio Etching. Regardless of the patterning method, transferring the pattern into high-index materials (like TiO_2, GaN, or Silicon) requires Deep Reactive Ion Etching (DRIE). Achieving vertical, smooth sidewalls for pillars with aspect ratios exceeding 10:1 is non-trivial [2, 3, 18]. Tapered sidewalls, a common etching effect, change the effective index of the waveguide mode, introducing phase errors that differ from the theoretical unit-cell simulations. Furthermore, surface roughness introduces scattering losses that accumulate over the depth of the pillar, significantly reducing the transmission efficiency of the final device.

B. Inverse Design and Co-Design

The traditional "forward design" workflow—mapping a target phase profile to a pre-computed library of unit cells—relies heavily on the local periodicity approximation and intense numerical computation. While effective for simple lenses, this approach struggles to address multi-objective challenges such as achromaticity, wide-angle efficiency, or polarization multiplexing, where the required scattering response often lies outside the capabilities of simple geometric primitives. To overcome these limits, the field is increasingly pivoting toward computational inverse design and end-to-end co-design frameworks.

Topology Optimization. Inverse design treats the metasurface layout not as an assembly of discrete pillars, but as a continuous distribution of material that can be optimized pixel-by-pixel [4]. Techniques such as the adjoint method—originally developed for aerodynamic shape optimization—allow researchers to calculate the gradient of a "figure of merit" with respect to the permittivity of every voxel in the simulation volume. By iteratively updating the geometry via standard optimization methods such as gradient descent, the algorithm discovers non-intuitive, freeform nanostructures that resemble random organic voids or labyrinths. These computer-generated geometries can exploit multiple scattering channels and near-field coupling to achieve efficiencies or bandwidths that are physically impossible with standard cylindrical pillars.

Deep Learning for Accelerated Design. While topology optimization yields high-performance devices, it is, like unit-cell libraries, computationally expensive, often requiring hundreds of full-wave simulations per optimization run. Deep learning offers a route to accelerate this process, an example shown in Figure 8. Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) can be trained on datasets of metasurface simulations to generate high-efficiency designs in

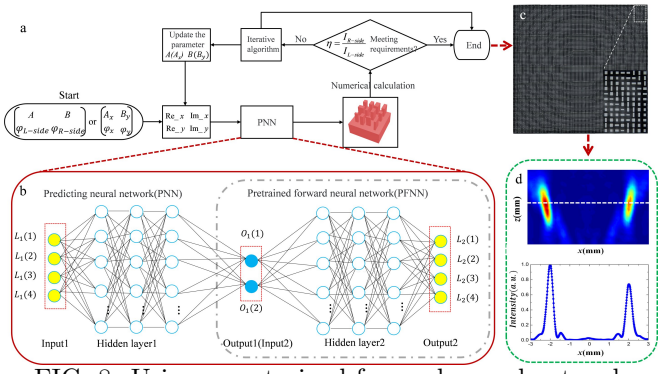

FIG. 8: Using a pretrained forward neural network (PFNN) to predict metasurfaces designed [21].

milliseconds. Furthermore, "surrogate models"—neural networks trained to predict the optical response of a given geometry—can replace the slow Maxwell's solvers in the optimization loop [4, 20], enabling the exploration of vast design spaces that would be intractable with conventional electromagnetic theory.

However, the results of deep learning systems must be taken with extra scrutiny as modern generative AI techniques often hallucinatex and still lag behind classical optimization methods in terms of accuracy. Similar to modern deep learning trends however, scaling up the data trained for these generative models will most likely result in high precision designs.

End-to-End Optical-Digital Co-Design. Perhaps the most transformative direction is the shift from designing "perfect" optical components to designing optimal "optical-digital systems." In classical optics, the goal is to form a sharp image on the sensor. In co-design, the goal is to capture the most information, even if the raw image looks blurry or distorted to the human eye. By modeling the metasurface and the image reconstruction neural network as a single differentiable pipeline, researchers can jointly optimize the physical nanostructure and the digital software [12, 22]. This approach has led to breakthroughs like "Nano-3D" [12], where the metasurface is intentionally designed to encode depth cues into the point spread function, allowing a simple monocular camera to function as a 3D depth sensor. This paradigm suggests that the future of flat optics lies not just in better physics, but in the tight integration of photonics and computation.

Summary. As material innovations and lithographic advances continue to expand the performance horizon, the most significant leaps will likely emerge from a shift toward holistic system-level engineering rather than component-level optimization. By embracing inverse design frameworks and end-to-end co-design, engineers can trade optical perfection for computational robustness. This evolution marks the critical transition of the field from exploring exotic physical phenomena to delivering reliable, mass-manufacturable optical hardware.

VI. CONCLUSION

The development of optical metasurfaces represents a fundamental paradigm shift in photonics, moving the control of light from volumetric accumulation to surface-based phase discontinuities. As this review has detailed, the field has rapidly matured from the initial derivation of the generalized laws of reflection and refraction in 2011 to the demonstration of high-performance devices that rival or exceed the capabilities of conventional refractive optics [1–3, 8]. By leveraging the subwavelength mastery of phase, amplitude, and polarization, metasurfaces have unlocked functionalities—such as single-layer achromatic focusing, polarization multiplexing, and flat-optics holography—that were previously impossible in compact form factors.

Looking forward, the trajectory of metasurface research is poised to converge with the broader trends of computational imaging and semiconductor manufacturing. As fabrication processes scale via deep-UV and nanoimprint lithography, and as design methodologies evolve through deep learning and topology optimization, metasurfaces are set to become the standard optical interface for augmented reality, autonomous sensing, and complex photnoic systems [2, 13, 15]. We are witnessing the dawn of an era where optical complexity is no longer defined by the curvature of glass, but by the information density encoded into a silicon chip.

ACKNOWLEDGEMENTS

We would like to thank Professor Hui Cao and Graduate Teaching Fellow Mert Ercan for their invaluable lectures and support throughout the semester.

* jeffrey.wei@yale.edu

[1] N. Yu and F. Capasso, Flat optics with designer metasurfaces, Nature Materials **13**, 139 (2014).
[2] A. I. Kuznetsov, M. L. Brongersma, et al., Roadmap for optical metasurfaces, ACS Photonics **11**, 816 (2024).
[3] S. M. Kamali, E. Arbabi, A. Arbabi, and A. Faraon, A review of dielectric optical metasurfaces for wavefront control, Nanophotonics **7**, 1041 (2018).
[4] Z. Li, R. Pestourie, Z. Lin, S. G. Johnson, and F. Capasso, Empowering metasurfaces with inverse design: Principles and applications, ACS Photonics **9**, 2178 (2022).
[5] A. Iyer, A. Alu, and A. Epstein, Metamaterials and metasurfaces—historical context, recent advances, and future directions, IEEE Transactions on Antennas and Propagation **68**, 1223 (2020).

[6] A. Shaltout, N. Kinsey, J. Kim, R. Chandrasekar, J. Ndukaife, A. Boltasseva, and V. Shalaev, Development of optical metasurfaces: Emerging concepts and new materials, Proceedings of the IEEE **104**, 1 (2016).

[7] loveideaet.click, Product detail page, https://loveideaet.click/product_details/117838381.html (2025), accessed: 2025-12-12.

[8] N. Yu, P. Genevet, M. A. Kats, F. Aieta, J.-P. Tetienne, and F. Capasso, Light propagation with phase discontinuities: Generalized laws of reflection and refraction, Science **334**, 333 (2011).

[9] E. Cohen, H. Larocque, F. Bouchard, F. Nejadsattari, Y. Gefen, and E. Karimi, Geometric phase from aharonov–bohm to pancharatnam–berry and beyond, Nature Reviews Physics **1**, 437 (2019).

[10] T. Wu, Z. Liu, W. Cao, H. Zhang, D. Yang, and Z. Yang, All-dielectric huygens' metasurface for wavefront manipulation in the visible region, Materials **14**, 5967 (2021).

[11] Y. Kim, T. Choi, G.-Y. Lee, et al., Metasurface folded lens system for ultrathin cameras, Science Advances **10**, eadr2319 (2024).

[12] B. Li, J. Wu, Y. Xu, Y. Zhang, Z. Zhu, N. Yu, and Q. Sun, Nano-3d: Metasurface-based neural depth imaging, arXiv preprint arXiv:2503.15770 (2025).

[13] M. Gopakumar, G.-Y. Lee, et al., Full-colour 3d holographic augmented-reality displays with metasurface waveguides, Nature **629**, 791 (2024).

[14] S. Shah et al., All-dielectric metasurface polarimeter for full stokes polarization measurements, ACS Photonics **9**, 4509 (2022).

[15] M. Faraji-Dana, E. Arbabi, A. Arbabi, et al., Compact folded metasurface spectrometer, Nature Communications 10.1038/s41467-018-06495-5 (2018).

[16] J. G. George et al., Overcoming challenges in fabrication of beam shaping meta-optics using deep ultraviolet lithography for industrial applications, Journal of Optical Microsystems **4**, 041402 (2024).

[17] N. Matsumoto, T. Hamasaki, and K. Shibuya, A dielectric metasurface optical system for augmented reality displays, in CLEO: Science and Innovations (2022).

[18] V. J. Einck et al., Scalable nanoimprint lithography process for manufacturing visible metasurfaces composed of high aspect ratio tio_2 meta-atoms, ACS Photonics 10.1021/acsphotonics.1c00609 (2021).

[19] Paul Scherrer Institute (PSI), Nanoimprint lithography, https://www.psi.ch/en/lnq/nanoimprint-lithography (2025), accessed: 2025-12-12.

[20] J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark, and M. Soljačić, Nanophotonic particle simulation and inverse design using deep learning, Science Advances **4**, eaar4206 (2018).

[21] Phys.org, Metasurfaces designed by a bidirectional deep neural network for generating quantitative field distributions, https://phys.org/news/2023-03-metasurfaces-bidirectional-deep-neural-network.html (2023), accessed: 2025-12-12.

[22] S. Colburn and A. Majumdar, Metasurface optics for full-color computational imaging, Science Advances **4**, eaar2114 (2018).