



Sub-100 Femtosecond All-Optical Modulation Beyond Electron–Phonon Limits

Cite as

Nano-Micro Lett.

(2026) 18:315

Renxian Gao^{1,2}, Jiayu Li³, Xiaoxiang Dong¹, Yonglin He¹, Wenbin Chen³, Peiwen Ren¹, Xiaoyu Zhao², Ming-De Li⁴ ✉, Zhilin Yang¹ ✉

Received: 27 November 2025

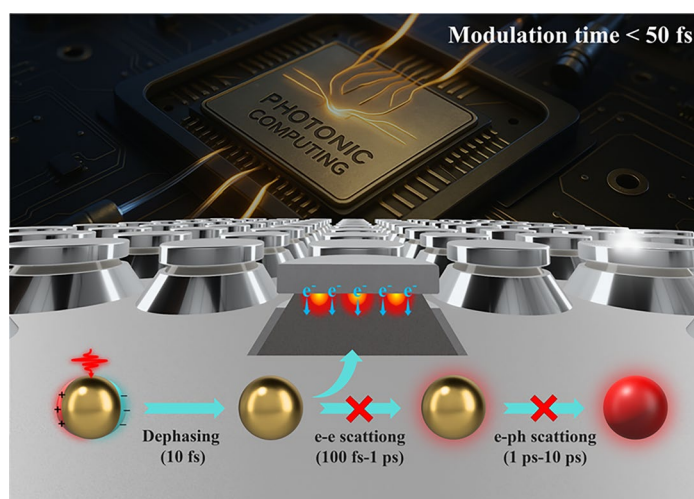
Accepted: 4 March 2026

© The Author(s) 2026

HIGHLIGHTS

- Engineered Ag–Si interfacial plasmonic modes enable deconvolution-verified sub-100 fs all-optical modulation, overcoming the temporal bottleneck imposed by electron–phonon relaxation in conventional plasmonic structures.
- A unified electromagnetic–thermal physical model reveals an interface-governed nonthermal carrier extraction pathway faster than electron–electron and electron–phonon scattering, quantitatively explaining the observed sub-100 fs dynamics.

ABSTRACT Ultrafast all-optical modulators are central to the advancement of next-generation photonic computing and signal-processing systems. However, the intrinsic electron–phonon relaxation bottleneck in plasmonic materials has long constrained modulation speeds to the picosecond regime, hindering the realization of sub-100 fs modulation. Here, we report a metastructured silver–single-crystal silicon nanodisk antenna that delivers experimentally resolved sub-100 fs all-optical modulation. Distinct from conventional planar metal–semiconductor junctions, the nanodisk architecture spatially co-localizes plasmonic energy deposition with the metal–semiconductor transfer boundary within a nanoscale-confined volume. This configuration markedly shortens hot-carrier transport pathways and preferentially activates interfacial carrier extraction during the earliest relaxation stage, thereby establishing an interface-dominated modulation pathway that precedes electron–phonon thermalization. By enabling modulation on timescales comparable to intrinsic electronic response limits, this work establishes a physical foundation for ultrafast photonic modulation, including femtosecond free-space photonic computing architectures, temporal optical gating, and other ultrafast systems constrained by carrier or cavity lifetimes.



KEYWORDS Sub 100 fs all Optical modulation; Silver; Silicon metastructure; Interfacial hot; Carrier dynamics

Renxian Gao, Jiayu Li, and Xiaoxiang Dong contributed equally to this work.

✉ Ming-De Li, mdli@stu.edu.cn; Zhilin Yang, zlyang@xmu.edu.cn

¹ College of Physical Science and Technology, Xiamen University, Xiamen 361005, People's Republic of China

² College of Materials and Environmental Engineering, Hangzhou Dianzi University, Hangzhou 310018, People's Republic of China

³ Key Laboratory for Preparation and Application of Ordered Structural Materials of Guangdong Province, Department of Chemistry, Shantou University, Shantou 515063, People's Republic of China

⁴ School of Physics and Optoelectronic Engineering & Guangdong Provincial Key Laboratory of Sensing Physics and System Integration Applications, Guangdong University of Technology, Guangzhou, Guangdong 510006, People's Republic of China

Published online: 07 April 2026



SHANGHAI JIAO TONG UNIVERSITY PRESS

Springer

1 Introduction

The ever-increasing demand for high-speed optical computing and all-optical signal processing has driven significant advancements in ultrafast optical modulators. These devices play a pivotal role in enhancing data throughput, energy efficiency, and integration density in next-generation photonic circuits [1–6]. Plasmonic all-optical modulators have emerged as promising candidates due to their ability to confine and manipulate light at deep-subwavelength scales, with ultrafast response times and strong potential for high-density photonic integration [7–9]. However, despite the theoretical potential of plasmonic systems, the intrinsic relaxation mechanisms associated with hot-carrier dynamics impose severe limitations on modulation speeds, constraining the realization of all-optical switching below 100 fs [7–10]. Overcoming these fundamental constraints demands innovative plasmonic structure designs that enable efficient carrier extraction and achieve ultrafast dynamics on the tens-of-femtoseconds timescale.

The modulation speed of plasmonic devices is inherently governed by the complex interplay of nonequilibrium carrier dynamics [11–13]. Upon photoexcitation, plasmonic systems exhibit a cascade of ultrafast processes, including plasmon dephasing and hot-electron generation, followed by electron–electron scattering within tens to hundreds of femtoseconds, electron–phonon thermalization over several picoseconds, and ultimately phonon dissipation into the surrounding environment on nanosecond timescales [14–16]. While electron–electron interactions contribute to rapid thermalization within the electronic subsystem, the subsequent coupling to the lattice introduces a bottleneck that significantly limits the recovery time of plasmonic modulators [17, 18]. In recent years, hot-carrier-assisted ultrafast modulation mechanisms have been extensively investigated in plasmonic metasurface and metal–semiconductor or metal–2D-material hybrid nanostructures [10, 12, 14, 19, 20]. In these systems, non-radiative plasmon decay generates nonequilibrium hot carriers within the metal, and interfacial extraction or energy transfer is employed to accelerate the onset of the transient optical response. However, despite the presence of a hundred-femtosecond fast component in many reported transient traces, the dynamics are typically followed by a pronounced sub-picosecond or picosecond relaxation tail governed by electron–phonon coupling and lattice heating.

Such lattice-mediated thermo-optic contributions persist beyond the initial electronic excitation and ultimately constrain both the achievable modulation contrast and the recovery speed. Parallel efforts based on engineered nanocavities and nonlinear optical materials have enabled sub-picosecond switching by leveraging strong optical Kerr effects, yet the ability to achieve sub-100 fs reset times remains challenging due to the persistent influence of electron–phonon interactions [21–25].

To address these challenges, we introduce a metastructured nanophotonic antenna capable of sub-100 fs all-optical modulation, with the underlying concept and design principles illustrated in Fig. 1a, b. The device is built upon a silver–single-crystal silicon nanodisk metastructure antenna (SSDMA), which exploits the interfacial plasmonic confinement at the metal–semiconductor boundary to activate an ultrafast, nonthermal hot-electron transfer pathway. By precisely engineering the near-field distribution at the Ag–Si boundary, the structure spatially co-localizes plasmonic energy deposition with the carrier-transfer interface within a nanoscale-confined volume. This configuration markedly shortens hot-carrier transport pathways and preferentially activates interfacial carrier extraction during the earliest relaxation stage, enabling femtosecond-scale carrier transfer that competes with electronic thermalization and precedes substantial lattice heating. Consequently, the transient optical response is governed by an interface-dominated electronic pathway rather than lattice-mediated thermo-optic effects, allowing modulation to occur on timescales approaching intrinsic electronic response limits. Beyond demonstrating sub-100 fs all-optical modulation, this work establishes an interface-governed physical mechanism for overcoming electron–phonon-limited temporal constraints in plasmonic systems and provides a generalizable design framework for ultrafast photonic components operating near intrinsic carrier or cavity-lifetime limits.

2 Experimental Section

2.1 Materials

Single-crystal silicon wafers ($\langle 100 \rangle$ orientation, resistivity 1–10 Ω cm) were obtained from Zhongsheng Silicon Technology Co., Ltd. (China). Polystyrene (PS) nanospheres with an initial diameter of 500 nm were purchased from Aladdin

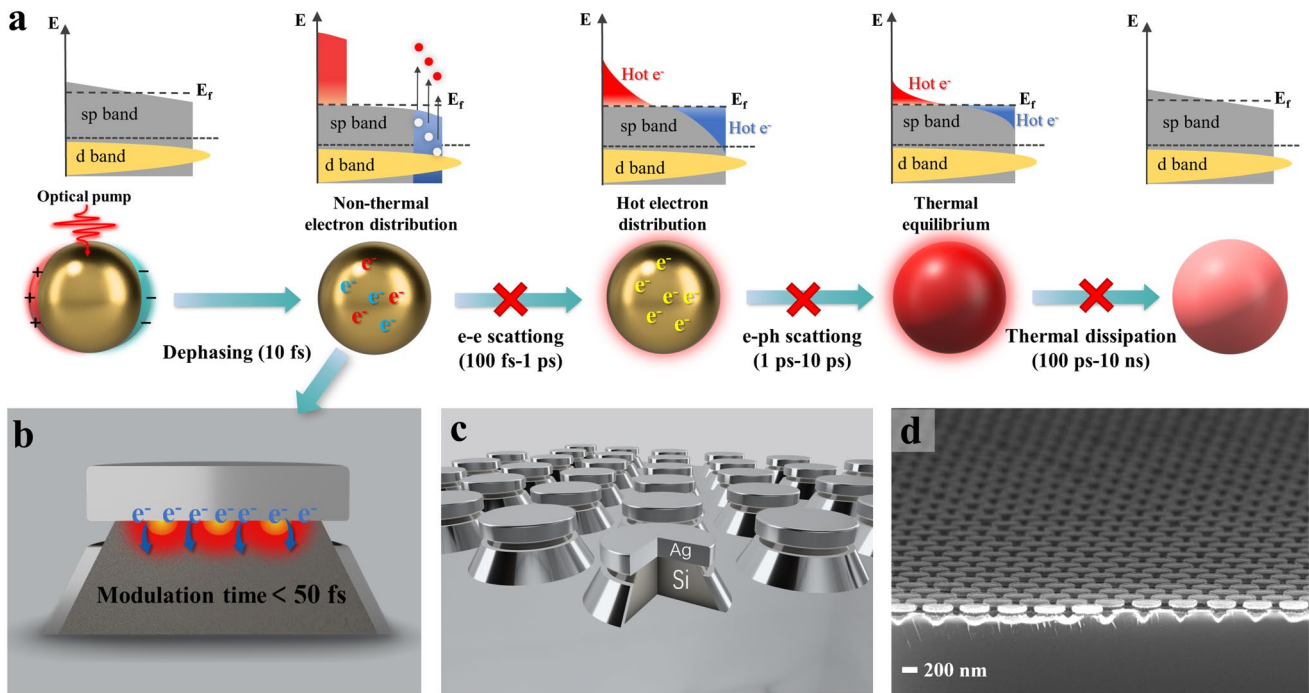


Fig. 1 Schematic diagrams and scanning electron microscopy image. **a** Schematic diagram illustrating the dynamics of surface plasmon relaxation. **b** Schematic diagram illustrates the rapid lossless relaxation pathway of surface plasmon resonance-induced nonthermal equilibrium electrons in SSDMA. **c** Schematic diagram illustrates the structure of the SSDMA, highlighting the complex arrangement of single-crystal silicon nanodisks integrated with overlaid silver nanodisks. **d** Scanning electron microscopy image of the metastructure

Biochemical Technology Co., Ltd. (China). High-purity silver pellets (99.99%) for electron-beam evaporation were supplied by Zhongnuo Advanced Material Co., Ltd. (China). High-temperature-resistant polyimide tape used for template removal was obtained from DuPont™ (USA). Gases used in the RIE processes, including argon (Ar, 99.999%), oxygen (O₂, 99.999%), and sulfur hexafluoride (SF₆, 99.99%), were provided by Hangzhou Jingong Gas Co., Ltd. (China). Deionized (DI) water (18.2 MΩ cm) was produced using a Milli-Q purification system. All chemicals and reagents used for substrate cleaning—acetone (ACS grade), isopropanol (IPA, ≥ 99.7%), and ethanol (≥ 99.5%)—were purchased from Sinopharm Chemical Reagent Co., Ltd. (China).

2.2 Template-Assisted RIE Lithography for SSDMA Metastructure Fabrication

To achieve high-precision fabrication of metal–single-crystal-silicon interface localized-field metastructure arrays, we employed a rigorously controlled reactive-ion etching (RIE) workflow. The overall process is outlined in Fig. S1.

A monolayer of hexagonally close-packed polystyrene (PS) nanospheres was first assembled at the air–water interface to serve as a soft lithographic template. The sphere diameter was subsequently reduced in an RIE system, enabling precise tuning of the structural pitch and aperture. The size-adjusted PS monolayer was then transferred onto the single-crystal silicon wafer and used as a mask in a second RIE etching step (Sentech RIE), during which selective anisotropic etching of the silicon produced a periodic nanopillar topography beneath the sphere array. Progressive extension of the etching duration yielded well-defined and uniform single-crystal silicon nanopillars, as shown in step 2 of Fig. S1. After removing the PS spheres using high-temperature-resistant adhesive film, a clean silicon nanopillar array was obtained (step 3). Finally, the pre-patterned silicon template was introduced into an electron-beam evaporation system (Atemd-500), where a conformal silver nanofilm was deposited at a rate of 4 Å s⁻¹ under high vacuum. Increasing the film thickness gradually decorated the silicon nanopillars with a continuous plasmonic coating, culminating in the formation of the SSDMA architecture.

3 Results and Discussion

3.1 Design Principle and Structural Characterizations

To achieve sub-100 fs ultrafast all-optical modulation across multiple discrete wavelengths, a metastructure array was carefully designed. This array included silver nanodisks placed on top of an array of monocrystalline silicon nanodisks (see Fig. 1b, c). The metastructure array with the metal-monocrystalline silicon interface was fabricated using an inductively coupled plasma etching technique, and further information is provided in Fig. S1. The structural morphology of the metasurface, as depicted in Fig. 1d, demonstrates a high level of conformity, with a schematic illustration shown in Fig. 1c. Specifically, the silicon nanodisk arrangement follows a hexagonal close-packed pattern with a periodicity of 500 nm, while the silver nanodisks exhibit a diameter of 300 nm. The distinctive taper design is characterized by a larger base diameter for the silicon nanotapers than their top diameter, facilitating partial encasement by the silver film. The total height of the nanotapers measures 250 nm, and the silver nanodisks have a thickness of 40 nm. Scanning electron microscopy (SEM) images of a sample from different viewing angles are included in Fig. S2.

3.2 Steady-State Optical Response and Resonance Mode Analysis

Prior to transient optical characterization, we performed steady-state reflectance spectroscopy together with corresponding finite-element electromagnetic simulations of the SSDMA. As shown in Fig. 2a, the experimental reflectance spectrum (red solid line) exhibits good qualitative agreement with the simulated spectrum (purple dashed line). In the experimental data, three well-defined reflectance minima are observed at 514, 572, and 660 nm, with the deepest minimum at 660 nm and progressively weaker minima at shorter wavelengths. In the numerical simulation, the corresponding resonance dips are located at 521, 570, and 687 nm, respectively. The slight discrepancies between experiment and simulation mainly arise from unavoidable fabrication-related deviations and differences in optical constants between the fabricated structures and the idealized simulation model. To elucidate the near-field resonance modes associated with these dips, we

simulated the localized electric-field distributions under normally incident plane waves at the corresponding wavelengths. The results (Fig. 2b) reveal distinct surface plasmon resonance modes, with the localized fields concentrated at the silver–silicon nanodisk interface. This spatial configuration of plasmonic hotspots is expected to shorten the relaxation pathway for hot-electron injection into single-crystal silicon, thereby enabling femtosecond-scale optical modulation. Subsequent transient optical measurements further test this hypothesis and provide direct experimental validation of the proposed mechanism. Moreover, the spatial arrangement of localized fields at resonant wavelengths on the silver–silicon interface offers a pathway toward multi-frequency ultrafast plasmon-driven all-optical modulation. To further substantiate the design of the SSDMA metastructure, we numerically compared the photoinduced electric-field intensity at the silver–silicon interface between the SSDMA and pure silver nanodisk arrays, as shown in Fig. S3.

3.3 Femtosecond Pump–Probe Probing of Interface-Localized Plasmon Dynamics

To assess the transient all-optical modulation capabilities of the metastructure, we performed transient optical experiments utilizing a femtosecond pump–probe system. The experimental configuration included a pulsed laser with a primary wavelength of 800 nm and an 84 fs pulse duration. This laser could produce pump pulses at the specified wavelength following amplification through an optical parametric amplifier. The pump light was accurately directed onto the sample through a lens, leading to excitation with a power density of $51 \mu\text{J cm}^{-2}$. Concurrently, a second beam was directed through a sapphire crystal to generate a super-continuum probe pulse that encompassed wavelengths between 430 and 830 nm. Upon interaction with the sample, the probing light was detected using a spectrometer. The acquired signals were then analyzed with differential circuitry, enabling a comprehensive evaluation of the sample response to transient optical modulation. The schematic of the transient pump–probe optical setup is shown in Fig. S4.

Based on the near-field simulation results illustrated in Fig. 2b, the plasmonic electric fields generated by incident light with wavelengths of 572 and 660 nm are clearly

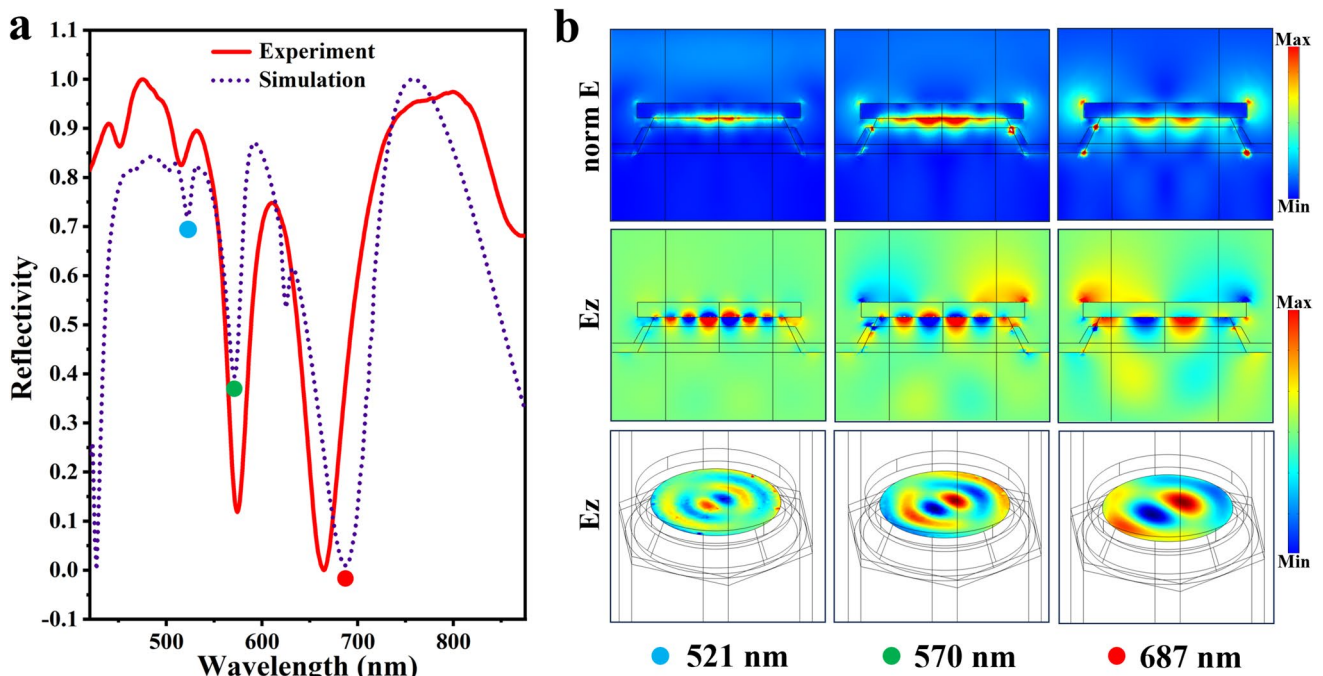


Fig. 2 Steady-State Reflectance Spectrum, Numerical Simulation Spectrum, and Electric Field Distribution. **a** Reflectance spectrum of SSDMA arrays, with the continuous red line representing the measured experimental data and the purple dashed line indicating the corresponding results from finite element electromagnetic simulations. **b** Finite element analysis of electric field distributions induced by monochromatic planar waves. The simulation presents spatially resolved electric field intensities (E-field modulus) and the corresponding Z-component distributions for incident wavelengths of 521 nm, 570 nm, and 687 nm. The top row shows the E-field modulus across the simulation domain, highlighting peak field enhancements. The middle row displays the Z-component of the electric field, while the bottom row provides a three-dimensional view of the Z-component

primarily concentrated at the interface between silver and silicon. Therefore, these wavelengths were selected for pump pulses to induce transient excitation dynamics, in contrast to the 400 nm wavelength pump light that activates surface plasmon resonance modes located further from the silver-silicon interface. Figure 3a–c shows the numerical simulation results illustrating the distribution of the surface plasmon electric field at wavelengths of 572, 660, and 400 nm. Figure 3d–f displays the transient reflection spectra of the sample when excited by pulsed light at the wavelengths shown in Fig. 3a–c. Figure 3d, e reveals multiple ultrafast dynamic signals within the wavelength range of 450 to 700 nm. The observed signals are consistent with the previously discussed steady-state spectra, providing clear evidence of the modulation of the resonance absorption modes of the surface plasmon resonance at the silver-silicon interface for central wavelengths of 514, 572, and 660 nm. Figure 3d, e demonstrates distinct femtosecond fast component transient

signals, as well as slower transient signals typically associated with nonequilibrium electron thermal diffusion induced by surface plasmons, with relaxation times at the picosecond level. To determine the source of the femtosecond fast component signals, a comparative analysis was conducted utilizing a 400 nm wavelength pump laser, as illustrated in Fig. 3f. The femtosecond fast component transient signals observed in Fig. 3d, e are lacking in this figure. Additionally, Fig. 3c reveals that when the sample is subjected to a 400 nm incident pump laser, the localized plasmonic field is predominantly restricted to the edges of the silver nanodisks, with hotspots located away from the silver-silicon interface. The findings from these comparative experiments confirm that the femtosecond fast components observed in Fig. 3d, e are related to the ultrafast dynamics of surface plasmons at the interface between silver and silicon.

The ultrafast dynamics of surface plasmon resonance typically span from a few femtoseconds to several hundred

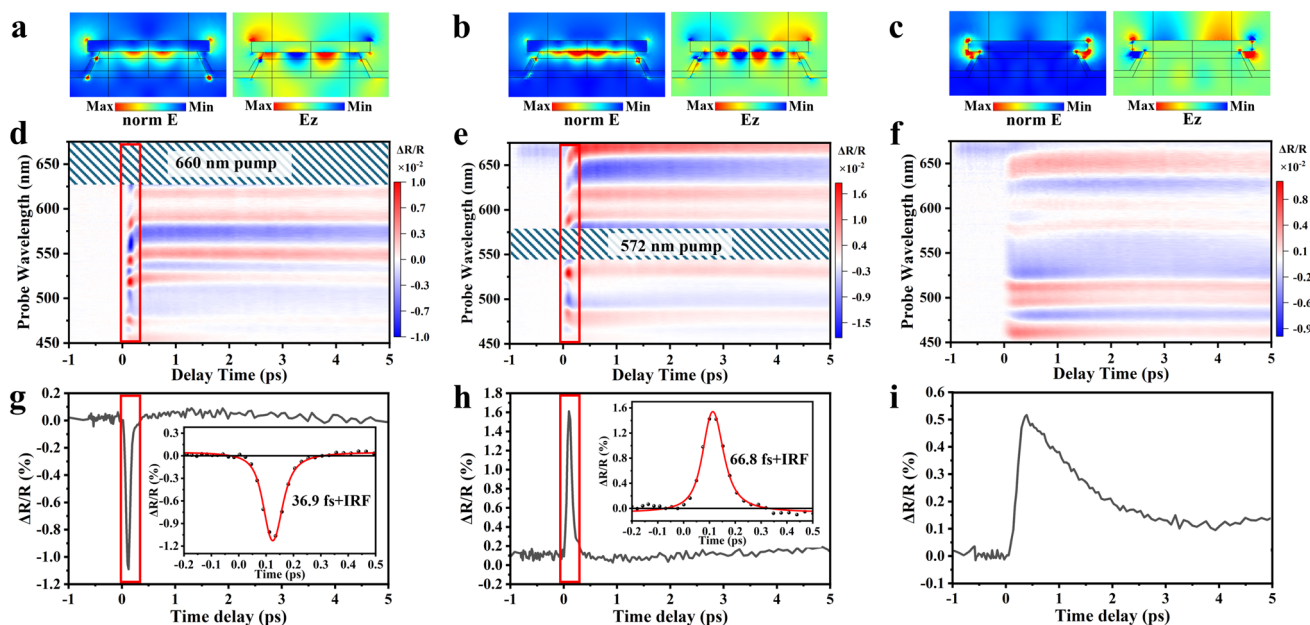


Fig. 3 Transient optical measurement. **a–c** Near-field numerical simulation results of the sample corresponding to incident laser wavelengths of 660, 572, and 400 nm. **d–f** Transient reflection spectra of the sample in the range of 450–675 nm with pump laser wavelengths of 660, 572, and 400 nm, where the maximum delay time is 5 ps. **g** Transient dynamic spectral lines with a pump wavelength of 660 nm and a probe wavelength of 559 nm, the solid red line in the inset corresponds to a fit considering the convolution between IRF and the Lorentz function. **h** Transient dynamic spectral lines with a pump wavelength of 572 nm and a probe wavelength of 589 nm, the solid red line in the inset corresponds to a fit considering the convolution between IRF and the Lorentz function. **i** Transient dynamic spectral lines with a pump wavelength of 400 nm and a probe wavelength of 514 nm

picoseconds, influencing the temperature and density of free carriers and phonons in metals and thereby modifying their dielectric function, which leads to pronounced nonlinear optical Kerr effects [26–28]. However, the rate of energy transfer from hot electrons to phonons in metals is intrinsically slow, restricting the modulation speed of conventional plasmonic nanostructures to the picosecond regime [29, 30]. To overcome this temporal bottleneck, we analyzed in detail how the spatial distribution of plasmonic modes governs the transient carrier dynamics. Comparing the hotspot distributions in Fig. 3a, b with the resonance-field localization in Fig. 3c reveals a strong correlation between the appearance of sub-100 fs transient components in Fig. 3d, e and the effective confinement of the plasmonic field at the silver–silicon interface. Specifically, the interface-localized plasmonic resonance mode promotes efficient Landau damping, generating high-energy nonequilibrium hot electrons [14–16]. Importantly, hot-carrier collection across a Schottky junction follows the established three-step picture (generation–transport–collection), and is constrained by

both the energetic condition $E - E_F > \Phi_B$ and the interfacial transmission selectivity [17, 31–33]. For the Ag–Si junction, the Schottky barrier height lies in the sub-eV range (typically $\Phi_B \sim 0.6 - 0.8$ eV for n-type Si). Under visible excitation, the Landau-damping-generated nonthermal distribution contains a high-energy tail that can satisfy $E - E_F > \Phi_B$, making interfacial charge transfer into Si energetically accessible [34, 35]. Because the plasmonic hotspot is co-localized with the Ag–Si interface, the hot-carrier transport distance is minimized, allowing carriers to reach the boundary before substantial electron–phonon energy transfer develops and thereby enabling a sub-100 fs interfacial dielectric perturbation that precedes the conventional electron–phonon-mediated picosecond pathway. Consistent with this picture, all three resonance modes in Fig. 2b (514, 572, and 660 nm) exhibit strong Ag–Si interface localization, so excitation at either 572 or 660 nm can trigger an interface-mediated dielectric perturbation that simultaneously influences the interface-localized resonances, as manifested by the pronounced femtosecond components highlighted in Fig. 3d, e.

In contrast, when the plasmonic excitation is localized away from the Ag–Si interface (as in Fig. 3c), the generated hot electrons cannot efficiently transfer into the semiconductor. Instead, they dissipate energy through electron–electron and electron–phonon scattering within the metal, producing slower picosecond-scale relaxation dynamics (Fig. 3f, i). This scattering-dominated pathway fundamentally limits the achievable modulation speed. From a design perspective, the plasmonic configurations in Fig. 3a, b dramatically enhance the spatial overlap between near-field hotspots and the carrier-injection region. Such co-localization minimizes the transport distance of hot electrons and greatly improves injection efficiency, thereby enabling multi-wavelength, sub 100 fs modulation of the transient optical response.

To precisely ascertain the all-optical modulation speed of the sample, we analyzed the transient dynamics curves at probe wavelengths of 559 and 572 nm extracted from Fig. 3d and e, as depicted in Fig. 3g, h. To quantitatively extract the intrinsic ultrafast response, the measured transient signals were modeled as the convolution of an independently measured Gaussian instrument response function (IRF) and a single exponential sample response. The IRF width was fixed to the cross-correlation FWHM of 100.8 fs, ensuring that no instrument-dependent parameter was allowed to float during fitting. Under this rigorously constrained deconvolution procedure, the probe wavelength resolved dynamics yield characteristic time constants of 37 ± 9 and 67 ± 14 fs at 559 and 589 nm, respectively. The corresponding transient kinetic traces with error bars are presented in Fig. S5, and detailed fitting procedures and numerical validation are provided in Section S6. Because the fastest extracted timescale approaches the IRF width, it represents an IRF-conditioned estimate rather than a strictly resolved decay constant. The intrinsic response is therefore conservatively assigned to a sub-100-fs regime. Despite this limitation, the fits consistently show that the Ag–Si interfacial hot-electron-induced permittivity perturbation occurs well before electron–phonon thermalization, confirming sub-100 fs dynamics substantially faster than conventional lattice-mediated relaxation. In addition to the response time, we quantitatively evaluated the modulation depth and switching contrast. At a pump power density of $51 \mu\text{J cm}^{-2}$, the peak transient reflectance changes $(\Delta R/R)_{peak}$ reach 1.1% and 1.6% at the

559 and 572 nm probe wavelengths, respectively (Fig. 3g, h). We then assessed the switching contrast using an on–off ratio metric. As shown in Fig. 2b, the maximum on–off ratio exceeds 100 within the first 5 ps after excitation. The on–off ratio is defined as:

$$\text{On-off ratio} = \frac{|(\Delta R/R)_{peak}|}{|(\Delta R/R)_{off}|} \quad (1)$$

where $(\Delta R/R)_{off}$ is defined as the maximum residual modulation magnitude within the 0.5–5 ps delay window after the initial peak. To the best of our knowledge, this value is among the highest reported for plasmonic ultrafast all-optical modulation systems [12, 14, 19–21].

3.4 Electron-Blocking Validation and Power-Dependent Ultrafast Dynamics

To provide additional evidence supporting the hypothesis that the femtosecond all-optical modulation signal is a consequence of ultrafast hot-electron injection induced by surface plasmons at the silver-silicon interface, we designed a comparative sample incorporating a 30 nm thick aluminum oxide insulating layer. This insulating layer was precisely positioned between the silver nanodisks and the monocrystalline silicon wafer utilizing an atomic layer deposition system, with the specific purpose of impeding the transfer of hot electrons from the silver surface to the monocrystalline silicon. The schematic of the design is shown in Fig. 4a, b, illustrating the relaxation pathways of hot carriers in samples both with and without an aluminum oxide insulating layer. Figure 4c displays the transient dynamics curves obtained from ultrafast spectroscopy measurements of the silver-silicon nanodisk samples with and without the aluminum oxide insulation layer. The inclusion of the aluminum oxide layer hinders the rapid transfer of hot electrons to the conduction band of monocrystalline silicon through the silver-monocrystalline silicon interface. As a result, there is a noticeable predominance of slower component dynamics, specifically in the heat transfer process between hot electrons and phonons, which impacts the modulation speed of the sample. This observation is clearly demonstrated in the transient spectra depicted in Fig. 4c, where the rate of recovery within the 0–5 ps timeframe is significantly slower for the sample with

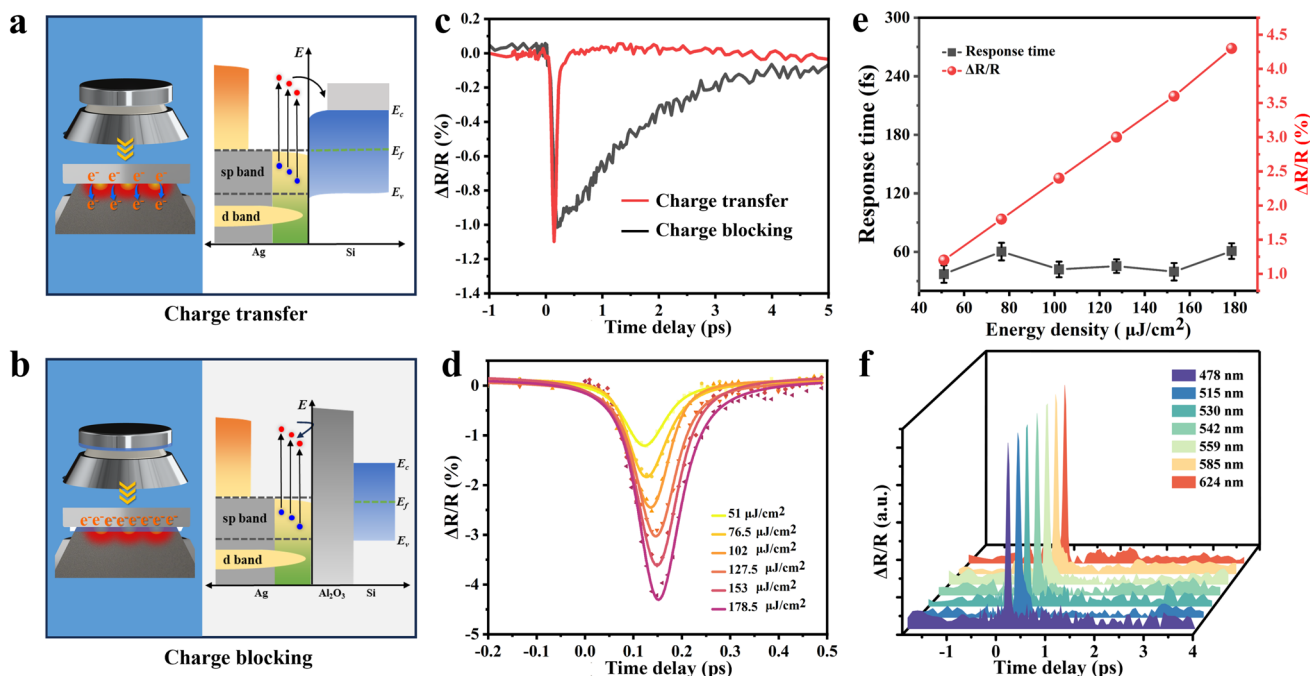


Fig. 4 Schematic illustration of the hot electron relaxation and transient optical measurements. **a, b** Schematic illustration of the hot electron relaxation pathways in samples with and without an alumina insulating layer. **c** Transient dynamic spectra of SSDMA with and without an alumina insulating layer, excited by a 660 nm pump laser. **d** Transient dynamic spectra of SSDMA samples at different pump power densities, with a pump wavelength of 660 nm. **e** Power dependence of modulation depth and speed. Error bars show fitting uncertainties. **f** Transient dynamic spectra at various probe wavelengths, extracted from the dataset presented in Fig. 3d

the alumina insulating layer than for that without it. Overall, these comparative experiments provide compelling evidence that sub-100 fs modulation of SSDMA is primarily driven by the ultrafast transfer of nonequilibrium electrons at the silver-single crystal silicon interface. This transfer induces rapid changes in the interface dielectric constant, enabling sub-100 fs ultrafast modulation of the probe light.

The power dependence of the transient all-optical modulation performance of the SSDMA was further investigated, as shown in Fig. 4d. Under 660 nm excitation, the pump fluence was varied from 51 to 178.5 $\mu\text{J cm}^{-2}$. The transient responses at different pump fluences were analyzed using the Gaussian–Lorentzian deconvolution described above to extract intrinsic modulation times and their dependence on modulation depth (Fig. 4e). The data show that the all-optical modulation depth of the sample increases linearly from 1.2% to 4.3%, indicating that the ultrafast modulation signal does not stem from the nonlinear effects of silicon. Additionally, as the pump

power increases, the modulation speed remains within an ultrafast timescale of less than 100 fs. To exclude potential resonance-specific artifacts, transient dynamics were further extracted at multiple probe wavelengths (Fig. 4f). Upon resonant excitation of the Ag–Si interfacial plasmon at 660 nm, Landau-damping induced interfacial carrier extraction transiently renormalizes the interfacial permittivity, causing all probe modes with near fields co-localized at this boundary to exhibit consistent sub-100 fs dynamics. These results further indicate that the sub-100 fs modulation is governed by a unified interfacial carrier dynamics mechanism.

3.5 Ultrafast Energy-Transfer Simulation Based on Coupled Wave-Optics and Two-Temperature Modeling

To further verify the activation of the lossless ultrafast energy-transfer channel, we established a coupled multiphysics

simulation framework that integrates the COMSOL Wave Optics Module with the general form of the two-temperature model (TTM). This approach enables simultaneous determination of the spatial distribution of electromagnetic absorption within the SSDMA metastructure and the subsequent temporal evolution of the electron and lattice temperatures. A full description of the numerical procedure is provided in Section S7. The electromagnetic loss density $Q(x, y, z)$ was first obtained from the Wave Optics Module under excitation wavelengths of 660 and 400 nm. This spatially resolved loss profile was then modulated by the temporal Gaussian envelope of the pump pulse to generate a time-dependent volumetric heat source, $S(x, y, z, t) = Q(x, y, z)I(t)$, which serves as the excitation term in the TTM [14, 35–39]. The transient non-equilibrium dynamics were computed by solving the coupled TTM equations using the finite-element method. The electron temperature T_e and lattice temperature T_l evolve according to:

$$C_e(T_e)\frac{\partial T_e}{\partial t} = \nabla \cdot (\kappa_e \nabla T_e) - G_{ep}(T_e - T_l) - G_{es}(T_e - T_s) + S(x, y, z, t), \quad (2a)$$

$$C_l\frac{\partial T_l}{\partial t} = \nabla \cdot (\kappa_l \nabla T_l) + G_{ep}(T_e - T_l), \quad (2b)$$

where $C_e = \gamma T_e$ and C_l are the electron and lattice heat capacities, κ_e and κ_l are the corresponding thermal conductivities, G_{ep} is the electron–phonon coupling constant of Ag, and G_{es} is an effective volumetric electron–silicon interfacial coupling coefficient ($\text{W m}^{-3}\text{K}^{-1}$) that parameterizes ultrafast interfacial energy extraction from the Ag electron subsystem into silicon. In the present simulations, $G_{es} = 2 \times 10^{16} \text{ W m}^{-3}\text{K}^{-1}$, obtained by converting literature-reported interfacial electron–substrate conductances for metal/Si interfaces into an equivalent volumetric form using the characteristic interfacial localization length in the SSDMA geometry [40, 41], see Supplementary Materials Sect. 7 for details. In this formulation, the silicon substrate is treated as an effective thermal reservoir maintained at the ambient temperature $T_s = 293 \text{ K}$. This approximation reflects the fact that, on the sub-100-fs timescale resolved in the present pump-probe measurements, the dominant energy-transfer pathway is interfacial extraction from Ag electrons into a much larger silicon heat capacity. The coupling term $G_{es}(T_e - T_s)$ therefore represents net energy diffusion from the Ag electron subsystem into a semi-infinite silicon substrate, without explicitly resolving the transient temperature evolution within silicon. This reduced description isolates

the experimentally relevant interfacial energy-transfer channel while avoiding additional unconstrained degrees of freedom (see Section S7 for details).

Figure 5a–d presents the evolution of the electronic temperature under 660 and 400 nm excitation during the first 5 ps. The results reveal that the electron subsystem within the silver nanostructure remains at an elevated temperature even at approximately 750 fs when only intrinsic electron–phonon coupling is considered, indicating that conventional plasmonic relaxation pathways intrinsically limit the attainable modulation speed. To overcome this bottleneck, the interfacial coupling term $G_{es}(T_e - T_s)$ was incorporated to account for the ultrafast transfer of hot electrons into the silicon substrate. As shown in Fig. 5e, f, inclusion of this interfacial channel substantially accelerates the cooling of the electron subsystem and yields dynamics that faithfully reproduce the experimentally observed sub-100 fs response. The simulations performed under 660 nm excitation also exhibit excellent agreement with the experimental transient results, and full datasets are provided in Fig. S6.

4 Conclusions

In summary, we demonstrated a metastructured plasmonic antenna capable of experimentally resolved sub-100 fs all-optical modulation, which surpasses the electron–phonon relaxation bottleneck that limits conventional plasmonic devices. By engineering the spatial confinement of interfacial plasmonic modes, a nonthermal hot-electron extraction pathway is activated that competes with electronic thermalization and precedes substantial lattice heating. This interface-governed ultrafast energy-transfer regime is quantitatively supported by femtosecond transient spectroscopy and a coupled electromagnetic-thermal physical model, establishing a clear mechanistic origin for the observed sub-100 fs dynamics. More broadly, this work identifies interfacial carrier extraction as an effective strategy for overcoming intrinsic electron–phonon-limited temporal constraints in plasmonic nanostructures and provides a physically grounded framework for the design of ultrafast optoelectronic components operating near intrinsic carrier-dynamics limits.

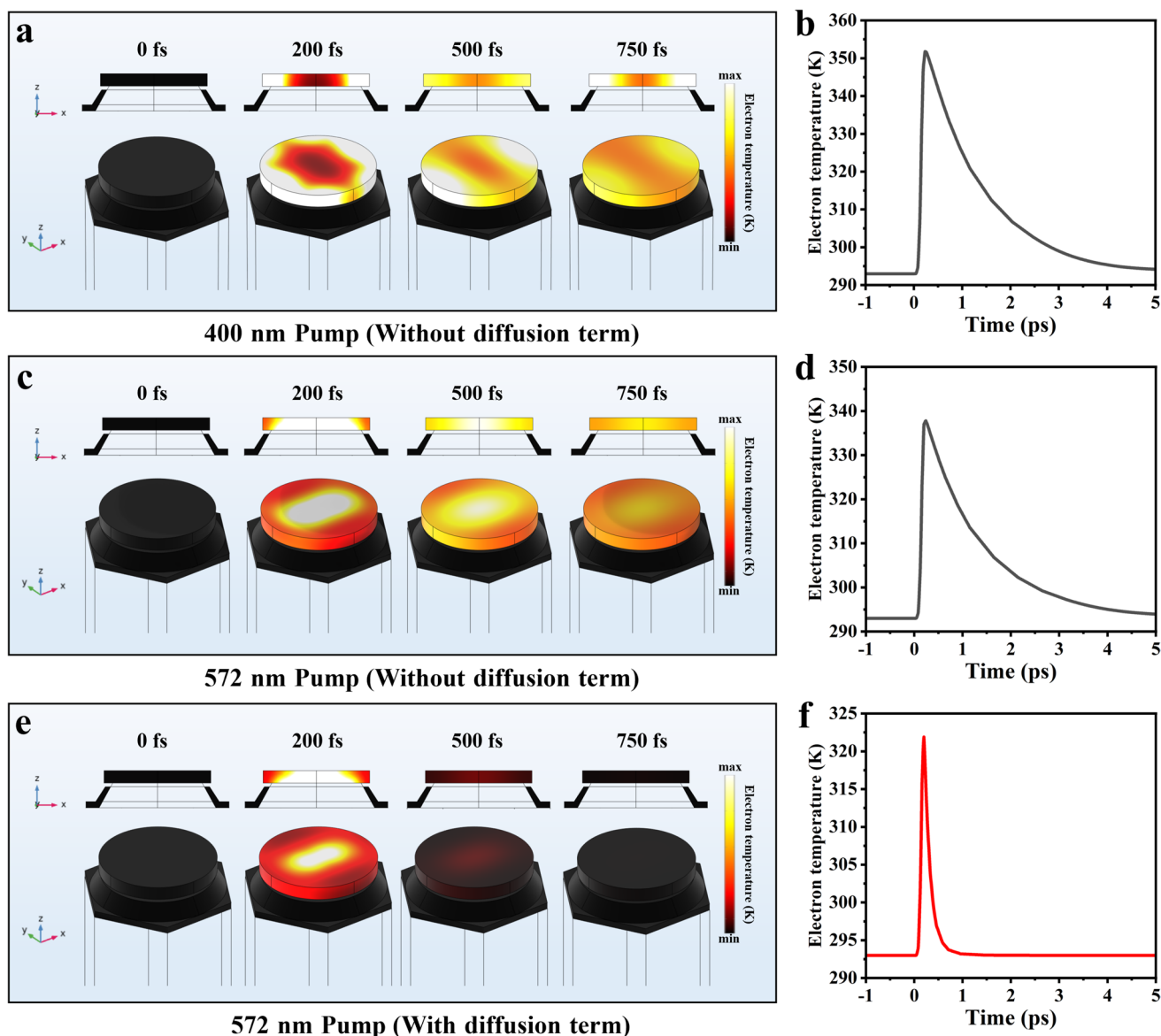


Fig. 5 Time-resolved thermal electron temperature simulation. **a** Electron temperature of the SSDMA structure at different time points under 400 nm pump wavelength excitation. **b** Time-resolved temperature spectrum of hot electrons under 400 nm pump wavelength excitation. **c** Electron temperature of the SSDMA structure at different time points under 572 nm pump wavelength excitation. **d** Time-resolved temperature spectrum of hot electrons under 572 nm pump wavelength excitation. **e** Electron temperature of the SSDMA structure at different time points under 572 nm pump wavelength excitation, including the diffusion term. **f** Time-resolved temperature spectrum of hot electrons, including the diffusion term, under 572 nm pump wavelength excitation

Acknowledgements This work was supported by Scientific and Technological Innovation 2030-"Quantum Communication and quantum computer" Major Project (2023ZD0300304), National Natural Science Foundation of China (12174324, 1240040904, 12404358, 92161118 and 22273057) and the National Key Research and Development Program of China (2021YFA1201502). The Natural Science Foundation of Zhejiang Province (LQN25A040010). Hong Kong Joint Laboratory for

Preparation and Application of Ordered Structural Materials of Guangdong Province (2023B1212120011).

Author's Contribution ZY, ML and RG conceived the idea RG manufactured the samples and coordinated overall research effort RG, JL and XD performed the pump-probe experiment RG and YH performed the numerical simulations WC, PR and XZ, supervised the experimental work RG, ML and ZY, wrote the manuscript All authors discussed the results and commented on the manuscript.

Declarations

Conflict of interest The authors declare no interest conflict. They have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40820-026-02166-z>.

References

1. P.L. McMahon, The physics of optical computing. *Nat. Rev. Phys.* **5**(12), 717–734 (2023). <https://doi.org/10.1038/s42254-023-00645-5>
2. J. Hu, D. Mengu, D.C. Tzarouchis, B. Edwards, N. Eng-heta et al., Diffractive optical computing in free space. *Nat. Commun.* **15**(1), 1525 (2024). <https://doi.org/10.1038/s41467-024-45982-w>
3. B.J. Shastri, A.N. Tait, T. Ferreira de Lima, W.H.P. Pernice, H. Bhaskaran et al., Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* **15**(2), 102–114 (2021). <https://doi.org/10.1038/s41566-020-00754-y>
4. Q. Guo, R. Sekine, L. Ledezma, R. Nehra, D.J. Dean et al., Femtojoule femtosecond all-optical switching in lithium niobate nanophotonics. *Nat. Photonics* **16**(9), 625–631 (2022). <https://doi.org/10.1038/s41566-022-01044-5>
5. J. Xie, J. Yan, H. Han, Y. Zhao, M. Luo et al., Photonic chip based on ultrafast laser-induced reversible phase change for convolutional neural network. *Nano-Micro Lett.* **17**(1), 179 (2025). <https://doi.org/10.1007/s40820-025-01693-5>
6. T.Y. Teo, X. Ma, E. Pastor, H. Wang, J.K. George et al., Programmable chalcogenide-based all-optical deep neural networks. *Nanophotonics* **11**(17), 4073–4088 (2022). <https://doi.org/10.1515/nanoph-2022-0099>
7. J. Yang, X. Zhang, Optical fiber delivered ultrafast plasmonic optical switch. *Adv. Sci.* **8**(10), 2100280 (2021). <https://doi.org/10.1002/advs.202100280>
8. H. Wang, Z. Hu, J. Deng, X. Zhang, J. Chen et al., All-optical ultrafast polarization switching with nonlinear plasmonic metasurfaces. *Sci. Adv.* **10**(8), eadk3882 (2024). <https://doi.org/10.1126/sciadv.adk3882>
9. S. Kunwar, S. Pandit, J.-H. Jeong, J. Lee, Improved photoresponse of UV photodetectors by the incorporation of plasmonic nanoparticles on GaN through the resonant coupling of localized surface plasmon resonance. *Nano-Micro Lett.* **12**(1), 91 (2020). <https://doi.org/10.1007/s40820-020-00437-x>
10. M. Ono, M. Hata, M. Tsunekawa, K. Nozaki, H. Sumikura et al., Ultrafast and energy-efficient all-optical switching with graphene-loaded deep-subwavelength plasmonic waveguides. *Nat. Photonics* **14**(1), 37–43 (2020). <https://doi.org/10.1038/s41566-019-0547-7>
11. K.F. MacDonald, Z.L. Sámsón, M.I. Stockman, N.I. Zheludev, Ultrafast active plasmonics. *Nat. Photonics* **3**(1), 55–58 (2009). <https://doi.org/10.1038/nphoton.2008.249>
12. A. Basiri, M.Z.E. Rafique, J. Bai, S. Choi, Y. Yao, Ultrafast low-pump fluence all-optical modulation based on graphene-metal hybrid metasurfaces. *Light. Sci. Appl.* **11**, 102 (2022). <https://doi.org/10.1038/s41377-022-00787-8>
13. Q. Guo, Y. Yao, Z.-C. Luo, Z. Qin, G. Xie et al., Universal near-infrared and mid-infrared optical modulation for ultrafast pulse generation enabled by colloidal plasmonic semiconductor nanocrystals. *ACS Nano* **10**(10), 9463–9469 (2016). <https://doi.org/10.1021/acs.nano.6b04536>
14. A. Schirato, M. Maiuri, A. Toma, S. Fugattini, R. Proietti Zaccaria et al., Transient optical symmetry breaking for ultrafast broadband dichroism in plasmonic metasurfaces. *Nat. Photonics* **14**(12), 723–727 (2020). <https://doi.org/10.1038/s41566-020-00702-w>
15. M. Kauranen, A.V. Zayats, Nonlinear plasmonics. *Nat. Photonics* **6**(11), 737–748 (2012). <https://doi.org/10.1038/nphoton.2012.244>
16. C. Clavero, Plasmon-induced hot-electron generation at nanoparticle/metal-oxide interfaces for photovoltaic and photocatalytic devices. *Nat. Photonics* **8**(2), 95–103 (2014). <https://doi.org/10.1038/nphoton.2013.238>
17. M.L. Brongersma, N.J. Halas, P. Nordlander, Plasmon-induced hot carrier science and technology. *Nat. Nanotechnol.* **10**(1), 25–34 (2015). <https://doi.org/10.1038/nnano.2014.311>
18. J.G. Liu, H. Zhang, S. Link, P. Nordlander, Relaxation of plasmon-induced hot carriers. *ACS Photonics* **5**(7), 2584–2595 (2018). <https://doi.org/10.1021/acsphotonics.7b00881>
19. M. Taghinejad, H. Taghinejad, Z. Xu, K.-T. Lee, S.P. Rodrigues et al., Ultrafast control of phase and polarization of light expedited by hot-electron transfer. *Nano Lett.* **18**(9), 5544–5551 (2018). <https://doi.org/10.1021/acs.nanolett.8b01946>
20. M. Taghinejad, H. Taghinejad, Z. Xu, Y. Liu, S.P. Rodrigues et al., Hot-electron-assisted femtosecond all-optical modulation in plasmonics. *Adv. Mater.* **30**(9), 1704915 (2018). <https://doi.org/10.1002/adma.201704915>
21. H. Harutyunyan, A.B.F. Martinson, D. Rosenmann, L.K. Khorashad, L.V. Besteiro et al., Anomalous ultrafast dynamics of hot plasmonic electrons in nanostructures with hot spots. *Nat.*

- Nanotechnol. **10**(9), 770–774 (2015). <https://doi.org/10.1038/nnano.2015.165>
22. L.H. Nicholls, T. Stefaniuk, M.E. Nasir, F.J. Rodríguez-Foruña, G.A. Wurtz et al., Designer photonic dynamics by using non-uniform electron temperature distribution for on-demand all-optical switching times. *Nat. Commun.* **10**, 2967 (2019). <https://doi.org/10.1038/s41467-019-10840-7>
 23. G.A. Wurtz, R. Pollard, W. Hendren, G.P. Wiederrecht, D.J. Gosztola et al., Designed ultrafast optical nonlinearity in a plasmonic nanorod metamaterial enhanced by nonlocality. *Nat. Nanotechnol.* **6**(2), 107–111 (2011). <https://doi.org/10.1038/nnano.2010.278>
 24. M. Taghinejad, W. Cai, All-optical control of light in micro- and nanophotonics. *ACS Photonics* **6**(5), 1082–1093 (2019). <https://doi.org/10.1021/acsphotonics.9b00013>
 25. H. Liu, W. Dong, H. Wang, L. Lu, Q. Ruan et al., Rewritable color nanoprints in antimony trisulfide films. *Sci. Adv.* **6**(51), eabb7171 (2020). <https://doi.org/10.1126/sciadv.abb7171>
 26. G.V. Hartland, Optical studies of dynamics in noble metal nanostructures. *Chem. Rev.* **111**(6), 3858–3887 (2011). <https://doi.org/10.1021/cr1002547>
 27. G. Li, S. Zhang, T. Zentgraf, Nonlinear photonic metasurfaces. *Nat. Rev. Mater.* **2**(5), 17010 (2017). <https://doi.org/10.1038/natrevmats.2017.10>
 28. M. Abb, Y. Wang, C.H. de Groot, O.L. Muskens, Hotspot-mediated ultrafast nonlinear control of multifrequency plasmonic nanoantennas. *Nat. Commun.* **5**, 4869 (2014). <https://doi.org/10.1038/ncomms5869>
 29. R. Rouxel, M. Diego, F. Medeghini, P. Maioli, F. Rossella et al., Ultrafast thermo-optical dynamics of a single metal nano-object. *J. Phys. Chem. C* **124**(28), 15625–15633 (2020). <https://doi.org/10.1021/acs.jpcc.0c04709>
 30. R. Gao, Y. He, D. Zhang, G. Sun, J.-X. He et al., Gigahertz optoacoustic vibration in Sub-5 nm tip-supported nano-optomechanical metasurface. *Nat. Commun.* **14**(1), 485 (2023). <https://doi.org/10.1038/s41467-023-36127-6>
 31. G. Tagliabue, A.S. Jermyn, R. Sundararaman, A.J. Welch, J.S. DuChene et al., Quantifying the role of surface plasmon excitation and hot carrier transport in plasmonic devices. *Nat. Commun.* **9**(1), 3394 (2018). <https://doi.org/10.1038/s41467-018-05968-x>
 32. G. Liu, Y. Lou, Y. Zhao, C. Burda, Directional damping of plasmons at metal-semiconductor interfaces. *Acc. Chem. Res.* **55**(13), 1845–1856 (2022). <https://doi.org/10.1021/acs.accounts.2c00001>
 33. C. Jia, X. Li, N. Xin, Y. Gong, J. Guan et al., Interface-engineered plasmonics in metal/semiconductor heterostructures. *Adv. Energy Mater.* **6**(17), 1600431 (2016). <https://doi.org/10.1002/aenm.201600431>
 34. R. Sundararaman, P. Narang, A.S. Jermyn, W.A. Goddard III., H.A. Atwater, Theoretical predictions for hot-carrier generation from surface plasmon decay. *Nat. Commun.* **5**, 5788 (2014). <https://doi.org/10.1038/ncomms6788>
 35. A.M. Brown, R. Sundararaman, P. Narang, W.A. Goddard, H.A. Atwater, *Ab initio* phonon coupling and optical response of hot electrons in plasmonic metals. *Phys. Rev. B* **94**(7), 075120 (2016). <https://doi.org/10.1103/physrevb.94.075120>
 36. A. Block, M. Liebel, R. Yu, M. Spector, Y. Sivan et al., Tracking ultrafast hot-electron diffusion in space and time by ultrafast thermomodulation microscopy. *Sci. Adv.* **5**(5), eaav8965 (2019). <https://doi.org/10.1126/sciadv.aav8965>
 37. M. Bonn, D.N. Denzler, S. Funk, M. Wolf, S.-S. Wellershoff et al., Ultrafast electron dynamics at metal surfaces: competition between electron-phonon coupling and hot-electron transport. *Phys. Rev. B* **61**(2), 1101–1105 (2000). <https://doi.org/10.1103/physrevb.61.1101>
 38. J. Bin Lee, K. Kang, S.H. Lee, Comparison of theoretical models of electron-phonon coupling in thin gold films irradiated by femtosecond pulse lasers. *Mater. Trans.* **52**(3), 547–553 (2011). <https://doi.org/10.2320/matertrans.m2010396>
 39. P. Bresson, J.-F. Bryche, M. Besbes, J. Moreau, P.-L. Karsenti et al., Improved two-temperature modeling of ultrafast thermal and optical phenomena in continuous and nanostructured metal films. *Phys. Rev. B* **102**(15), 155127 (2020). <https://doi.org/10.1103/physrevb.102.155127>
 40. D.G. Cahill, W.K. Ford, K.E. Goodson, G.D. Mahan, A. Majumdar et al., Nanoscale thermal transport. *J. Appl. Phys.* **93**(2), 793–818 (2003). <https://doi.org/10.1063/1.1524305>
 41. R. Stoner, H. Maris, Kapitza conductance and heat flow between solids at temperatures from 50 to 300 K. *Phys. Rev. B* **48**(22), 16373–16387 (1993). <https://doi.org/10.1103/physrevb.48.16373>
- Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.