Nanofocusing in a metal–insulator–metal gap plasmon waveguide with a three-dimensional linear taper

Hyuck Choo[1†]*, Myung-Ki Kim[1,2†], Matteo Staffaroni[2], Tae Joon Seok[2], Jeffrey Bokor[2,3], Stefano Cabrini[3], P. James Schuck[3], Ming C. Wu[2] and Eli Yablonovitch[2]*

The development of techniques for efficiently confining photons on the deep sub-wavelength spatial scale will revolutionize scientific research and engineering practices. The efficient coupling of light into extremely small nanofocusing devices has been a major challenge in on-chip nanophotonics because of the need to overcome various loss mechanisms and the on-chip nanofabrication challenges. Here, we demonstrate experimentally the achievement of highly efficient nanofocusing in an Au–SiO$_2$–Au gap plasmon waveguide using a carefully engineered three-dimensional taper. The dimensions of the SiO$_2$ layer, perpendicular to the direction of wave propagation, taper linearly below 100 nm. Our simulations suggest that the three-dimensional linear-tapering approach could focus 830 nm light into a 2×5 nm^2 area with ≤ 3 dB loss and an intensity enhancement of 3.0×10^4. In a two-photon luminescence measurement, our device achieved an intensity enhancement of 400 within a 14×80 nm^2 area, and a transmittance of 74%.

Nanofocusing processes are essential for maximizing the performances of a new class of nanoscale optical/photonic devices, in particular on-chip components such as ultrafast, highly efficient nanoscale on-chip lasers[1–4], detectors[5,6], modulators[7,8], light-emitting diodes[9–11] and single-photon sources[12,13]. The utilization of surface plasmon polaritons (SPPs)—propagating electron oscillations along the interface between a metal and a dielectric medium—has been proposed as a feasible mechanism for overcoming the diffraction limits of conventional focusing devices[14,15]. To achieve efficient on-chip nanofocusing, various SPP structures with linearly tapering designs, such as metallic V-trenches[16], as well as tapered insulator–metal (IM) and metal–insulator–metal (MIM) structures[17–23], have been theoretically proposed to minimize the losses that occur during nanofocusing processes (by searching for optimal tapering angles). Some of these have been demonstrated experimentally[16,17,19,23].

However, because of the challenges involved in on-chip nanofabrication, most experimental studies have been limited to the demonstration of two-dimensional nanofocusing processes that compress energy in one dimension normal to the propagation axis, and form a focused line instead of a point. Thus far, the best experimentally demonstrated two-dimensional nanofocusing efficiency is ~3 dB (50%) through a 45 nm × 2 μm slit aperture[23]. To further increase the field enhancement and localization, one must consider three-dimensional nanofocusing, in which light is compressed in two dimensions normal to the propagation axis. However, current three-dimensional nanofocusing devices— tapered metal-coated fibres and metal tips—are stand-alone devices[24–34] and their on-chip implementation is challenging. In another recent three-dimensional approach, researchers reported a three-dimensional plasmonic dimple structure for nanofocusing on the vertical edge of a chip[35]; however, the work lacks quantitative data.

In this work, we demonstrate the design, fabrication and characterization of a highly efficient, on-chip three-dimensional MIM nanoplasmonic photon compressor (3D NPC) that can be readily integrated with other on-chip nanophotonic components. Among the proposed nanofocusing approaches, we have further developed the linearly tapered MIM concept proposed by Conway and Yablonovitch[36] as well as by Pile and Gramotnev[22]. An optimized, linearly tapered MIM gap plasmon waveguide could theoretically reduce the excessive losses that occur during nanofocusing processes (Supplementary Section S4). In simulation, the coupling loss and maximum E^2 enhancement of the 3D NPC were predicted to be 2.5 dB and ~3.0×10^4, respectively, for the case where light was compressed from a 200×500 nm^2 area into a 2×5 nm^2 area. We realized the 3D NPC on a chip by using electron beam-induced deposition (EBID) and demonstrated highly localized light confinement using two-photon photoluminescence (TPPL) techniques. From the TPPL measurements, we experimentally estimated an intensity enhancement of 400 within a 14×80 nm^2 cross-sectional area and a coupling efficiency of -1.3 dB (or 74% transmittance).

Device concept

The 3D NPC structure is presented schematically in Fig. 1. The MIM–SPP configuration includes upper and lower gold layers with an intermediate SiO$_2$ layer. We utilize the fundamental anti-symmetric MIM–SPP mode, which forms an antisymmetric electric-field distribution about the intermediate dielectric layer and has no theoretical cutoff. This SPP mode is suitable for nanofocusing applications because its profile scales down with the size of the MIM cross-section, in contrast to a conventional dielectric fibre probe (Supplementary Sections S1,S2). The device has three functional sections: (i) a diffraction-limited spot-sized rectangular body that serves as an input coupler; (ii) a tapering section as a microscale to nanoscale optical-mode converter; and (iii) a sub-100 nm scale
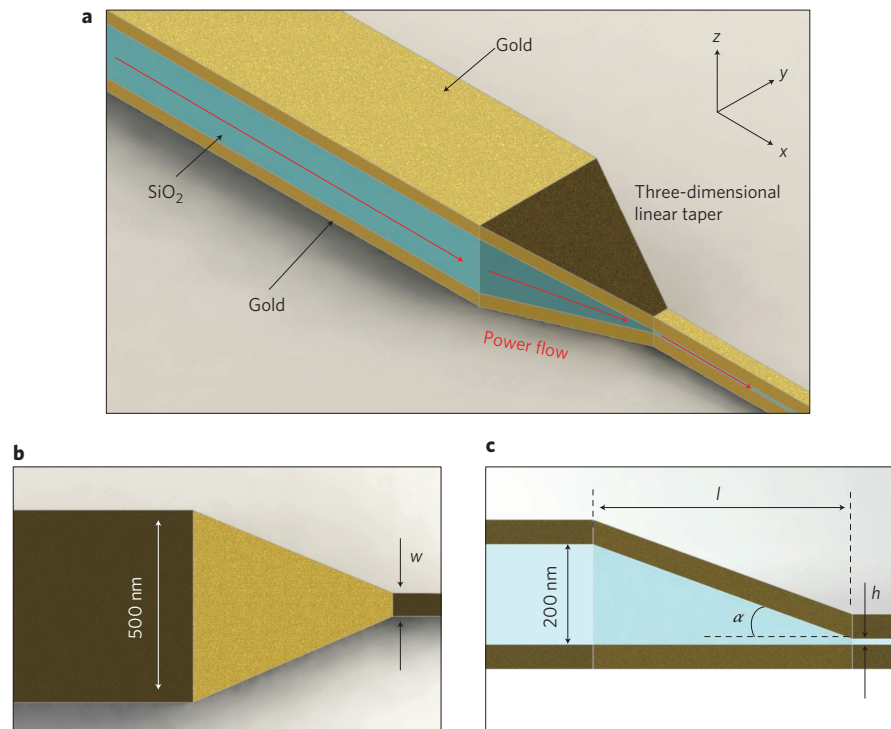
Figure 1 | The 3D NPC. a–c, Perspective (**a**), top (**b**) and side (**c**) views of a schematic illustration of a 3D NPC structure. The 3D NPC is essentially configured as a three-dimensional linearly tapered MIM-based SPP guide composed of upper and lower gold plates with an intermediate dielectric SiO_2 layer.

MIM–SPP focusing waveguide/tip. The optical light that couples into the MIM–SPP waveguide through the rectangular aperture at the open end of the input coupler is converted into SPPs with an efficiency exceeding 50% (refs 37,38). In the tapering section, where mode conversion occurs, the horizontal and vertical dimensions (perpendicular to the direction of SPP propagation) of the MIM–SPP waveguide decrease linearly to the sub-100 nm scale MIM–SPP waveguide/tip. The SPPs undergo a gradual conversion along this three-dimensional linear taper, and form a highly localized hot spot inside the sub-100 nm scale waveguide.

Design optimization
To maximize the field confinement in the sub-100 nm scale waveguide/tip, we first optimized the dimensions of the 3D NPC by running a series of finite-difference time-domain (FDTD) simulations. Because of the diffraction-limited spot size, we fixed the width and height of the SiO_2 layer of the rectangular body to 500 nm and 200 nm, respectively. To achieve an optimal design of the three-dimensional taper, we varied the vertical tapering angle α of the coupler while holding the ratio of its width and height constant, and also examined a range of SiO_2 heights h for the sub-100 nm scale waveguide (Fig. 1c and Methods).

The thicknesses of the upper and lower gold plates were set to 50 nm, approximately twice the skin depth ($\sim$25 nm) of the gold, at 360 THz (ref. 39). The propagation length of the fundamental antisymmetric MIM–SPP mode at 360 THz (wavelength = 833 nm) in the MIM body was calculated to be 7.2 μm (Supplementary Section S2); this sets the maximum length of the MIM plasmonic waveguide. At the sub-100 nm scale waveguide, where h is 14 nm, this length was reduced to 0.88 μm.

A device with a sub-100 nm scale waveguide of infinite length was examined to understand the influence of the tapering geometry on the coupling efficiency in the absence of cavity resonance. Figure 2a shows the coupling efficiency as a function of α for four sub-100 nm scale waveguides with h = 30, 15, 5 and 2 nm. The coupling efficiency was calculated by comparing the total power

(P_T/P_B) through the cross-section in the body (P_B) and in the sub-100 nm scale waveguide (P_T) immediately before and after the tapered section. For the P_B calculation, we modified the geometry of the body so that it was infinitely long without a tapered section. Regardless of the value of h, the highest coupling efficiency was calculated for $\alpha = 20°$ (83%, 77%, 66% and 56% for h = 30, 15, 5 and 2 nm, respectively). For large α, coupling was predominantly limited by scattering due to the severely mismatched wave vectors[21,22,36]. On the other hand, small α was less than optimal, and coupling was limited by the absorption losses in the longer tapered section[21,22,36]. The resistive and scattering losses in the tapered section were balanced for α between 15° and 30° in all four cases, as shown in Fig. 2a.

Figure 2b shows the calculated power flux profiles along the x–y and x–z planes for $\alpha = 20°$ and h = 15 nm. The power flux was enhanced by $\sim$20 dB at the junction between the tapered section and the sub-100 nm scale waveguide. Once the SPPs reached the sub-100 nm scale waveguide, the intensity decayed rapidly along the direction of propagation due to the extremely short propagation length—less than 0.94 μm at the tip with h = 15 nm and w = 38 nm (Supplementary Section S2).

Figure 2c shows the E^2 enhancement and coupling efficiency at the junction between the tapered section and the sub-100 nm scale waveguide as a function of h for $\alpha = 20°$. The relative E^2 enhancement was again calculated in the same manner as the coupling efficiency, and maximum E^2 values were compared. Figure 2c shows that the maximum intensity was strongly enhanced for values of h less than 10 nm, even though the coupling loss increased significantly. This behaviour results from the high spatial localization of the field and the increase in the effective refractive index, n_{eff} (increase of which results in slow propagation; see Supplementary Section S2). For a sub-100 nm scale waveguide, 2.0 nm tall and 5.0 nm wide, the maximum intensity enhancement was predicted to be as high as $\sim$3.0 × 10^4 at $\alpha = 20°$ with a coupling loss of 2.5 dB.

Figure 2d,e presents the cross-sectional distributions of E^2 (a vector plot of the electric field is overlaid in Fig. 2d) along
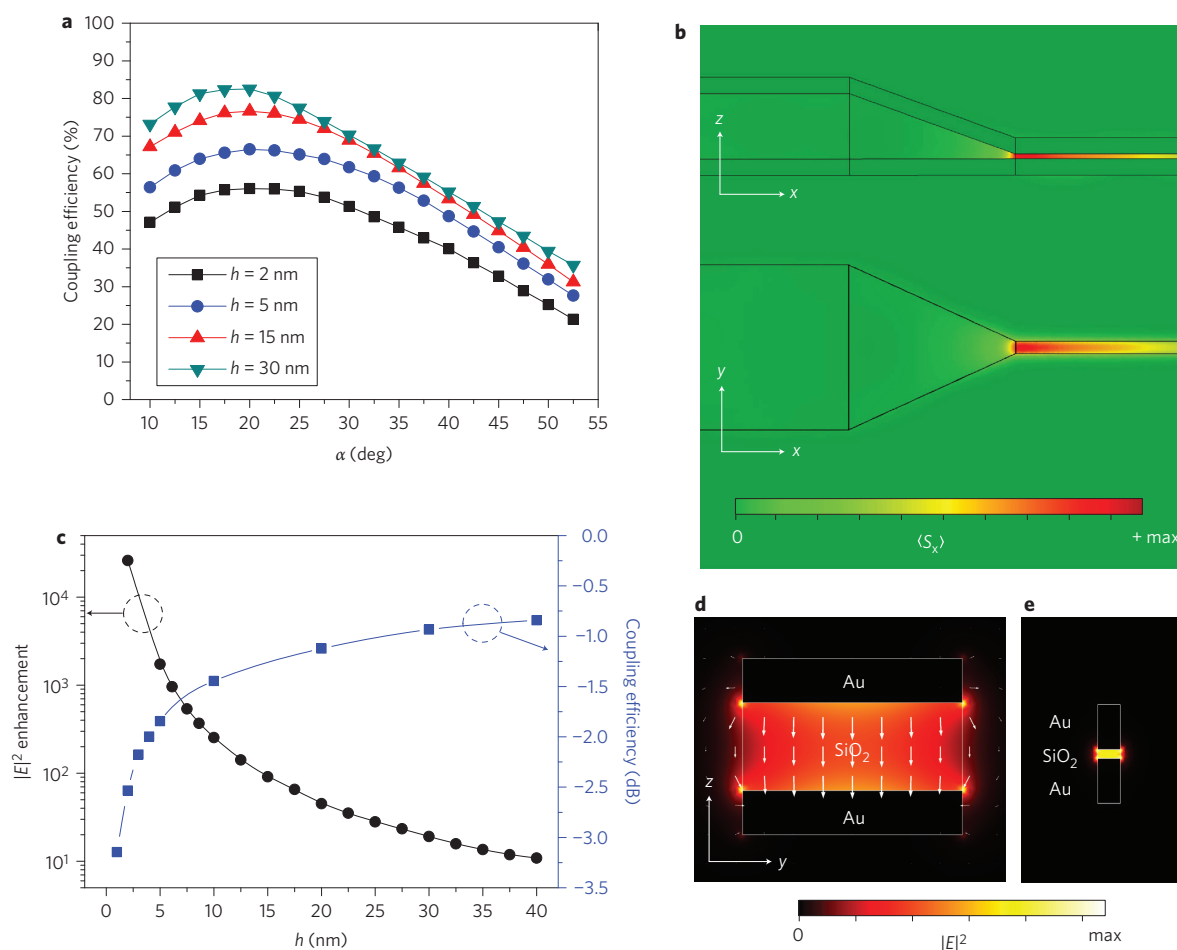
Figure 2 | Simulation of a 3D NPC with an infinite tip. a, Calculated coupling efficiency as a function of vertical taper angle α for SiO_2 tip thickness values h of 30, 15, 5 and 2 nm in the 3D NPC with an infinite tip. **b**, Time-averaged power flux profiles along the x–z and x–y planes for $\alpha = 20°$ and $h = 15$ nm under conditions in which power excitation occurred on the left end of the guide. **c**, E^2 enhancement and coupling efficiency on a dB scale as a function of h for $\alpha = 20°$. **d,e**, Cross-sectional side views of the E^2 profiles along the y–z plane for $h = 200$ and 10 nm, respectively. The E-field vector plot is overlaid in Fig. 2d.

the body and the sub-100 nm scale waveguide of the 3D NPC, for $h = 200$ nm and 10 nm, respectively. The images clearly show that the energy of the fundamental antisymmetric MIM–SPP mode was extremely concentrated between the two metal plates inside the SiO_2 layer. The corresponding photon density at the hotspot increased, particularly as the dimensions of the SiO_2 spacer were decreased below 100 nm.

Sub-100 nm scale waveguides of finite length (or focusing tips), as shown in Fig. 3a, generate cavity resonances due to constructive and destructive interference between the forward- and backward-propagating SPPs, which further enhances the electric field. Our theoretical study shows that the resonance spatial properties depend strongly on L_{tip} (defined in Fig. 3a) as well as h.

Figure 3b shows the E^2 enhancement as a function of L_{tip} for $h = 5$, 10 and 15 nm. The cavity resonant characteristics were investigated by varying L_{tip} from 20 nm to 380 nm while holding α fixed at 20° and the length of the 3D NPC body at 1.5 µm. For each h, several resonance peaks were observed over the range of L_{tip} values examined (Fig. 3b). The overall trend of the results showed that the intensity enhancement decreased as L_{tip} increased, and, consequently, photon density decreased. At this deep sub-wavelength scale, the slope of the peak intensity enhancement in Fig. 3b becomes large because the cross-sectional modal area (or h) is extremely small (Supplementary Section S5).

The height of the dielectric layer h strongly influenced the calculated resonance properties. The value of n_{eff} at the tip varied

as a function of h, and this resulted in a unique spatial period (ΔL_{tip}) between the resonant peaks for each h. From Fig. 3b, ΔL_{tip} values between the resonant peaks for $h = 5$, 10 and 15 nm were 60, 100 and 120 nm, respectively. From these values, we estimated n_{eff} in the tip cavity region using the Fabry–Perot resonance relation $n_{eff} = \lambda/(2\Delta L_{tip})$. The values of n_{eff} were ~6.9, 4.1 and 3.5 for $h = 5$, 10 and 15 nm, respectively. These values agreed well with their counterparts in the case of the infinite tip, that is, $n_{eff} = 6.7$, 4.0 and 3.5 for $h = 5$, 10 and 15 nm, respectively (Supplementary Section S2).

Figure 3c shows the intensity distributions for $h = 15$ nm and $L_{tip} = 130$, 250 and 380 nm. The standing wave patterns, which were not observed in the 3D NPC with an infinite tip, were clearly visible in the finite tip. The maximum intensity enhancement at each hot spot was three to five times larger than the corresponding value observed in the infinite tip. For $h = 15$ nm and $\alpha = 20°$, the coupling efficiency was calculated to be 77%, and the intensity enhancement exceeded 400 ($L_{tip} < 150$ nm).

Fabrication and characterization

Based on the optimal design described in the previous section, we fabricated a 3D NPC structure on a fused silica substrate using EBID, focused ion-beam milling, and electron-beam evaporation (see Methods and Supplementary Section S6). The final dimensions were chosen by considering the potential challenges involved in nanofabrication and characterization. Figure 4a presents a scanning
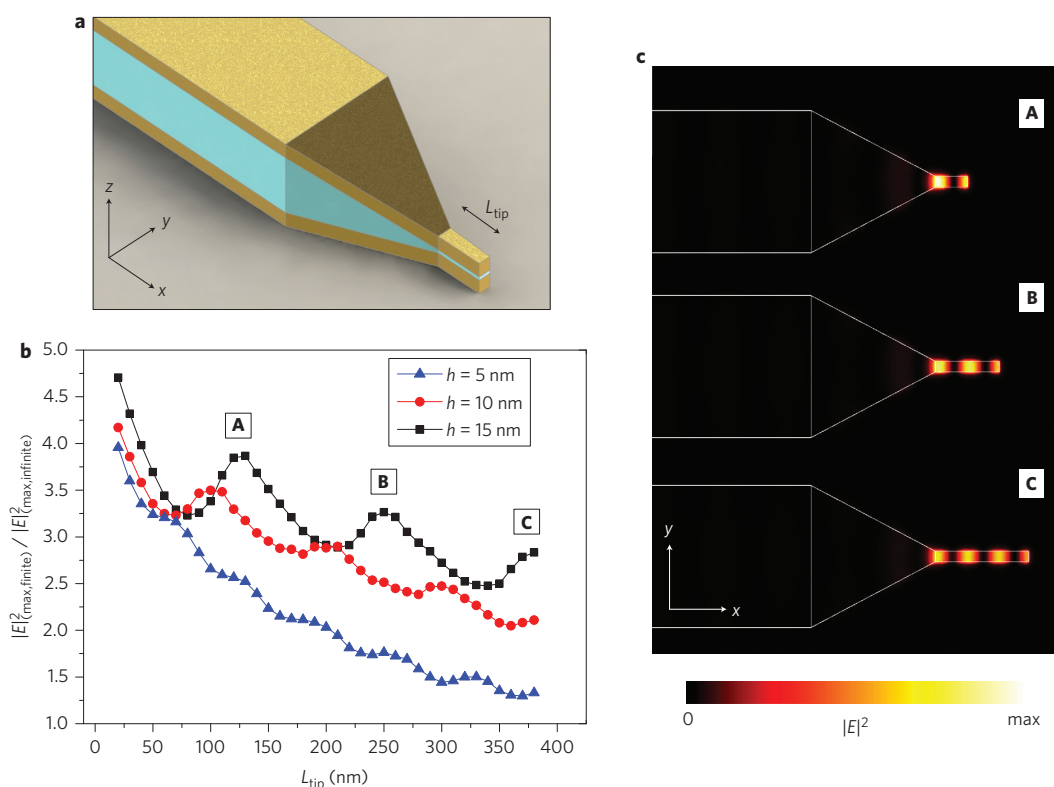
Figure 3 | Simulation of a 3D NPC with a finite tip. a, Schematic illustration of a 3D NPC structure with finite tip. **b**, Theoretical E^2 enhancement ratio in the 3D NPC with a finite or infinite tip, for $h = 5$, 10 and 15 nm. **c**, Cross-sectional top views of the E^2 profiles along the x–y plane under three conditions of resonance, for $h = 15$ nm, as shown in **b**.
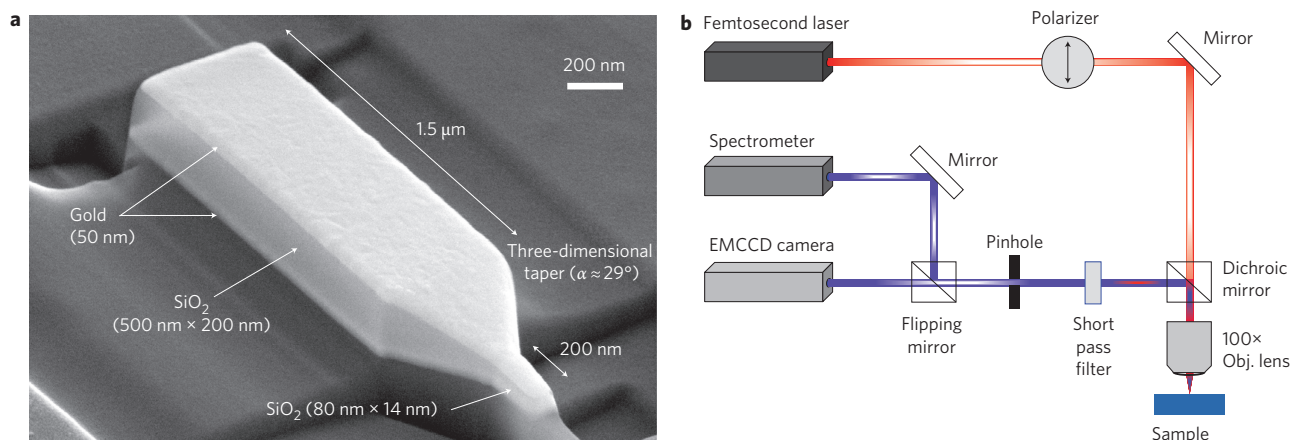

Figure 4 | Fabricated 3D NPC and optical characterization setup. a, SEM image of a fabricated 3D NPC structure. **b**, TPPL measurement setup. The sample was excited using a focused laser beam with a selected polarization. The TPPL emission was directly imaged using an EMCCD camera and spectrally analysed using a spectrometer.

electron microscopy (SEM) image (collected at 45°) of the fabricated NPC structure. The minimum width and height of the intermediate dielectric (SiO₂) layer were measured as $w = 80$ nm and $h = 14$ nm. The tapering angle was 29°, which lay within the high coupling efficiency (>70%) range of 15–30° determined by our theoretical study (Fig. 2a).

TPPL measurement techniques[40,41] in a confocal setup (Fig. 4b) were used to demonstrate the intensity enhancement at the tip of the fabricated NPC (see Methods). For excitation, we delivered 830 nm, 120 fs Ti:sapphire laser pulses (54 MHz repetition rate) at an average power of 210 μW to the back aperture of the NPC. Figure 5a–d presents electron-multiplying charge-coupled

device (EMCCD) images of the TPPL emission profiles for four excitation scenarios. The excitation light polarized along the x-axis (x-polarization) clearly indicates that the maximum photon compression into the tip is observed only when the light is properly coupled into the back aperture of the rectangular body, thereby generating SPPs that travel through the three-dimensional tapered coupler (Fig. 5d) and are concentrated in the tip. In the other three scenarios, the field enhancement at the tip was minimal (Fig. 5a–c). Note that the x-polarized excitation light generated the desired transverse magnetic (TM) antisymmetric plasmonic mode and expected enhancement in the MIM structure (Supplementary Section S7). Figure 5e shows
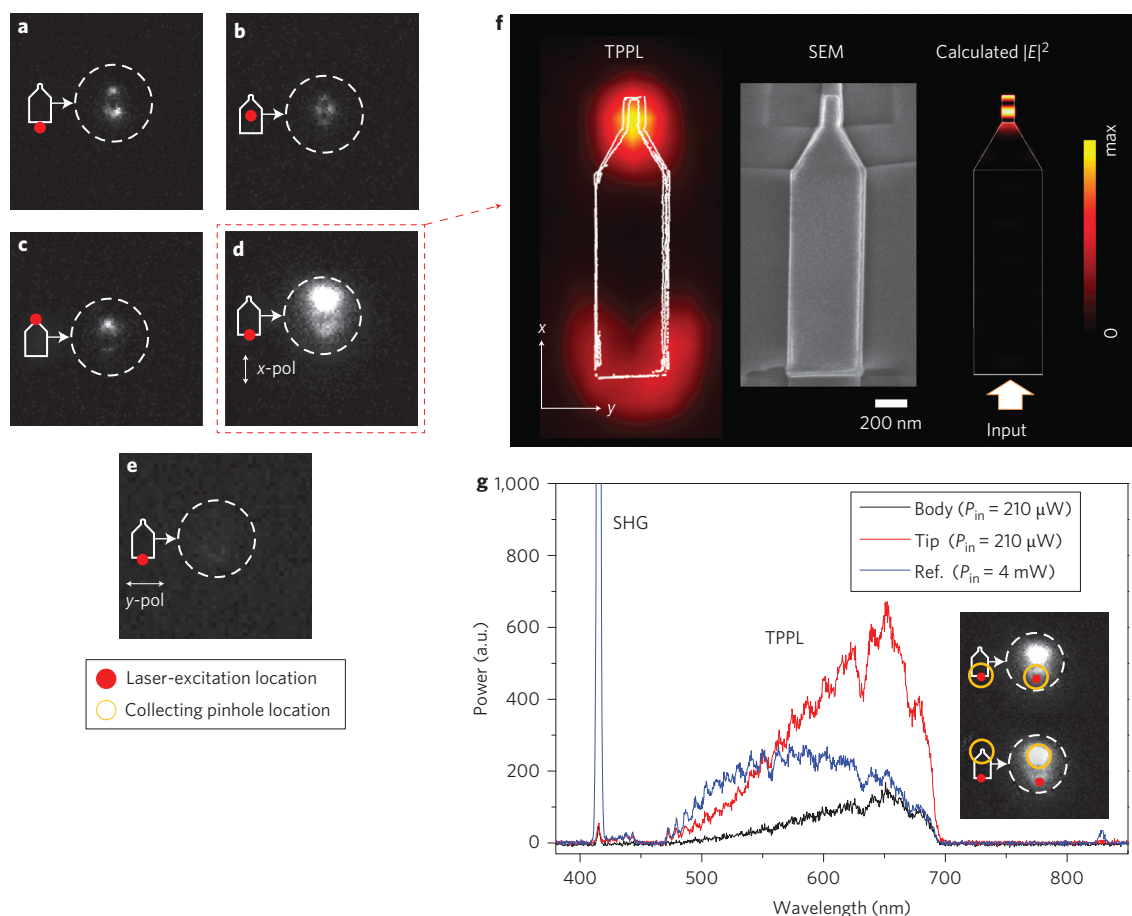
Figure 5 | Optical characterization of the 3D NPC. a–d, EMCCD images of TPPL emission for four different excitation scenarios. Red dots indicate positions of laser excitation. The laser beams were *x*-polarized, and excitation power was fixed at 210 μW. **e**, TPPL image of the device for *y*-polarized excitation at the same excitation position and power as in **d**. **f**, Unsaturated intensity map of **d** superimposed onto the contours of an SEM image of the device, and top-view SEM image of the characterized sample and simulated $|E|^2$ intensity profile for a sample with the dimensions of the fabricated sample. **g**, Emission spectra of the 3D NPC body (black curve) and tip (red curve), and a smooth gold film (blue curve) for reference. Inset: we intentionally aligned and misaligned the pinhole (yellow circle) relative to the excitation beam position (red dot).

a TPPL image of the device during excitation with *y*-polarized light at the same excitation location and power as in the *x*-polarization case shown in Fig. 5d. This clearly proves that the coupled mode responsible for strong enhancement in the 3D NPC structure shown in Fig. 5d is the antisymmetric TM plasmonic mode (Supplementary Section S7).

The location of the brightest TPPL emission spot in Fig. 5d was more accurately estimated by superimposing the unsaturated intensity map captured by the EMCCD camera onto the contour of the SEM image of the device, as shown in Fig. 5f. This clearly shows a maximum TPPL emission in the tip of the 3D NPC structure. Weaker TPPL emission, observed at the location of the laser excitation (near the back aperture of the rectangular body), resulted from local confinement at the edges. The rightmost figure in Fig. 5f shows the simulated E^2 profile for dimensions the same as those of the fabricated sample. The location of the maximum E^2 value in the simulation result agrees well with the location observed in the measured intensity map on the left; however, the nanoscale interference profile in the tip, as revealed by the simulation, could not be observed in the experimentally obtained TPPL intensity map due to the diffraction limit of the imaging system.

By intentionally aligning and misaligning the 150-μm-diameter pinhole (in the confocal setup shown in Fig. 4b) and the excitation beam, we could selectively excite the structure at the base and collect the emission spectrum from the tip. Alternatively, we could excite and collect at the same location, as shown in the insets of Fig. 5g. Although only a small fraction of the 210 μW excitation power coupled from the air into the body of the 3D NPC and reached the tip, the average TPPL intensity from the tip was six times stronger than the average intensity generated from the base of the rectangular body.

Discussion

The extreme enhancement in the 3D NPC originates from its unique, cleverly implemented tapering geometry. To demonstrate and quantify the intensity enhancement originating from geometric effects, the strength of the TPPL emission from a nearby flat gold film (from which no significant field enhancements would be expected because of the surface geometry) was measured as a reference and then compared with that of the NPC.

The first step in the analysis was to calculate the total intensity enhancement factor, that is, the ratio between the incident excitation intensity and the intensity generated in the tip ($\alpha_{\mathrm{tip/inc}}$):

$$\alpha_{\mathrm{tip/inc}} = \frac{E_{\mathrm{tip}}^2}{E_{\mathrm{inc}}^2} = \frac{E_{\mathrm{tip}}^2}{E_{\mathrm{body}}^2} \cdot \frac{E_{\mathrm{body}}^2}{E_{\mathrm{inc}}^2} = \alpha_{\mathrm{tip/body}} \cdot \alpha_{\mathrm{body/inc}} \qquad (1)$$

The enhancement factor from our simulations compared the intensity before and after the tapered section, and is given by $\alpha_{\mathrm{tip/body}}$.

Hence, for an apple-to-apple comparison, we rearrange the variables in equation (1) to obtain

$$\alpha_{\text{tip/body}} = \frac{E_{\text{tip}}^2}{E_{\text{body}}^2} = \frac{E_{\text{tip/inc}}^2}{E_{\text{body/inc}}^2} = \frac{E_{\text{tip}}^2}{E_{\text{inc}}^2} \cdot \frac{E_{\text{inc}}^2}{E_{\text{body}}^2} = \alpha_{\text{tip/inc}} \cdot \alpha_{\text{inc/body}} \quad (2)$$

where $\alpha_{\text{tip/inc}}$ is given by[40]

$$\alpha_{\text{tip/inc}} = \left(\frac{A_{\text{ref}}}{A_{\text{tip}}} \times \frac{\langle \text{TPPL}_{\text{tip}} \rangle}{\langle \text{TPPL}_{\text{ref}} \rangle} \right)^{1/2} \cdot \frac{\langle P_{\text{ref}} \rangle}{\langle P_{\text{inc}} \rangle} \approx 125 \quad (3)$$

A_{ref} and A_{tip}, which are the surface areas of the reference flat gold film and the tip from which the TPPLs originated, were 126,000 nm^2 and 5,500 nm^2, respectively (Supplementary Section S8). The ratio between the time-averaged TPPL intensities from the tip and the reference film ($\langle \text{TPPL}_{\text{tip}} \rangle / \langle \text{TPPL}_{\text{ref}} \rangle$) was estimated to be 1.8 based on the TPPL spectra shown in Fig. 5g. $\langle P_{\text{ref}} \rangle$ and $\langle P_{\text{inc}} \rangle$, the average excitation power incident on the 3D NPC and on the flat gold surface, were 210 µW and 4 mW, respectively.

An estimate for $\alpha_{\text{inc/body}}$ (the E^2 enhancement factor of the field intensity inside the body of the MIM plasmonic waveguide compared to that of the incident excitation beam) was obtained by considering the coupling efficiency between the vertically incident Gaussian beam and the MIM–SPP waveguide mode in addition to the SPP modal area and n_{eff} inside the MIM plasmonic waveguide (Supplementary Section S8). $\alpha_{\text{inc/body}}$ is given by

$$\alpha_{\text{inc/body}} = (1/\eta) \cdot \left(\frac{A_{\text{m}}^{\text{body}}}{A_{\text{ref}}} \right) \cdot \left(\frac{\varepsilon_{\text{SiO}_2}}{n_{\text{eff}}^{\text{body}}} \right) \approx 3.2 \quad (4)$$

The coupling efficiency η, the modal area $A_{\text{m}}^{\text{body}}$ and the effective refractive index $n_{\text{eff}}^{\text{body}}$ of the body MIM–SPP mode were estimated to be 0.15, 44,000 nm^2 and 1.63, respectively, using FDTD and finite-element method simulations (Supplementary Sections S2,S7). Equation (2) was then evaluated over the given values to obtain $\alpha_{\text{tip/body}} \approx 400$, which was found to be in excellent agreement with the simulation result of 410. The coupling efficiency (or transmittance) through the tapered section without the cavity effect was estimated to be > -1.3 dB (74%; see Supplementary Section S9). In Fig. 5g, second harmonic generation (SHG) was weakly detected from the tip of the 3D NPC structure, in contrast to the highly enhanced TPPL signal. This results from the fact that SHG emission preserves the direction (parallel to the detection plane in our case) and coherence of the excitation plasmonic waves, whereas the TPPL emission is highly isotropic (Supplementary Section S10). In the case of a flat gold surface, a strong SHG signal could be detected due to back scattering/reflection from the substrate. This direction-dependent result provides additional strong evidence that the energy localization at the tip resulted from the SPP-mode energy guided through the linearly tapered 3D MIM plasmonic waveguide.

In summary, we have demonstrated a highly efficient on-chip three-dimensional nanofocusing structure. Optimal tapering angles and lengths were found in simulation to minimize the intrinsic scattering and resistive absorption losses observed during nanofocusing. The optimized design was realized on chip using an EBID technique, and the implemented 3D NPC showed highly localized light confinement. We strongly believe that this highly efficient three-dimensional nanoplasmonic photon compressor will be useful in a variety of on-chip nanoscale photonic/plasmonic applications.

Methods

Device fabrication. The three-dimensional taper of the NPC was fabricated in two steps: (i) the vertical taper was created using electron-beam-induced deposition and (ii) the lateral taper was defined using focused ion-beam milling. Using electron beam-induced deposition of SiO$_2$, a vertical SiO$_2$ taper structure (tapering angle ~29°) was created on top of an evaporated gold film on a fused silica coverslip. The deposition parameters, including the scan patterns and durations, as well as the intensity of the electron beam, were adjusted to generate the designed tapering geometry of the SiO$_2$. Following electron beam-induced SiO$_2$ deposition, another 50-nm-thick gold layer was evaporated onto the device (Supplementary Section S6). Finally, a lateral taper and the device geometry were defined using focused ion-beam milling (see Supplementary Movie). The purity of the gold used for evaporation exceeded 99.99%. The evaporation rate was 0.5 Å s^{-1}, and the root-mean-square and peak-to-peak roughness values for the deposited gold film were 1.44 nm and 8 nm, respectively, characterized using atomic force microscopy. The final fabricated photon confinement area measured 14×80 nm^2. The complete set of dimensions for the fabricated NPCs is given in Fig. 4a.

Optical measurements. The fabricated NPC was tested for photonic confinement by imaging and measuring the TPPL emission profile in a confocal setup. Using a 100×0.90 NA objective lens, a 120 fs Ti:sapphire laser at 830 nm (full-width at half-maximum (FWHM) ~400 nm) was focused at the base of the NPC. Direct imaging of the TPPL emission was performed using a thermoelectrically cooled EMCCD camera. Spectral measurements were collected using a liquid nitrogen-cooled spectrometer. The pinhole in the setup had a diameter of 1.5 mm, which appeared as a 150 µm aperture in the field of view. At normal incidence, the photons scattered along the edges of the base coupled into the NPC, and generated SPPs along the interfaces of the gold and SiO$_2$, travelling towards the tip of the NPC. By adjusting (or intentionally misaligning) the placement of the excitation beam with respect to the pinhole position, three measurements at different excitation and collection conditions were collected on the NPC: (i) excitation and collection at the base; (ii) excitation and collection at the tip; and (iii) excitation at the base and collection at the tip. Finally, as a reference value, we measured the intensity enhancement from the unprocessed area of the evaporated gold film. Note that SHG was considerably suppressed in the base-excitation/tip-collection measurement.

Simulation details. The simulations shown in Figs 2, 3 and 5f were conducted numerically using a commercially available FDTD software package (CST microwave studio, http://www.cst.com). The propagation lengths and effective refractive index described in the text and the Supplementary Information were calculated using the eigenmode solver in the RF module of the COMSOL software (http://www.comsol.com). For all calculations, we assumed a free-space frequency of 360 THz, which corresponds to a wavelength of 833 nm. The complex permittivity of gold and the refractive index of SiO$_2$ were $\varepsilon_{\text{Au}} = -26.4 + 1.65i$ and $n = 1.45$, respectively. We fixed the width (w_{body}) and height (h_{body}) of the SiO$_2$ layer of the rectangular body to 500 nm and 200 nm, respectively. We then varied (i) the vertical tapering angle α of the coupler and (ii) the SiO$_2$ height h of the sub-100 nm scale waveguide to maximize the energy compression at the sub-100 nm scale waveguide (Fig. 1c). The ratio between the width and height of the cross-sectional dielectric area was held constant along the entire waveguide ($w/h = w_{\text{body}}/h_{\text{body}}$) throughout the tapered section. Accordingly, the numerical values of the parameters α and h yielded all other dimensions, including the length [$l = (200 \text{ nm} - h) \cdot \tan^{-1}(\alpha)$] and width [$w = (500 \text{ nm}/200 \text{ nm}) \times h$] of the tapered section. The selection of α and h determined the length of the tapering section and the width of the sub-100 nm scale waveguide, as shown in Fig. 1.

References

1. Hill, M. T. *et al.* Lasing in metallic-coated nanocavities. *Nature Photon.* **1,** 589–594 (2007).
2. Noginov, M. A. *et al.* Demonstration of a spaser-based nanolaser. *Nature* **460,** 1110–1112 (2009).
3. Oulton, R. F. *et al.* Plasmon lasers at deep subwavelength scale. *Nature* **461,** 629–632 (2009).
4. Khajavikhan, M. *et al.* Thresholdless nanoscale coaxial lasers. *Nature* **482,** 204–207 (2012).
5. Tang, L. *et al.* Nanometre-scale germanium photodetector enhanced by a near-infrared dipole antenna. *Nature Photon.* **2,** 226–229 (2008).
6. Neutens, P., Van Dorpe, P., De Vlaminck, I., Lagae, L. & Borghs, G. Electrical detection of confined gap plasmons in metal–insulator–metal waveguides. *Nature Photon.* **3,** 283–286 (2009).
7. Nikolajsen, T., Leosson, K. & Bozhevolnyi, S. I. Surface plasmon polariton modulators and switches operating at telecom wavelengths. *Appl. Phys. Lett.* **85,** 5833–5835 (2004).
8. Cai, W., White, J. S. & Brongersma, M. L. Compact, high-speed and power-efficient electrooptic plasmonic modulators. *Nano. Lett.* **9,** 4403–4411 (2009).
9. Anger, P., Bharadwaj, P. & Novotny, L. Enhancement and quenching of single-molecule fluorescence. *Phys. Rev. Lett.* **96,** 113002 (2006).
10. Taminiau, T. H., Stefani, F. D., Segerink, F. B. & Van Hulst, N. F. Optical antennas direct single-molecule emission. *Nature Photon.* **2,** 234–237 (2008).

11. Okamoto, K. *et al.* Surface-plasmon-enhanced light emitters based on InGaN quantum wells. *Nature Mater.* **3,** 601–605 (2004).
12. Yuan, Z. et al. Electrically driven single-photon source. *Science* **295,** 102–105 (2002).
13. Akimov, A. V. *et al.* Generation of single optical plasmons in metallic nanowires coupled to quantum dot. *Nature* **450,** 402–406 (2007).
14. Maier, S. A. A. *Plasmonics: Fundamentals and Applications* (Springer, 2007).
15. Gramotnev, D. K. & Bozhevolnyi, S. I. Plasmonics beyond the diffraction limit. *Nature Photon.* **4,** 83–91 (2010).
16. Volkov, V. S. *et al.* Nanofocusing with channel plasmon polaritons. *Nano Lett.* **9,** 1278–1282 (2009).
17. Vernon, K. C., Gramotnev, D. K. & Pile, D. F. P. Adiabatic nanofocusing of plasmons by a sharp metal wedge on a dielectric substrate. *J. Appl. Phys.* **101,** 104312 (2007).
18. Verhagen, E., Kuipers, L. & Polman, A. Plasmonic nanofocusing in a dielectric wedge. *Nano Lett.* **10,** 3665–3669 (2010).
19. Schnell, M. *et al.* Nanofocusing of mid-infrared energy with tapered transmission lines. *Nature Photon.* **5,** 283–287 (2011).
20. Davoyan, A. R., Shadrivov, I. V., Zharov, A. A., Gramotnev, D. K. & Kivshar, Y. S. Nonlinear nanofocusing in tapered plasmonic waveguides. *Phys. Rev. Lett.* **105,** 116804 (2010).
21. Gramotnev, D. K., Pile, D. F. P., Vogel, M. W. & Zhang, X. Local electric field enhancement during nanofocusing of plasmons by a tapered gap. *Phys. Rev. B* **75,** 035431 (2007).
22. Pile, D. F. P. & Gramotnev, D. K. Adiabatic and nonadiabatic nanofocusing of plasmons by tapered gap plasmon waveguides. *Appl. Phys. Lett.* **89,** 04111 (2006).
23. Choi, H., Pile, D. F. P., Nam, S., Bartal, G. & Zhang, X. Compressing surface plasmons for nano-scale optical focusing. *Opt. Express* **17,** 7519–7524 (2009).
24. Stockman, M. I. Nanofocusing of optical energy in tapered plasmonic waveguides. *Phys. Rev. Lett.* **93,** 137404 (2004).
25. Bouhelier, A., Beversluis, M., Hartschuh, A. & Novotny, L. Near-field second-harmonic generation induced by local field enhancement. *Phys. Rev. Lett.* **90,** 013903 (2003).
26. Ropers, C. *et al.* Grating-coupling of surface plasmons onto metallic tips: a nanoconfined light source. *Nano Lett.* **7,** 2784–2788 (2007).
27. Verhagen, E., Kuipers, L. & Polman, A. Enhanced nonlinear optical effects with a tapered plasmonic waveguide. *Nano Lett.* **7,** 334–337 (2007).
28. Lindquist, N. C., Nagpal, P., Lesuffleur, A., Norris, D. J. & Oh, S.-H. Three-dimensional plasmonic nanofocusing. *Nano Lett.* **10,** 1369–1373 (2010).
29. Verhagen, E., Polman, A. & Kuipers, L. (Kobus). Nanofocusing in laterally tapered plasmonic waveguides. *Opt. Express* **16,** 45–57 (2008).
30. Gramotnev, D. K., Vogel, M. W. & Stockman, M. I. Optimized nonadiabatic nanofocusing of plasmons by tapered metal rods. *J. Appl. Phys.* **104,** 034311 (2008).
31. Hecht, B. *et al.* Scanning near-field optical microscopy with aperture probes: fundamentals and applications. *J. Chem. Phys.* **112,** 7761–7774 (2000).
32. Novotny, L. & Hafner, C. Light propagation in a cylindrical waveguide with a complex, metallic dielectric function. *Phys. Rev. E* **50,** 4094–4106 (1994).
33. Issa, N. A. & Guckenberger, R. Optical nanofocusing on tapered metallic waveguides. *Plasmonics* **2,** 31–37 (2007).
34. Novotny, L. & Hecht, B. *Principles of Nano-Optics* (Cambridge Univ. Press, 2006).
35. Vedantam, S. *et al.* A plasmonic dimple lens for nanoscale focusing of light. *Nano Lett.* **9,** 3447–3452 (2009).
36. Conway, J. *Efficient Optical Coupling to the Nanoscale* PhD thesis, Univ. California, (2006).
37. Worthing, P. T. & Barnes, W. L. Efficient coupling of surface plasmon polaritons to radiation using a bi-grating. *Appl. Phys. Lett.* **79,** 3035–3037 (2001).
38. Feng, N.-N. & Negro, L. D. Plasmon mode transformation in modulated-index metal–dielectric slot waveguides. *Opt. Lett.* **32,** 3086–3088 (2007).
39. Reuter, G. E. H. & Sondheimer, E. H. The theory of the anomalous skin effect in metals. *Proc. R. Soc. Lond. A* **195,** 336–364 (1948).
40. Schuck, P. J., Fromm, D. P., Sundaramurthy, A., Kino, G. S. & Moerner, W. E. Improving the mismatch between light and nanoscale objects with gold bowtie nanoantennas. *Phys. Rev. Lett.* **94,** 017402 (2005).
41. Beversluis, M. R., Bouhelier, A. & Novotny, L. Continuum generation from single gold nanostructures through near-field mediated intraband transitions. *Phys. Rev. B* **68,** 115433 (2003).

Author contributions

H.C. fabricated the 3D NPC devices, performed the experiments, and carried out TPPL measurements. S.C. and P.J.S. assisted with device fabrication and measurement collection. M.K. conducted the simulations and, with H.C., analysed the experimental data. M.S. and T.J.S. assisted with the simulations. J.B., M.W. and E.Y. provided in-depth discussion of the project. H.C. and M.K. wrote the manuscript.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permission information is available online at http://www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.C. and E.Y.

Competing financial interests

The authors declare no competing financial interests.