Optimization of coupling from a sub-wavelength nanoaperture to the fundamental Gaussian mode

Muthiah Annamalai and Michael Vasilyev*

Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX 76019, USA

(Received 29 March 2010; final version received 25 August 2010)

We minimize scattering losses and maximize coupling to a given free-space mode from a sub-wavelength metal aperture surrounded by periodic and aperiodic corrugations via optimum choice of the parameters of the structure. We show that introduction of aperiodicity into the corrugations enables coupling to a Gaussian mode \( T > 64\% \), which is twice as high as that obtainable from the best periodic-corrugation design. These results indicate the possibility of using such metal nanostructures for quantum- and nonlinear-optical applications.

Keywords: subwavelength structures; surface waves; optical nanocavities

1. Introduction

The sources of single photons on demand are important for quantum information processing, such as quantum key distribution, quantum communication via teleportation, and quantum computing based on linear optics. Making such sources involves building structures in which one unit of active material (e.g. atom, quantum dot, etc.) interacts only with a small set of optical modes, which requires the use of small-volume microcavities. In these structures, spontaneous emission can be enhanced and a considerable portion of it can be channeled into one optical mode (Purcell effect), if the ratio \( Q/(V/\lambda^3) \) is high, where \( Q \) is the cavity Q-factor, \( V \) is the mode volume, and \( \lambda \) is the emission wavelength in the medium. In addition, high collection efficiency of the emitted photons is required.

One promising approach to building single-photon sources, as well as to obtaining enhanced nonlinear-optical effects is to use metal nanocavities. Indeed, in contrast to semiconductor or dielectric resonators, the metal cavities can easily confine light to extremely small volumes \( (V \ll \lambda^3) \) with finesse on the order of 10\(^2\). This can be done, for example, by using an open-ended metal box resonator with dimensions \( \approx a \times \lambda/2^{3/2} \times \lambda/2^{1/2} \), where \( a \ll \lambda \), because its fundamental TE\(_{101}\) mode does not have a cut-off [1]. Placement of a fluorescent source at electric-field maximum in such a metal nanocavity would result in dramatic enhancement of the light–matter interaction due to very tight light confinement. The critical issue for such sub-wavelength structures, however, is the efficiency of coupling the light in and out of them into a desired free-space mode without much loss due to scattering to other modes. Surrounding the open end of the nanocavity (sub-wavelength nanoaperture) with periodic corrugations can concentrate the emission pattern into a narrow range of angles, as was demonstrated experimentally in [2].

In this paper, we investigate quantitatively how much light can be coupled from the nanoaperture to a well-defined macroscopic mode (e.g. a fundamental Gaussian beam) and how much is lost due to scattering into higher-order modes. The possibility of good coupling would make the metal nanocavity structures compatible with low-loss fiber delivery (the fiber mode is close to Gaussian). In [3,4], we extended the first-principles approach of [5] and [6] to calculate 30% coupling efficiency and 30% loss to higher-order modes (with the remaining 40% reflected back into the nanocavity) for the slit and corrugation parameters taken from [5] and corresponding to maximum total coupling of light from the sub-wavelength slit into the free space. This set of parameters optimized for total coupling may not, however, be optimum for Gaussian-mode coupling (i.e. scattering to higher-order modes may be too high). This calls for a separate calculation, where the corrugation period \( d \) and depth \( h \), number of corrugations on each side of the slit \( N \), and position of Gaussian beam waist \( z_0 \), are optimized for various wavelengths of light \( \lambda \), with the explicit purpose of maximizing the Gaussian-mode coupling and minimizing the loss due to scattering into other modes.

*Corresponding author. Email: vasilyev@uta.edu
Preliminary results of this optimization work were shown at a conference talk [4], and the detailed report is presented in this paper. In addition, we describe an extension of our model to aperiodic corrugations, which results in improved performance and also allows one to maximize the coupling at multiple wavelengths (e.g., useful for enhancing nonlinear-optical interactions).

2. Model of nanoaperture with periodic corrugations

We consider the 2D arrangement shown in Figure 1 (a), where a TEM excitation from a sub-wavelength slit made in a metal film, couples to the free space. The slit is surrounded by a periodic pattern of corrugations (grooves) symmetrically on both sides. References [5] and [6] investigated the interaction of such a structure with plane waves and, consequently, were only concerned with relative enhancement of light transmission through it compared to the non-corrugated case. We, on the other hand, study the interaction of this structure with the fundamental Gaussian mode and are interested in the actual value of the coupling coefficient and the percentage of loss due to scattering into other modes.

We start with the approach of [3] where the field emitted by the sub-wavelength slit of width \( a = 40 \text{ nm} \) is calculated similar to [5] and then projected onto a cylindrical Gaussian beam. In accordance with [5] and [6], we assume perfect metal boundary conditions (lossless metal) and the groove/slit width \( a \ll \lambda \). Then for p-polarized light, the grooves act as metal slab waveguides supporting only the fundamental TEM mode. We consider the light emerging from the slit through the corrugated face of the metal film (serving as the output coupler from the nanocavity), ignoring the opposite side of the film (which can be incorporated later on by simple Fabri-Perot formula). A self-consistent system of equations [5] calculates the field amplitudes \( E_\alpha \) at each corrugation with index \( \alpha = -N, \ldots, +N \) \((\alpha = 0 \text{ is the slit})\). The total transmittance from the slit to free space is \( T_{\text{total}} = 1 - |1 - E_0/A_0|^2 \), where \( A_0 \) is the amplitude of field traveling inside the slit.

Thus, we need to solve the system of equations from [5],

\[
(G_{\alpha\alpha} - \epsilon_\alpha) E_\alpha + \sum_{\beta \neq \alpha} G_{\alpha\beta} E_\beta = 2iA_0 \delta_{\alpha0},
\]

where \( \alpha \) and \( \beta \) indices are running from \(-N\) to \(+N\), with \( \epsilon_\alpha = 1/\tan k h \) for \( \alpha \neq 0 \) and \( \epsilon_\alpha = -i \) for \( \alpha = 0 \), \( k = 2\pi/\lambda \).

\[
G_{\alpha\beta} = \frac{i k}{2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi_\alpha^*(x) H_{0}^{(1)}(k|x-x'|) \phi_\beta(x') dx dx',
\]

(2)

mode functions \( \phi_\alpha(x) = 1/\sqrt{a} \) within the \( \alpha \)th corrugation and \( \phi_\alpha(x) = 0 \) outside of it, and \( H_{0}^{(1)} \) being the Hankel function.

The lowest-order (fundamental) 1D Gaussian mode TEM0 has the form

\[
g(x,z) = \text{Gaussian}(x,z) = \frac{\exp\left(-\frac{1}{2}\frac{kx^2}{a_0^2 - (z-z_0)^2}\right)}{\sqrt{\frac{\pi a_0^2}{2}} \left[1 + (z-z_0)^2/(ka_0^2)^2\right]},
\]

(3)

where \( a_0 \) is the beam waist size also called the 1/e intensity half-width, and \( z_0 \) is the focal position where this beam waist occurs. We assume the mode to be propagating in the \(-z\) direction (for \(+z\), replace \( k \) by \(-k\), and to be normalized to 1: \( \int_{-\infty}^{+\infty} |g(x,z)|^2 dx = 1 \). With \( N \) grooves on each side of the central slit, the main equation (Equation (1)) is a matrix equation representing a linear system of \((2N+1)\) equations for \((2N+1)\) unknowns \( E_\alpha \), coupled via Green’s function matrix \( G_{\alpha\beta} - \epsilon_\alpha \) to the input excitation \( A_0 \).
An important fact is that the resulting matrix $G_{a\beta} - \epsilon_a$ has a Toeplitz structure, in addition to being symmetric around the main diagonal. This allows fast evaluation of the matrix elements, which helps in subsequent extensive optimizations.

It is convenient to calculate the coupling $T$ to the above Gaussian mode by the far-field overlap integral over azimuthal angle $\theta$:

$$T = |t|^2 = \frac{\int_{-\pi}^{\pi} |g(\theta)|^2 H_y^{\text{Norm}}(\theta)\,d\theta}{|A_0|^2}. \quad (4)$$

In Equation (4), the Gaussian profile is

$$g(\theta) = \sqrt{\frac{(ka_0)^2}{\pi}} e^{ikz_0 \cos \theta} e^{-(ka_0 \theta)^2/2},$$

with the sign convention such that $z_0 < 0$ is in front of the corrugated side, and

$$H_y^{\text{Norm}}(\theta) = \sqrt{\frac{ka}{2\pi}} \frac{\sin[(ka \sin \theta)/2]}{(ka \sin \theta)/2} \sum_{a=-N}^{N} E_a e^{-ikda \sin \theta}$$

is the far-field pattern of the emerging radiation. We note that the fields in Equations (5) and (6) are normalized so that $\int_{-\pi}^{\pi} |g(\theta)|^2 \, d\theta = 1$ and $\int_{-\pi}^{\pi} |H_y^{\text{Norm}}(\theta)|^2 \, d\theta = |A_0|^2 T_{\text{total}}$.

In the optimization, we can adjust the wavelength $\lambda$, groove depth $h$ and period $d$, and location of the Gaussian beam waist $z_0$ to meet the requirements on the coupling $T$, loss $L$, total transmittance to free space $T_{\text{total}} = T + L$, and reflectance back into the slit $R = 1 - T_{\text{total}}$.

Let us note that the assumptions made in this model restrict the quantitative accuracy of its predictions for real metals with complex refractive indices. However, its simplicity offers orders-of-magnitude speed-up compared to the full numerical solutions and enables quick observation of key parameter dependencies and rapid coarse optimization of the corrugated structures. The fine tuning still requires full numerical modeling and experimental feedback.

3. Model extensions to corrugations with aperiodic spacing and non-uniform depth

A nanoaperture with aperiodic corrugations may provide additional degrees of freedom to further increase the coupling to a Gaussian mode. In the aperiodic model, the groove period is allowed to vary ($d_a \neq d_p$ in general), but we still assume the symmetric geometry ($d_a = d_{-a}$) around the central groove 0. Intuitively, aperiodic grooves can help adjust the phase of the radiation to better match the Gaussian mode TEM$_{0p}$, increasing the overall coupling. Another reason to study the aperiodic case is to model realistic nanostructures, which are fabricated within certain practical tolerance.

The equations presented above need only a slight modification for the aperiodic case, and the numerical results are easily obtained. Specifically, to capture the effect of aperiodic groove width, $d_a$ should be replaced by $\sum_{i=1}^{d} d_i$ in the calculation of the elements $G_{a\beta}$ in Equation (2) and of the far-field in Equation (6).

The non-uniform groove depth model is exactly the same for periodic or aperiodic groove widths, and only the values $\epsilon_a$ are changed from $1/\tan kh$ to $1/\tan kh_a$ (while keeping $h_a = h_{-a}$). Thus, the depth non-uniformity can be imposed independently of the aperiodicity of the groove spacing.

The number of parameters that can be optimized in the aperiodic spacing and/or non-uniform depth case scales with the number of grooves and can be quite large. Since the search space over various parameters of the design is disjoint (certain designs are forbidden) and abundant with local maxima, we use direct search techniques like genetic algorithms (GA) combined with local search (gradient descent, Nelder-Mead) methods in a hybrid scheme to achieve the non-obvious optimal designs reported here. Hybrid GA schemes are superior to simple GA in terms of avoiding premature convergence and maintaining rich diversity in population to explore several local maxima, and climb the hills in design space that lead to the global maximum. Hybrid GA along with the aperiodic spacing and non-uniform depth models enables the following design capabilities:

- maximize or minimize any of $T, T/L, R = 1 - T_{\text{total}}$ at a given wavelength for aperiodic groove spacing and depth;
- optimization constrained by multiple criteria at a given wavelength;
- joint optimization of resonant structures for multiple wavelengths.

4. Results

First, we present the results corresponding to the periodic corrugations. Given the fact that in the practical situation the minimum obtainable Gaussian-beam waist size should scale proportionally to the wavelength, we consider two cases: in one the beam waist $a_0 = 1400 \text{ nm} \times \lambda/560 \text{ nm}$ scales with the wavelength (scaled-waist case) and in the other it is fixed at $a_0 = 1400 \text{ nm}$ (fixed-waist case). Both cases are chosen to yield the same result at wavelength $\lambda = 560 \text{ nm}$. The results of the calculations, optimized over corrugation depth $h$ and position of Gaussian beam waist $z_0$, are
shown in Figure 1(b) and Figure 2 for the scaled-waist case and in Figure 3 for the fixed-waist case, respectively. It is clear from Figures 1(b), 2 and 3 that, unlike the total coupling to free space $T_{\text{total}}$, the coupling $T$ to a Gaussian beam has only one main maximum corresponding to the optimum corrugation period $d = 0.85 \ldots 0.9 \lambda$ for the scaled-waist and $d = 0.8 \ldots 0.85 \lambda$ for the fixed-waist cases. In the fixed-waist case, the maximum ratio $T/L$ of coupling to loss is realized for long wavelengths, which is a byproduct of having a constant beam waist $a_0$ for all wavelengths (this means that the Gaussian beam has wider acceptance angle at longer wavelengths, which leads to a better overlap). In the scaled-waist case, the far-field acceptance angle is fixed, and the optimum wavelength is no longer at the boundary.

The values of coupling $T$ further optimized over the wavelength are shown in Figure 4 for various $N$ in the scaled- (a) and fixed-waist (b) cases. Up to 41% coupling in the scaled-waist case (up to 35% in the fixed-waist case) can be obtained at $N = 16$. The fully (and independently) optimized $T$ and $T/L$ (including optimization over corrugation period $d$) are shown in Figure 5(a) for both scaled and fixed beam waists. The curves show that there is very little need to go beyond $N = 4 \ldots 5$. It is interesting to note that for $N > 3$, the optimum position of the beam waist $z_0$ is in front of the metal film, i.e. the emergent field is slightly converging.

We can further improve the coupling and coupling-to-loss ratio by the use of aperiodic corrugations. The optimal coupling for $\lambda = 560$ nm in the periodic corrugation case considered above is $T = 30\%$, $T/L = 1.1$ with $N = 8$ grooves, groove period $d = 495$ nm and depth $h = 98$ nm. Optimization for maximum coupling $T$ at the same wavelength in the aperiodic corrugations model yields $T = 44\%,$
Figure 4. Coupling coefficient $T$, optimized over groove depth $h$, waist position $z_0$, and wavelength $\lambda$, shown as a function of the corrugation period $d$ for the (a) scaled-waist case $a_0 = 1400 \text{ nm} \times \lambda/560 \text{ nm}$ and (b) fixed-waist case $a_0 = 1400 \text{ nm}$. (The color version of this figure is included in the online version of the journal.)

$T/L = 1.15$, with $N=6$, groove spacing $d_{1-6} = [218, 503, 510, 534, 162, 324] \text{ nm}$ and depth $h = 101 \text{ nm}$, as shown in upper Figure 5(b). If the groove depth is also varied for optimum coupling, we get the best results of $T = 64\%$, $T/L = 3.8$ with $N=9$, groove spacing $d_{1-9} = [238, 145, 331, 152, 387, 455, 60, 514, 508] \text{ nm}$ and depth $h_{1-9} = [99, 95, 100, 94, 100, 12, 100, 105, 112] \text{ nm}$, as shown in lower Figure 5(b), demonstrating our $2\times$ improvement compared with the periodic case. Similar results are obtained at other wavelengths as well, indicating feasibility of high coupling ($T>50\%$) for any wavelength chosen in our range.

Comparing geometries with and without periodicity for the same number of grooves, we have also observed that the aperiodic corrugations offer not only extra degrees of freedom for optimization of nanocavity structures, but also similar design tolerances to fabrication errors. We have checked the design space in the neighborhood of the optimal geometry and found slowly degrading performance (coupling $T$ and coupling-to-loss ratio $T/L$) within a few nanometers from each geometry parameter. Figure 6 shows the result of perturbing the optimum designs of aperiodic corrugations (with varying groove spacing and depth) randomly within ranges of $\pm 1 \text{ nm}$ and $\pm 4 \text{ nm}$. The robustness of our designs to perturbations also indicates that we have indeed achieved a significant global maximum in the design space, without being misled by the local maxima or numerical artifacts.

We believe that the advantage of the aperiodic structures over the periodic ones mainly comes from their flexibility in shaping the wavefront curvature, whereas the periodic structures can only produce a very simple phase profile. As illustrated by Figure 7, the periodic corrugations produce nearly linear phase profiles with opposite slopes for $\alpha < 0$ and $\alpha > 0$. As a result, the magnitudes $|E_w|$ of the optimized periodic structure have to decay quickly with $\alpha$, in order to avoid the formation of far-field side lobes corresponding to these tilted wavefronts. The aperiodic structures, on the other hand, are not restricted to this simple phase dependence; for example, the phase profile for $|\alpha| = 2, 3, 4, \text{ and } 5$ in the aperiodic case with fixed depth and $N=6$ is parabolic. As a result, aperiodic structures support a larger number of corrugations with substantially non-zero values of $|E_w|$, which provides better flexibility in beam shaping and also enables smaller overall-size structures for the same total number of corrugations. A similar mechanism of improved performance of aperiodic structures compared to periodic ones has been recently observed for sub-wavelength dielectric gratings [7].

Using the aperiodic-corrugation model with uniform groove depth, we have also looked into the possibility of simultaneous optimization of the geometry for two or more wavelengths. This, for example, can be useful for short-pulse operation or for nanocavity-enhanced nonlinear interactions of several waves. The model is then used to optimize our nanocavity parameters of interest (total transmission, coupling, reflection, transmission to loss ratio, cavity field enhancement, etc.). Since the wavelengths involved might be quite different, we use the scaled-waist approach. In general, we expect that the structure optimized for several wavelengths should contain corrugations with several different periods. Under the constraint on the total number of corrugations, this is best approximated by an aperiodic corrugated structure. Once the rough parameters of a structure for reasonable coupling of all wavelengths have been found, the fine optimization becomes the trade-off between the number of supported wavelengths and the sacrifice in coupling efficiency.

We started by finding the geometry with $N=6$ for simultaneous reasonably high coupling at two wavelengths: $T = 20\%$ at both $\lambda_1 = 650 \text{ nm}$ and $\lambda_2 = 775 \text{ nm}$. We observed that the results can be improved in the cases where one wavelength is a multiple of the other. For
example, we could obtain $T = 30\%$ for $N = 9$ at both $\lambda_1 = 250\text{ nm}$ and $\lambda_2 = 1000\text{ nm}$. Although these preliminary results do not reach the highest coupling values of the single-wavelength case, they have so far been only optimized over a small parameter sub-space and are expected to improve after more extensive optimization.

We have also obtained the designs of the aperiodic nanocavity [8] optimized for three-wavelengths needed for efficient $\chi^{(2)}$ nonlinear optical interactions, such as optical parametric amplification, sum-frequency generation, etc. For these applications, different wavelengths may need to be optimized for different parameters. For example, in the case of sum-frequency generation for up-conversion of mid-infrared beams, one may want to build a single-, double-, or triple-resonant cavity for the signal, idler, and pump beams, with prime requirements of high coupling efficiency $T$ for one wave (e.g. the signal at $\lambda_s = 4100\text{ nm}$), high

Figure 5. (a) Coupling $T$ and coupling-to-loss ratio $T/L$ optimized over wavelength $\lambda$, waist position $z_0$, groove period $d$ and depth $h$ for both scaled-waist and fixed-waist cases. (b) Geometries optimized for maximum coupling $T$ at $\lambda = 560\text{ nm}$ with $N = 6$ (aperiodic model: varying groove spacing, uniform groove depth) and with $N = 9$ (aperiodic model: varying both groove spacing and depth). (The color version of this figure is included in the online version of the journal.)

Figure 6. Coupling $T$, optimized for wavelength $\lambda = 560\text{ nm}$ over all parameters of aperiodic model, and its sensitivity with respect to random variations of parameters of the optimum geometry within $\pm 1\text{ nm}$ and $\pm 4\text{ nm}$. (The color version of this figure is included in the online version of the journal.)
total transmission $T_{\text{total}}$ for another (e.g. for idler at $\lambda_i = 955 \text{ nm}$), and high cavity field enhancement for the third (e.g. high reflectance $(1 - T_{\text{total}})$ for the pump at $\lambda_p = 775 \text{ nm}$). An aperiodic model can satisfy such complicated requirements. In our preliminary work, we have found geometry with $N = 7$ that yields $T = 38\%$ for the signal, $T_{\text{total}} = 39\%$ for the idler, and $T_{\text{total}} = 63\%$ (i.e. $R = 99.4\%$) for the pump in this scenario.

5. Conclusion

To summarize, we have performed full-scale optimization of the coupling from a sub-wavelength metal slit surrounded by periodic or aperiodic corrugations into a fundamental Gaussian mode. We have optimized the values of the parameters of the metal nanostructure, such as the number of corrugations, their depth and spacing, and location of the Gaussian beam waist, to find maximum coupling $T$ and coupling-to-loss ratio $T/L$ for a given wavelength. With periodic corrugations, we have achieved maximum $T = 35\%$ and $T/L = 1.7$ for a Gaussian beam with waist size $a_0 = 1400 \text{ nm}$ independent of the wavelength, as well as $T = 41\%$ and $T/L = 1.14$ for that with beam waist size linearly proportional to the wavelength. Aperiodic corrugations improve this performance by approximately $2 \times$, provide tolerance to fabrication errors, and enable multi-wavelength optimization. A version of our solver was recently used to predict and optimize the emission patterns from a fabricated nanocavity (see figure 1(b) in [9]).

Acknowledgements

This work was supported in part by the Texas Higher Education Coordinating Board ARP program, the Air Force Office of Scientific Research Grant FA9550-06-1-0413, and the DARPA Contract No. HR0011-08-1-0063.

References