How small can “Nano” be in a “Nanolaser”? 

Jacob B. Khurgin¹,* and Greg Sun²

¹Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, 21218, USA, e-mail: jakek@jhu.edu
²Department of Physics, University of Massachusetts Boston, Boston, MA 02125, USA
*Corresponding author

Abstract

We show that the lasing threshold of the single mode metal-semiconductor nano-laser (spaser) is determined only by the photon absorption rate in the metal and exhibits very weak dependence on the composition, shape, size (as long as it is less than half-wavelength) and temperature of the gain medium. This threshold current is on the order of a few tens of micro-amperes for most semiconductor-metal combinations which leads to unattainably high threshold current densities for a substantially subwavelength laser (spaser). Therefore, in our view, surface plasmon emitting diodes, (SPEDs), operating far below “spasing” threshold may be a more viable option for the chip scale integrated nanophotonics.

Keywords: laser; plasmon; semiconductor.

1. Introduction

In the 50 years since the first semiconductor laser (SL) was demonstrated SLs have penetrated every corner of our everyday life and have become key components in long distance and local communication, display, data storage and processing and many other fields. The key performance characteristics of SLs are high efficiency, reliability, small size, and high speed. All these characteristics are engendered by the combination of the properties that is unique to SLs: the lasing transition is the allowed one with oscillation strength of the order of unity, and every atom in the lattice can participate in the lasing process. Because of this, even a small volume of semiconductor gain medium has gain high enough to compensate cavity losses, and packs enough power to deliver performance that no other type of laser can match. Therefore, one of the main thrusts in the development of SLs has been on shrinking their size. These developments have been enabled by the spectacular advances in nanofabrication. The rationale for reducing the size of the laser is three-fold – first of all it is the integration density; second, smaller lasers have lower threshold currents and thus operate at lesser power; third, the smaller the laser the faster it can operate. This happens because as the cavity gets smaller both the cavity round trip time and the device capacitance gets reduced.

Reduction of the SL size can only be achieved if both carriers and photons can be confined within a small volume while introducing no additional loss for either carriers or photons. Confinement of carriers to a few nanometers in one dimension as in Quantum wells (QWs) has been achieved as early as the 1980s, while 3-dimensional confinement on a slightly larger scale in quantum dots (QDs) has been successfully implemented since the 1990s. Development of high quality photonic crystal cavities has culminated in a single QD laser [1], whose operation serves as a proof that a tiny gain medium volume of $10^{-3}$ μm$^{3}$ is sufficient to achieve lasing threshold.

Confinement of photons, however, is a more challenging task, impeded by diffraction limit, according to which it is impossible to confine the photons of wavelength $\lambda$ in a dielectric medium with refractive index $n$ to a volume less than $(\lambda/2n)^3$. (Here the “dielectric” is a material with negligibly small conductivity at optical frequencies, i.e., this definition includes semiconductors.) In fact, the smallest lasers using all-dielectric structures all fall quite short of diffraction limit [1–3]. Therefore, the research has shifted from all-dielectric to metal-dielectric structures of various shapes, where the optical field combines with the oscillations of free carriers into the coupled modes – surface plasmons (SPs) that can be confined beyond diffraction limit [4, 5] in all three dimensions.

The interest in sub-wavelength lasers has been spiked by the revolutionary proposal by Bergman et al. [6] for a coherent emitter of SPs (spaser) consisting of a metal nanoparticle surrounded by a semiconductor gain material and the first estimate given in that work has shown that the population inversion required to compensate the loss in the metal was quite reasonable. Since then, a multi-faceted research effort, both theoretical [7–11] and experimental [12–24] has been devoted to spasing, but there had been only single demonstration of “laser-like” behavior in gold-dye complexes that were subwavelength in all three dimensions [12]. The indication of lasing consisted of some linewidth narrowing and vaguely nonlinear output behavior under intense pulsed optical pumping [12]. When it comes to developing a practical semiconductor spaser, a great number of diverse schemes had been successfully demonstrated, including mostly optical but also injection [15] pumping. Various geometries, such nanowires [13], double heterostructure mesas [14, 15], disc structures [16], nanopatches [17], and coaxial pillars [18–20] have been used, but they all have one feature in common: the device was larger than $\lambda/2n$ in at least one dimension. The threshold has been universally high and with a few exceptions, the pumping was either optical or required low temperature. A number of review works [7, 8, 21–24] recently published confirm the fact that truly sub-wavelength, i.e., $<\lambda/2n$ in all 3 dimensions, electrically pumped spaser has not been demonstrated, which, of course has not dampened the enthusiasm for the development of the elusive sub-wavelength coherent source.
At this point, as the work on nano-laser goes on for about a decade, perhaps it is worthwhile to pause and reflect on how far this road is going to take us, and whether there exists a hard limit to how far a coherent source can be shrunk. The question of how small can this ubiquitous “Nano” in “Nanolaser” be is a pressing contemporary question. This is precisely the objective of our short discourse in which we come to a rather spectacular and unexpected conclusion. We demonstrate that the threshold of true subwavelength spaser is almost independent  

of most of the parameters that as a rule greatly affect the threshold of conventional laser. This list includes size, shape, material composition, and temperature of gain medium. The lowest possible threshold current of true sub-λ spaser depends almost exclusively on the rate of free carriers scattering in the metal, $γ_{e}$, and the expression for this threshold current is strikingly simple, $I_{th}=\frac{n_{c}}{\gamma_{e}}$. Since for noble metals $γ_{e}$ is on the order of $10^{14}$ s$^{-1}$, the threshold current cannot be possibly $<10$ mA, which for small spaser volumes of less than $(\lambda/2\pi)^{3}$ leads to enormous current densities in excess of $10^{5}$ A/cm$^{2}$ even assuming that there is no losses in the device other than due to metal absorption.

2. Inherent loss in sub-wavelength cavities

We begin with a short and almost intuitive explanation of the above conclusion, and then provide results of rigorous calculations confirming it. As a starting point we note that for the case when the fields are confined within a volume that is much smaller than a wavelength, a so-called quasi-static limit applies, when one can neglect retardation effects (i.e., temporal derivatives) and solve the Poisson equation for the electric field in place of the wave equation, with insignificant magnetic field. This result follows from integrating Maxwell’s equation $\nabla \times H = j_{0} e^{-i \omega t} + j_{\omega} E$ inside the confinement volume with characteristic dimension $a << \lambda /n$ and using Stokes theorem to obtain order-of-magnitude relation $H = e_{0} \epsilon_{0}^{0} \omega \gamma_{e} n^{2} a^{2} /2$. The time averaged value of magnetic energy $U_{m} = H^{2}/2$ is related to the time-averaged value of electric energy $U_{e} E^{2} / 2$ as $U_{m} / U_{e} = (\pi n a d)^{2}$, hence for strongly sub-wavelength confinement $a << \lambda /n$ magnetic field does not play any significant role.

Since in the oscillating mode inside the dielectric the energy is stored half of the time in the form electrical energy, and the other half in the form of magnetic (inductive) energy, it follows that in all dielectric structures the mode cannot be confined within sub-wavelength volume, and will radiate, which of course is the alternative interpretation of the diffraction limit from the energy conservation point of view. If, however, the free carriers are present, the energy can be transferred from the electric field to the kinetic motion of the carriers (current) and back enabling self-sustaining oscillations (SPs) no matter how small is the volume [4, 5]. That is why metal or another conductor is necessary to break the diffraction limit. But breaking the limit comes at an exceedingly high price – since half of the time all energy is contained in kinetic motion of carriers, the overall effective rate of energy loss of the oscillating mode $γ_{eff}$ will be one half of the rate of energy loss inside the metal. The latter rate is twice the rate of momentum scattering inside the metal, $γ_{m}$, hence we arrive at startling conclusion that $γ_{eff} = γ_{m} + γ_{rad}$ for any mode that is significantly sub-wavelength in all 3 dimensions, no matter what is the exact size, shape, and dielectric composition. This has been first noted in [25] and shown to be true for various shapes in [26] where we have also shown that at frequencies that are lower than $γ_{m}$ (which is on the order of 10–20 THz for noble metals) [27] the mode losses decrease as magnetic field increases when conductivity current, rather than displacement current, becomes the dominant factor, which explains why the various sub-wavelength metal-dielectric structures are far less lossy at low frequencies than in optical range. In addition to the loss in the metal there is also a radiative loss of the mode $γ_{rad}$ that is in most cases significantly smaller than non-radiative loss for small particles and we can write for overall effective loss $γ_{eff} = γ_{m} + γ_{rad}$

3. Defining the threshold of a spaser

So, the loss in the sub-wavelength structure is quite high, but can it be compensated and then superseded by optical gain in the surrounding dielectric (semiconductor) to enable a spaser? Most of the researchers have approached this issue from the point of compensating positive imaginary part of the dielectric constant of the metal, $\epsilon_{m}^{0} = n_{m}^{2} \epsilon_{m}/\omega$ by the negative dielectric constant of gain medium (semiconductor) $\epsilon_{s}^{0} = -g_{s} n_{s}/\omega$, where $g_{s}$ is the gain coefficient. Neglecting the relatively small radiative loss, the compensation condition $\epsilon_{eff}^{0} + \epsilon_{m}^{0} = 0$ then leads to the threshold gain coefficient $g_{th} = m_{th} / n_{th}$, where $n_{th}$ for noble metal/semiconductor this amounts to about a few thousand per cm gain coefficient that is attainable at carrier densities of $< 10^{19}$ cm$^{-3}$ – a number that is high but is certainly achievable in state-of-the-art QW lasers [28]. It is this result that has supplied the much needed enthusiasm to the research community and commenced the race for the smallest-on-earth nanolaser.

While there is nothing wrong with the above estimate of the threshold carrier density, it is the threshold current density that is far more relevant when it comes to practical implementation, and in order to estimate it we should first consider proper definition of laser threshold. In conventional large scale lasers the threshold is always assumed to occur when the round trip loss $\Gamma_{r}$ is balanced by the round trip gain $G_{r}$, or, dividing both round trip gain and loss by round trip time one uses the threshold condition $g_{th} = \Gamma_{r}$ where $\Gamma_{r}$ is the aforementioned rate of energy loss of the oscillating mode and $g_{th}$ is gain per unit time, or the stimulated emission rate. The rate equation for the number of photons in the lasing mode $N_{p}$ and the number of carriers $N_{c}$ are [28].

$$\frac{dN_{p}}{dt} = (g_{th} - \gamma_{eff} )N_{p} + g_{th} n_{sp}$$

$$\frac{dN_{c}}{dt} = R_{p}^{-1} g_{th} N_{p} - \beta_{sp} n_{sp} - \gamma_{sp}$$

where $\gamma_{sp}$ is the spontaneous decay rate, and $g_{th} n_{sp}$ is the spontaneous emission into the lasing mode.
The other two parameters in (1) are the excess noise factor \( n_p \geq 1 \) and the fraction of spontaneous radiation going into the lasing mode \( \beta \). Note that the equations are written for the total number of particles in the laser and not for their densities. The simple relation between the stimulated and spontaneous emission terms in (1) is the effect of fundamental relation between Einstein’s B and A coefficients.

Now, when the total number of modes into which the spontaneous radiation can take place is very large, i.e., \( \beta \) is very small, and the spontaneous emission term \( g_{sp} n_p < \beta R_n \) on the right-hand side can be neglected leading to the conventional threshold condition \( g_{sp} = \gamma_{sp} \). However, the situation becomes drastically different when the laser cavity contains just a few, and ultimately a single mode that is strongly coupled to the gain medium. Then one can see immediately that the spontaneous photons emitted into the mode can no longer be ignored and definition of the threshold becomes murky. In fact, the conventional threshold condition cannot really be reached in the steady state.

What distinguishes laser radiation from that of a conventional incoherent source is the degree of its coherence usually characterized by the degree of second order coherence [29], which in laymen terms represents the relation between the number of coherent photons generated by the stimulated emission, and the incoherent spontaneously emitted photons. Therefore it has been suggested in [30] that a more suitable definition of lasing threshold, suitable for lasers of all sizes is the condition when the rate of stimulated photon emission begins to overtake the rate of spontaneous photon emission. The authors in [30] have proposed the condition of \( N_{sp} = 1 \) and experiments in [1] have confirmed that as the number of photons exceeds unity the second order coherence also approaches unity.

One can see from (1) we only need to slightly modify the above condition by an excess noise factor \( n_p \) that is close to unity in order to achieve perfect balance between spontaneous and stimulated photons. Furthermore, it is easy to see that when the newly-introduced threshold condition \( N_p = n_p \) is reached the stimulated emission rate becomes \( g_{sp} = \gamma_{sp}/2 \), i.e., equal to precisely one half of the effective loss rate, hence the total loss rate becomes \( \gamma_p = \gamma_{sp}/2 \). Since the cavity loss rate is the same as the observed emission linewidth, we can see that when \( N_p = n_p \), the linewidth of the emission gets reduced by a factor of two, which in itself is also a good gauge of laser threshold.

4. Threshold current of a true spaser is always higher than 10 \( \mu \)A

But what is the main implication of the above theory when applied to the spaser? By substituting the threshold condition into the second equation in Eq.(1), one obtains for the steady-state threshold pump rate

\[
R_{th} = \gamma_{sp} n_p + \frac{1}{2} (\beta^{-1}) \gamma_{sp} n_p + R_{SB} > \gamma_{sp} n_p
\]

(2)

Multiplying (2) by the electron charge yields our no-longer-surprising conclusion that the minimum threshold current of a laser is \( I_{th, min} = e \gamma_{eff} n_p > e \gamma_{sp} n_p \) which is now easily interpreted as follows: in order to achieve a reasonable degree of coherence one must support at least \( n_p \) photon in the lasing mode, which requires one electron-hole pair to be injected every \( 1/\gamma_{sp} n_p \) seconds. For a typical value of metal scattering rate of 10 fs [27] we obtain the aforementioned absolutely minimum value of threshold current of 16 \( \mu \)A.

In a relatively large scale laser with multiple modes, \( \beta \) is very small and obviously depends on mode shape and volume. Therefore the second term in (2) dominates the first one. Furthermore, the total nonradiative decay rate \( R_{SB} \) is proportional to the volume of the active medium and can be significant. As a result, in a relatively large scale laser the threshold current strongly depends on volume, shape, confinement factor, and through the relation between the gain, \( n_p \) and non-radiative recombination, also on the nature of the gain medium and temperature. But as the volume gets smaller, and gradually becomes sub-wavelength, only a single mode remains strongly coupled to the gain medium, and \( \beta \) approaches unity.

Also, the total number of carriers decreases with volume and so does the total nonradiative loss rate. For the spaser with \( \beta^{-1} \), we then obtain from (1) the rate equations for the density of SPs \( N_{sp} \) and carriers \( N_C \):

\[
\frac{dN_{sp}}{dt} = g_{sp} (N_{sp} + n_p)(\gamma_m + \gamma_{rad})N_{sp}
\]

\[
\frac{dN_C}{dt} = e I - g_{sp} (N_{sp} + n_p)
\]

(3)

where \( I = e R_p \) is the pump current. The steady state solution of (3) immediately yields for the output radiation power

\[
P_{out} = h \nu \gamma_{rad} N_{sp} \frac{\hbar \nu}{e \gamma_m + \gamma_{rad}} I
\]

(4)

i.e., an expected threshold-less characteristic. It should be noted that rather than using conventional dipole radiation out-coupling, one can consider near-field coupling the energy of SPs into plasmonic waveguides or simply near field focusing onto a target. That would improve out-coupling efficiency, but at the same time make threshold current only higher. The threshold-less characteristic (4) follows from a simple conservation argument: the decay into the SP mode is the dominant outlet for the decay of electron-hole pairs, whether it is via spontaneous or stimulated emission – hence all the input power goes in the SP mode with a portion of it emerging from the lasing cavity.

5. Modeling a spaser

To test our main conclusion – that threshold of the true sub-wavelength laser depends only on the metal loss, we have performed detailed modeling of a purported nanolaser of Figure 1 designed to operate in the 1300 nm range at the fundamental (dipole) SP mode of a prolate gold spheroid with the half axis ratio of \( b/a = 0.425 \) surrounded by the gain material, In\(_{33}\)Ga\(_{67}\)As that is lattice-matched to InP, with a thin (5 nm) spacer made of wider gap In\(_{52}\)Al\(_{48}\)As to avoid surface
recombination. This injection scheme is rather hypothetical, involving highly doped n and p layers of InP separated from each other by proton-bombarded material in order to provide current confinement. Practicality and efficiency of such an injection method are uncertain, perhaps a different type of confinement using nanoantennae is more practical, but at any rate, our goal is to provide the threshold current estimate for the best case scenario of 100% efficient injection and to show that even then the threshold is unsustainably high for true subwavelength lasers.

The threshold is plotted vs. the size of nanoparticle for 3 different temperatures in Figure 2. As expected, the threshold current exhibits weak size dependence rising $<20\%$ as the nanoparticle length is reduced from 90 to $<10$ nm. This increase occurs because the fraction of energy residing in the metal does increase slightly. To test that the threshold only weakly depends on the gain material we have performed calculations assuming different values of the transition strength in In$_{0.53}$Ga$_{0.47}$As [31] – for as long as these values were within the same order of magnitude the threshold changes were $<5\%$. This is easy to understand due to the fact that the only material-dependent variable in our expression for threshold is the excess noise factor, $n_{ep}=f_{c}(1-f_{c})(f_{v}-f_{c})$, where $f_{c,v}$ are Fermi factors of the conduction and valence bands. Since high gain is always required to compensate large metal loss, and the effective mass of electron is much less than that of a hole, conduction band Fermi factor $f_c$ is always close to unity at laser frequency and $n_{ep}=1$. Of course, using materials with transition strength orders of magnitude less than that in In$_{0.53}$Ga$_{0.47}$As would render lasing impossible as the satellite valleys in the conduction band would start being populated. But for most direct bandgap semiconductors [31] one can say that threshold barely depends on exactly what material is used.

But what is most unexpected to anyone familiar with SLs is extremely weak temperature dependence of the threshold current. Yet this result becomes obvious once one realizes that the threshold depends so strongly on temperature in conventional SLs because, as the tails of Fermi-Dirac distributions $f_{c,v}$ extend beyond the quasi-Fermi levels there exist substantial population of “idle” carriers at higher energies. The spontaneous recombination of all these “hot” electron-hole pairs increases current. In addition, the current is increased because Auger recombination is temperature dependent. But in the nano-laser or spaser just the electron-hole pairs that are resonant with the one and only mode contribute to the current – their number depends only on the width of the mode and exhibits very weak temperature dependence. Auger recombination is also not much of the factor because of the small total number of carriers. The only decrease in threshold may come from the reduction of metal scattering rate itself with temperature, but that decrease is relatively small [32–34]. Thus the temperature dependence of threshold is not nearly as strong as the exponential law in conventional SLs.

6. Discussing the repercussions

Having established the fact that the threshold current of a nanolaser whose size is sub-wavelength in all three dimensions is always on the scale of $10^{-5}$ A we need to consider the implications. First of all, it makes very little sense to shrink the size of the laser far beyond the wavelength – as the threshold current stays constant, the current density increases quadratically, reaching 1 MA/cm$^2$ for the nanoparticles that are about 50 nm long. That current density is clearly unsustainable even in the pulse mode, and one should not forget that we have not even touched the issue of the damage from excessive heating caused by dissipating upwards of $10^{11}$ W/cm$^2$.

As we have already mentioned, the main benefits of reducing the volume of a laser are reduction of current and increase in speed. This strategy works for as long as the dimensions are larger than wavelength, but then the Purcell’s effect [35] gets into the action and reduces spontaneous time proportionally to the volume $\tau_{spon}\sim(V/\lambda)\tau_{spon,0}$, where $\tau_{spon,0}$ is the bulk spontaneous time. (Note that in our derivation Purcell’s effect is present implicitly as total spontaneous emission stays constant while the volume shrinks.) Since the threshold current can be written as $I_{th}=eN_{sp}V/\tau_{spon,0}$ the threshold is no longer volume dependent, and shrinking the size only causes the same current to go into smaller volume – obviously this is not a good strategy.
It has been suggested [21] that high thresholds of nanolasers are nothing but the growing pains that would eventually be overcome as it happened with conventional injection lasers. Indeed in the early 1960s the room temperature threshold of the first injection lasers was 100’s of kA/cm² [36] but then rapid progress ensued and by the end of the decade, room temperature double heterostructure lasers [37] have been operated with threshold current densities of <1 kA/cm². But such a rapid process was based on the fact that from the very inception of SLs it had been well understood that the cause of high threshold was the absence of carrier and light confinement, and the method for overcoming this problem, double heterostructure had been identified [38, 39] right away. Once inevitable material growth related problem had been solved the threshold had quickly come down by orders of magnitude.

In contrast, our less than enthusiastic conclusion for the single mode spaser is based entirely on the best case assumptions, with all the carrier and plasmon confinement assumed to be already implemented. The high threshold current density is the consequence of two of the most basic laws of physics: Maxwell equations, and the relation between the first and second Einstein coefficients. Short of repealing these laws the only hope lies in reducing the metal loss, but this goal would essentially require engineering of new materials on the atomic level [40].

That leaves one to consider what are the alternatives to the true sub-wavelength on all three-dimensional nanolaser. Obviously, for the structures that are elongated along at least one of the dimensions the effective loss of the mode is decreased as less energy gets transferred into the kinetic motion of electrons and back. This is, in fact, the geometry of all the lasing devices in [13–20]. Therefore it is not unfeasible to obtain injection pumped lasing in a modal volume less than (λ/2n)³, but only with a cavity longer than λ/2n. In other words, the lasers with sub-diffraction mode cross-section of much less than (λ/2n)² are quite feasible.

Yet even with such a cavity the threshold will remain high, and to the best of our knowledge there is no clear direction for its reduction anywhere in sight.

7. Conclusions and recommendations

This brings us back to the beginning of this article where we have identified three major benefits of small size: integration density, low power dissipation, and high operational speed. Neither integration density nor low power dissipation is the prerogative of lasers – light emitting diodes are characterized by both small size and high quantum efficiency. Shrinking the size of the diode below the wavelength is bound to increase quantum efficiency even further as Purcell-enhanced radiative recombination would overwhelm nonradiative channels. What is most important is that Purcell enhancement would also increase the speed of the light-emitting diode, making it a good choice for many applications. Purcell factors in a plasmonic nanoparticle of 50 nm length can be a few hundred which would make the effective recombination time only a few picoseconds, which, while longer than 10 fs stimulated recombination time, is still sufficient for ultra-fast signal processing. In fact, it is not easy to envision any application of nano-laser (or spaser) which would require high degree of coherence. Therefore, in our view, the most promising avenue of development of on-chip sub-wavelength metal-semiconductor sources is not the spaser operating above threshold, but Surface Plasmon Emitting Diode (SPED), operating far below lasing threshold, yet having both speed and efficiency required in nanophotonics circuit applications.

References

[34] Parks GR, Lawrence WE, Christy RW. Intraband optical conductivity \( \sigma(\omega, T) \) of Cu, Ag, and Au: Contribution from electron-electron scattering. Phys Rev B 1981:23:6408.

Received May 22, 2012; accepted June 20, 2012