Ultrafast dynamics of optically-induced heat gratings in metals - more complicated than expected

Yonatan Sivan* and Marat Spector†

School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, P.O. Box 653, Israel, 8410501, and Department of Physics, Ben-Gurion University of the Negev, P.O. Box 653, Israel, 8410501

E-mail: sivanyon@bgu.ac.il

Abstract

Diffusion of heat in metals is a fundamental process which is crucial for a variety of applications of metal nanostructures. Surprisingly, however, ultrafast heat diffusion received only limited attention so far. Here, we study heat diffusion by identifying the underlying (femtosecond and few picosecond) time scales responsible for the generation and erasure of optically-induced transient Bragg gratings in thin metal films. We show that due to the interplay between the temporal and spatial nonlocal nature of the thermo-optic response, the heat diffusion dominates the temperature dynamics only when it occurs on a ~ 1 picosecond time scale; when it is slower or faster, the spatially-local (but temporally-nonlocal) effects of electron thermalization and electron-phonon energy transfer dominate the dynamics. Further, we show that heat diffusion affects

*To whom correspondence should be addressed
†School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, P.O. Box 653, Israel, 8410501
‡Department of Physics, Ben-Gurion University of the Negev, P.O. Box 653, Israel, 8410501
also the nonlinear optical response such that the overall change of the permittivity (hence, reflectivity) has a significant dependence also on the illumination period rather than only on the illumination intensity.

Keywords

Ultrafast heat dynamics, Plasmonics, Transient gratings

Intro

Heat generation and dynamics is central to our understanding light-matter interactions in metals, specifically for the purpose of separating thermal effects from (non-thermal) electronic effects. It is also critical for many applications and specifically, for the correct interpretation of the role of heat in nanophotonic applications such as surface femtochemistry and plasmon-assisted photocatalysis.

The standard model of heat generation and dynamics in metals is the well-known Two Temperature Model (TTM) in which a distinction is made between the electron subsystem and the phonon subsystem (each assigned by a temperature, $T_e$ and $T_{ph}$, respectively); Further, the electron system is assumed to become thermalized instantaneously, hence, before significant energy transfer between the electrons and the phonons occurred. Under these conditions, the dynamic equations for the electron and phonon temperatures are coupled via a simple linear term, namely,

\begin{equation}
C_e(T_e) \frac{\partial T_e(t)}{\partial t} = -G_{e-ph}(T_e - T_{ph}) + p_{abs}(\vec{r}; t), \\
C_{ph} \frac{\partial T_{ph}(t)}{\partial t} = G_{e-ph} (T_e - T_{ph}).
\end{equation}

Here, $C_e$ and $C_{ph}$ represent the heat capacity of the electrons and phonons, respectively; $G_{e-ph}$ is the electron-phonon coupling factor, representing the rate of energy exchange be-
between the electrons and the lattice; \( p_{\text{abs}} \), is the density of absorbed photon power. These equations describe the decay of the electron temperature following an initial heating stage (which sometimes for convenience is even skipped altogether) due to energy transfer to the phonons; in conjunction, it describes the phonon heating. The TTM was originally introduced in Ref. [16] and derived in detail from the semi-quantum mechanical Boltzmann equation for the few picosecond regime in Ref. [17]. Unfortunately, as well known, the assumption underlying the TTM is never strictly valid - the thermalization of the electron subsystem extends roughly over the first picosecond after the excitation pulse exited the metal\(^{17-20}\).

As a remedy, using a unique model that ensures energy conservation, we derived in Ref. [4,5] an extended version of the TTM (referred to below as the eTTM) from the semi-quantum mechanical Boltzmann Equation. In this model, the early stages of the thermalization of the electron subsystem are accounted for via the exact electron distribution and the total energy of the non-thermal electrons; the latter is characterized by a fast (pulse-duration limited) rise time and slow decay time, corresponding to the thermalization of the electron subsystem. This derivation confirmed the phenomenological model presented much earlier in Ref. [18,21] as well as the classical derivations in Ref. [22,23]. In contrast to the TTM, the eTTM allows the electron subsystem to be non-thermal, and only assumes (e.g., by adopting the relaxation time approximation, or adopting a rigorous many-body formulation\(^{17,24}\)) that the electron subsystem can be described by some temperature - the one to which the electron system would have relaxed if it was isolated from the photons and phonons, see discussion in Ref. [4,5]. In that sense, within the validity conditions of the Boltzmann description, the eTTM is an exact coarse-grained description of the energy dynamics; it is valid at all times (in particular, also before all the non-thermal energy is depleted) and the temperature varies instantaneously upon absorption of photons as it reflects the total energy of the electron system.

Most research on the heat dynamics in metals to date focussed on the derivation of
the thermal properties, the coupling dynamics and the temporal dynamics. However, quite peculiarly, early studies regularly ignored heat diffusion. It seems that this neglect suited most of the configurations studied experimentally. Indeed, on one hand, nanometric metal particles or thin films are characterized by a uniform electric field and hence, uniform temperatures. In fact, the temperature is uniform even for metal nanostructures extending to several tens or even hundreds of nanometers, due to the strong heat diffusion in metals. Furthermore, whenever a large beam was used for illumination, diffusion was a slow process (see Eq. (2) below) occurring only on the edges of the illumination spot, away from where the processes of interest occurred. However, on the other hand, it is easy to appreciate that there are several scenarios in which heat diffusion cannot be neglected. These include, for example, thick metal layers into which light penetration is far from being complete, metal-dielectric composites, diffusion of a localized heat spot in a thin film (studied in Ref. [31]), highly non-uniform illumination of micron-scale metal objects etc.

In this manuscript, we focus on a specific case of the last scenario mentioned above, namely, on the diffusion of a periodic heat pattern (aka transient Bragg gratings, TBGs) in a thin metal film. In this scenario (depicted schematically in Fig. 1), two pump pulses are interfered onto a metal film to create a periodic pattern of absorption; in turn, the generated heat and the temperature dependence of the metal permittivity gives rise to a periodic permittivity modulation that outlives the pulse, but self-erases due to heat diffusion. Such a pattern can enable transient reflectivity of guided modes in the metal film that can be turned
on and off on a sub-picosecond time scale, as for free-carrier gratings.\textsuperscript{34} This property is appealing for ultrafast switching applications, as it may enable switching speeds significantly faster than what is commonly achieved with free-carrier generation in semiconductors.\textsuperscript{35} To the best of our knowledge, TBGs in metal films were studied before only in the context of the slow (nanosecond and microsecond) phonon dynamics (see, e.g., Ref. [36–39]), where the electron thermalization and sub-picosecond temperature dynamics were conveniently ignored. Studies of faster dynamics were performed in Ref. [40] for time scales of several hundreds of picoseconds and in Ref. [28] on a scale of few tens of picoseconds. The neglect of the subpicosecond dynamics might have also originated from the un-availability of a proper theoretical tool valid in that temporal range.

![Figure 1: (Color online) A schematic illustration of an optically-induced temperature grating, and the dynamics of its contrast ($\Delta T(t)$, the temperature difference between red and yellow regions).](image)

Accordingly, using the (all-time valid) eTTM, we focus in the current manuscript on the subpicosecond and few picosecond spatio-temporal heat dynamics in optically-induced TBGs in metals, specifically, on the interplay between the numerous time scales - the pump duration (via $p_{abs}$), the thermalization rate (defined as $\Gamma^{NT}$, see Eq. (4) below), $e-ph$ energy exchange rate ($\Gamma^{Te} = G/C_e$, see Eq. (1)) on one hand, and the diffusion time $\tau_{diff}$ on the other hand. Based on standard (i.e., single temperature) heat equations, the latter is defined
using the second-order (spatial) moment of the solution of the diffusion equation subject to the periodic modulation, namely,

\[ \tau_{\text{diff}} \equiv \frac{d^2}{4\pi^2 D_e}, \quad D_e \equiv \frac{K_e}{C_e}. \]  \hspace{1cm} (2)

Here, \( D_e \) and \( K_e \) are the diffusion coefficient and thermal conductivity of the metal electrons; \( d \) is a characteristic length scale associated with the initial temperature non-uniformity; here, it is simply the illumination period. The magnitude of the diffusion constant \( D_e \) of metal electrons is approximately \( D_e \approx 100 - 300 \text{ cm}^2\text{s}^{-1} \) (as e.g., demonstrated experimentally\(^{31}\)). With sufficiently strong inhomogeneities of the illumination (specifically, small periods, in the current context), the heat diffusion can become as fast as a few hundreds of femtoseconds.

We show, somewhat surprisingly, that due to the non-trivial interplay between the temporally- and spatially-\textit{nonlocal} nature of the thermo-optic response of metals (namely, thermalization and \( e - ph \) energy transfer vs. heat diffusion), heat diffusion dominates the dynamics only in a narrow temporal regime, specifically, when the diffusion time (2) is between the thermalization (\( \sim 0.5\text{ps} \)) and \( e - ph \) heat transfer (\( \sim 1\text{ps} \)) time scales; when the diffusion time is different, the dynamics is dominated by the latter \textit{spatially-local but temporally-nonlocal} effects. Furthermore, we show that the gratings do not get totally erased, but rather persist for many picoseconds. This unique combination of spatial and temporal \textit{nonlocality} explains why the standard trade-off of the nonlinear optical response of materials between speed and strength\(^{41}\) can be surpassed by metals (see discussion in Ref. [42]), enabling the thermo-optic nonlinearity in metals to be both fast and strong. We also show, contrary to our initial expectations that TBGs in metals are unsuitable for many ultrafast switching applications, except in the extreme spatially-nonlocal limit.

Finally, we also show that the spatial nonlocal nature of the nonlinear thermo-optic response of the bulk metal (not to be confused with the \textit{structural} nonlocality in metal-dielectric composites, see Ref. [43]) affects the overall magnitude of the nonlinear response.
This newly observed effect might have been playing a role in the multiplicity of reported values of the intensity-dependent optical nonlinearity of metals and their composites (see e.g., Ref. [44,45]) and provides yet another motivation to avoid describing the nonlinear optical response by a nonlinear susceptibility, a concept which was derived in the context of spatially- and temporally-local (i.e., Kerr) electronic nonlinearities.

While the results of the current manuscript rely on a rather generic “minimal” model applied to metals only, in its last part we describe several aspects of the model that can be improved for more accurate modelling of specific metal systems, and refer briefly to the possibility of observing similar effects in other materials systems like semiconductors, 2D materials etc..

Spatio-temporal dynamics of the electronic and phonon temperatures

Model - the extended Two Temperature Model (eTTM)

We adopt here the extended Two Temperature Model (eTTM)\textsuperscript{4,5} which describes the spatio-temporal dynamics of $U_e^{NT}$, the non-thermal (NT) electron energy, as well as of the electron and phonon (lattice) temperatures. In this approach, $U_e^{NT}$ is determined by\textsuperscript{4,5,18,22;1,2}

$$
\frac{\partial U_e^{NT}(\vec{r}, t)}{\partial t} = - (\Gamma_e + \Gamma_{ph}) U_e^{NT}(\vec{r}, t) + \langle p_{abs}(\vec{r}, t) \rangle.
$$

(3)

Here, $\Gamma_e$ represents the rate of exchange of energy between the non-thermal electrons and the thermal ones, or in other words, it represents the thermalization rate of the electron

\textsuperscript{1}Note that in Eq. (3), we neglect diffusion of the non-thermal electrons. A proper determination of their contribution to the dynamics involves a full non-equilibrium description of the problem; it lies beyond the scope of the current study. See further discussion of ballistic electrons below.

\textsuperscript{2}Note that the eTTM does not capture the increase of rate of energy transfer from electrons to the lattice during the thermalization time, as discussed in Ref. [19]; however, this effect should have, at most, a modest quantitative effect on the issues discussed in the current work.
distribution as a whole. It was computed and measured to be on the order of several hundreds of femtoseconds (see e.g., Ref. [19,21,23,46] to name just a few of the studies of the thermalization). In Ref. [18] it was found experimentally that $\Gamma_e$ is independent of the laser fluence in the range of $2.5 - 200 \mu J/cm^2$, corresponding to an electron temperature rise of up to about 200K. Similarly, $\Gamma_{ph}$ represents the rate of energy transfer between the non-thermal electrons and the phonons. Due to the similarity in the physical origin, it is set to the same value chosen for the energy transfer rate between the thermal electrons and the phonons ($\Gamma_{T_{ph}}$), see below. Together they form the total decay rate of the NT energy (also referred to as the thermalization rate) denoted by

$$\Gamma^{NT} = \Gamma_e + \Gamma_{ph}. \quad (4)$$

The second term on the RHS, $\langle p_{abs}\rangle$, describes the time-averaged power density of absorbed photons. Using the Poynting theorem for dispersive media and assuming that all the absorbed photon energy is converted into heat, together with the assumption of a slowly varying envelope approximation in time, it can be written as

$$\langle p_{abs}(\vec{r}, t) \rangle = \frac{1}{2} \epsilon_0 \epsilon''_m \omega_{pump} \langle |\vec{E}(\vec{r}, t)|^2 \rangle, \quad (5)$$

where $\epsilon_m$ is the metal permittivity (and $\epsilon''_m$ is its imaginary part), $\omega_{pump}$ is the pump frequency, $\vec{E}$ is the local electric field and $\langle \rangle$ stands for time-averaging over the period.

The equations for the temperatures are

$$C_e(T_e) \frac{\partial T_e(\vec{r}, t)}{\partial t} = \nabla [K_e(T_e, T_{ph}) \nabla T_e] - G_{e-ph}(T_e - T_{ph}) + \Gamma_e \mathcal{U}_e^{NT}(\vec{r}, t),$$

$$C_{ph} \frac{\partial T_{ph}(\vec{r}, t)}{\partial t} = \nabla [K_{ph}(T_{ph}) \nabla T_{ph}] + G_{e-ph}(T_e - T_{ph}) + \Gamma_{ph} \mathcal{U}_e^{NT}(\vec{r}, t). \quad (6)$$

In these coupled heat equations, $\mathcal{U}_e^{NT}$ serves as the heat source, $C$ and $K$ are the heat capacities and thermal conductivities of the electrons and the lattice as denoted by subscripts.
and, \( T_{\text{ph}} \); \( G_{e-ph} \) (mentioned above) represents the rate of energy exchange between the electrons and the lattice which occurs on time scales defined as \( \Gamma_e \equiv G_{e-ph}/C_e \) and 
\( \Gamma_{T_{\text{ph}}} \equiv G_{e-ph}/C_{\text{ph}} \), respectively; Since \( \Gamma_e \gg \Gamma_{T_{\text{ph}}} \), the latter is essentially negligible. Also worth noting is the relative smallness of the phonon diffusion \( K_{\text{ph}} \), which is henceforth neglected. Also neglected is heat coupling to the environment and consequent cooling, as it occurs typically on time scales much longer than considered in the current work.

The model described above is used below to characterize the ultrafast heat diffusion in the periodically-illuminated metal film. We emphasize that in order to maintain simplicity and without loss of generality, inclusion of more advanced modelling (e.g., accounting more accurately for thermalization,\(^{20}\) for the temperature dependence of the optical and thermal parameters,\(^{47,48}\) for potential anisotropy of the parameters, for differences between different metals, for quantum mechanical effects etc.) is intentionally avoided in order to keep the generality and simplicity of the discussion. Those effects are discussed briefly at the end of manuscript and would be included in future studies.

**Numerical results**

Solving the complete eTTM requires knowledge of the temperature dependence of all parameters. In the current manuscript, we neglect the temperature-dependence of all parameters (most notable of which is of the heat capacity), as appropriate for low intensities whereby the temperature rise is modest.\(^{48}\) Any changes associated with such temperature dependence make only modest quantitative changes to the results shown below. Under these conditions, the solution of Eq. (3) was shown in Ref. [49] to be

\[
U_e^{NT}(\vec{r}, t) = \frac{\sqrt{\pi} \tau_{pump}}{2} p_{abs}(\vec{r}, t)e^{\frac{\Gamma^{NT} \tau_{pump}^2}{4 - \Gamma^{NT} t}} \left[ 1 + erf \left( \frac{t}{\tau_{pump}} - \frac{\Gamma^{NT} \tau_{pump}}{2} \right) \right].
\]

(7)

For a short pump pulse, \( U_e^{NT} \sim \delta(t) e^{-\Gamma^{NT} t} \), such that the pump pulse and NT energy peak together. For a longer pump pulse, the NT energy becomes maximal later than the
peak of the pump pulse. Furthermore, the non-instantaneous thermalization gives rise to a temporally smeared source for the electron temperature compared to the source appearing in the TTM$^{16}$ (which is just $\langle p_{abs} \rangle$ (5)).

We now solve Eqs. (3)-(6) for the case in which the metal film is optically thin (compared to the optical penetration (skin) depth into the metal) and the illumination is periodic ($p_{abs} \sim \cos(2\pi x/d)$). Accordingly, we assume that all quantities depend only on $x$ and $t$ and look at the dynamics within a single period $d$. In addition, we set the pump pulse to be the shortest time-scale in the system (i.e., $\tau_{pump} \ll \tau_{diff}, 1/\Gamma^{NT}, 1/\Gamma^{Te}, \ldots$); the latter choice is motivated a-posteriori by the weak dependence on the pump duration observed in further simulations (not shown). The parameters used in the solution of the $eTTM$ (3)-(6) are given in Table 1, suitable for Au; the parameters for Ag and Al are quite similar. We assume that the pump illuminates the sample near its plasmon resonance with a local field of $\sim 30MV/m$.\footnote{Note that we avoid specifying the local intensity, as it is a somewhat improper quantity to use in the context of metals. Indeed, the negative real part of the permittivity causes the fields within the metal to be primarily evanescent, hence, not to carry energy (such that the Poynting vector, hence, intensity vanish, at least in the absence of absorption). Instead, we use the local density of electromagnetic energy, by specifying the local electric field, which is easy to connect to the incoming field.}

Importantly, the comparisons below are performed for the same local field (hence, absorbed power density) within a single unit-cell\footnote{This allows us to avoid the complication associated with the strong angle and frequency dependence of the relation between the incoming and local fields.}.

Table 1: Parameters used in the numerical solution of the $eTTM$ (3)-(6).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_e$</td>
<td>$2 \cdot 10^4$</td>
<td>$Jm^{-3}K^{-1}$</td>
<td>Ref. [50]</td>
</tr>
<tr>
<td>$C_{ph}$</td>
<td>$2.5 \cdot 10^6$</td>
<td>$Jm^{-3}K^{-1}$</td>
<td>Ref. [50]</td>
</tr>
<tr>
<td>$G_{e-ph}$</td>
<td>$2.5 \cdot 10^{16}$</td>
<td>$Jm^{-3}K^{-1}s^{-1}$</td>
<td>Ref. [51]</td>
</tr>
<tr>
<td>$\Gamma_e$</td>
<td>$2 \cdot 10^{12}$</td>
<td>$s^{-1}$</td>
<td>Ref. [18]</td>
</tr>
<tr>
<td>$\Gamma^{NT}$</td>
<td>$2 \cdot 10^{12}$</td>
<td>$s^{-1}$</td>
<td>Ref. [18]</td>
</tr>
<tr>
<td>$\Gamma^{Te}$</td>
<td>$1.25 \cdot 10^{12}$</td>
<td>$s^{-1}$</td>
<td>Ref. [18]</td>
</tr>
<tr>
<td>$\Gamma^{T_{ph}}$</td>
<td>$10^{10}$</td>
<td>$s^{-1}$</td>
<td>Ref. [18]</td>
</tr>
<tr>
<td>$K_e$</td>
<td>315</td>
<td>$Wm^{-1}K^{-1}$</td>
<td>Ref. [50]</td>
</tr>
<tr>
<td>$\tau_{pump}$</td>
<td>100</td>
<td>$fs$</td>
<td>—</td>
</tr>
<tr>
<td>$\lambda_{pump}$</td>
<td>500</td>
<td>$nm$</td>
<td>—</td>
</tr>
</tbody>
</table>
Figs. 2(a)-(c) show maps of the spatio-temporal dynamics of $T_e$ for different illumination periods $d$ (hence, different diffusion times). Figs. 2(d)-(f) show the corresponding cross-sections of $T_e$ at $x_{\text{max}} = 0$ (where the illumination intensity is maximal), and at $x_{\text{min}} = d/2$ (no illumination; $U_e^{\text{NT}}(x_{\text{min}}) = 0$). In all cases, $T_e(x_{\text{max}})$ initially builds up due to thermalization (i.e., due to energy transfer from the NT electrons to thermal ones); This occurs on a time scale $\Gamma_e^{-1}$ which is longer than the pump duration, hence, this is referred to as a temporally-nonlocal effect. The maximal temperature rise with respect to room temperature is $\sim 90\%$, $\sim 70\%$, $\sim 55\%$ for the three cases we simulated. The temperature at $x_{\text{max}}$ then gradually decreases due to both $e-ph$ energy transfer and electron heat diffusion. The latter effect causes $T_e(x_{\text{min}})$ to build up as well, despite not being initially illuminated (hence, this is a spatially-nonlocal effect). Figs. 2(g)-(i) show the electron temperature grating contrast, defined as $\delta T_e^g \equiv T_e(x_{\text{max}}) - T_e(x_{\text{min}})$. We can identify 2 limits -

- The **spatially-local limit**, $1/\Gamma_e^{NT}, 1/\Gamma_e^{Te} \ll \tau_{\text{diff}}$, i.e., when the electron heat diffusion is much slower than the local relaxation of $T_e$ (see Figs. 2(a) & (d)). In this case, most of the electron heat is transferred to the lattice before any significant amount of heat reaches the minimum point from the maximum point. As a result, the high temperature region is localized and the electron temperature grating contrast is maximal.

- The **spatially-nonlocal limit**, $1/\Gamma_e^{NT}, 1/\Gamma_e^{Te} \gtrsim \tau_{\text{diff}}$, i.e., when the electron heat diffusion is faster than the local energy relaxation of NT electrons (see Figs. 2(c) & (f)). In this case, the fast diffusion of electron heat limits the heating of the illuminated regions, while promoting the heating of the un-illuminated regions (as also observed previously in Ref. [30]). Consequently, the high temperature region is relatively delocalized and the electron temperature grating contrast, $\delta T_e^g$, is minimal.

For any choice of parameters between these two extremes (e.g., Figs. 2(b) & (e)), a non-negligible amount of heat is transferred from the maximum point to the lattice before it reaches the minimum point.
Further simulations show that as the pump duration is increased (at a fixed pulse energy), the temporal maximum of $T_e(x_{\text{max}})$ drops only slightly and the temperature at the minimum remains roughly the same, so that the electron temperature grating contrast decreases only slightly as well. Accordingly, we do not pursue the study of longer pulses.

Figure 2: (Color online) Spatio-temporal dynamics of the temperature of the electrons $T_e$ for an optically thin metal film when the diffusion time (Eq. (2)) and period $d$ are (a) $\tau_{\text{diff}} = 3.6\,\text{ps}$ and $d = 1.6\,\mu\text{m}$, (b) $\tau_{\text{diff}} = 430\,\text{fs}$ and $d \approx 0.5\,\mu\text{m}$ and (c) $\tau_{\text{diff}} = 150\,\text{fs}$ and $d \approx 0.3\,\mu\text{m}$; temperature range plotted is 300K to 600K in all plots. (d)-(f) Cross-sections of the maps directly above, specifically, $T_e(x_{\text{max}})$ (solid line), and $T_e(x_{\text{min}})$ (dashed line). (g)-(i) The corresponding grating contrasts in log scale (blue solid lines), along with the exponential fits of the decay stage (dashed red lines).

In order to quantify the decay dynamics of the grating contrast, we plot in Fig. 3(a) the
temporal maximum of $T_e(x_{\text{max}})$ and $T_e(x_{\text{min}})$ for different values of $\tau_{\text{diff}}$. One can see a weak dependence of the maximal values except for very short diffusion where $T_e(x_{\text{max}})$ drops and $T_e(x_{\text{min}})$ increases. Accordingly, the grating contrast $\delta T_e^g$ decreases for shorter diffusion times (Fig. 3(b)).

We now define $\tau_g^e$ as the characteristic time scale in which the electron temperature contrast is erased. To determine $\tau_g^e$, we fit the decay stage of the grating contrast to an exponential function (see Fig. 2(g)-(i)). Potentially unexpectedly, one can see in Fig. 3(c) that when the diffusion time is long, the grating is erased much faster than the diffusion time (2); this happens due to the spatially-local but temporally-nonlocal effect of $e-ph$ energy transfer which occurs roughly on a $1/\Gamma T_e$ time scale$^5$. When the diffusion is faster, then, the grating erasure time drops slightly but is then limited by the thermalization rate, $1/\Gamma N T$. This effect originates again from the temporally-nonlocal nature of the thermo-optic response of the metal. In particular, the absorbed energy is “converted” into heat non-instantaneously, on the time scale of $1/\Gamma N T$, such that the grating contrast erasure is smeared over a similar time scale. Mathematically, it is indeed easy to see that in the extreme spatial nonlocal limit ($d \to 0$, hence, $\tau_{\text{diff}} \to 0$), the $e-ph$ coupling term is negligible with respect to the diffusion term, such that the dynamics of the electron temperature is dictated by the convolution between the temporally smeared source (7) and the impulse response of the system (namely, the Green’s function of the single temperature heat equation). Simulations performed with reduced thermalization time (as appropriate e.g., for graphite$^{53}$) indeed confirm this interpretation. The peaking time of the temperature at the minimum ($T_e(x_{\text{min}})$) exhibits a similar trend but slightly delayed; this happens because the temperature rise at $x_{\text{min}}$ starts only after the thermalization time $1/\Gamma N T$ (more accurately, a fit shows $\sim 1.35/\Gamma N T$).

$^5$The simulations show that the erasure time is can becomes slightly slower than $1/\Gamma T_e$; this is possible because unlike $\tau_g^e$, the $e-ph$ energy transfer time $1/\Gamma T_e$ is not defined exactly as the exponential decay time of the contrast decay.
We now look again at Fig. 2(g)-(i). Peculiarly, the later stages of the dynamics reveal that the rapid grating contrast slows down significantly. Additional numerical simulations of the complete eTTM (not shown) show that the diffusion proceeds at the phonon diffusion rate (as already shown in Ref. [31,40]); clearly, this occurs due to the coupling of the electrons to the phonons (the grating erasure is indeed complete if the electrons are artificially decoupled from the phonons). Within the few picosecond range, the residual contrast is \( \lesssim 1\% \), specifically, it is a few degrees for the local case but, as one might expect, it is much lower, a fraction of a degree, in the nonlocal case; naturally, the (residual) contrast decreases for a stronger electron thermal conductivity.

In order to complete the description of the heat dynamics, we also show the spatio-temporal dynamics of the phonon temperature. One can observe a behaviour reminiscent of that of the electron temperature in the local and nonlocal limits. Importantly, the actual phonon diffusion is faster than its intrinsic value because of the fast electron heat diffusion and the spatially-local \( e-ph \) energy transfer. We also observe the non-zero residual phonon temperature grating contrast in the few picosecond regime; this residual phonon grating contrast decreases...
contrast is, however, much stronger than for the electron temperature - it is 100%, 50%, 30% of the maximal phonon temperature rise, respectively, for the simulations in Fig. 2, specifically, at a few degree level.

Figure 4: (Color online) Same as Fig. 2 for $T_{ph}$; temperature range plotted is 300K to 308K in all plots.

**Permittivity dynamics**

We use the temperature-dependent permittivity model described in Ref. [31] to describe the spatio-temporal dynamics of the metal permittivity $\epsilon_m(T_e, T_{ph})$ induced by the change in the temperatures. This dynamics thus determines the thermo-optic nonlinear response of the
metal to the incoming light. Specifically, the probe pulse is centered at 850nm for which the contribution of interband transitions to the permittivity can be lumped into a single (temperature-independent) constant $\epsilon_\infty$. Accordingly, we refer below only to $\epsilon_{\text{m, intra}}(T_e, T_{ph})$.

Fig. 5 shows an initial rise of the permittivity due to the $T_e$ rise and a corresponding relative permittivity change below 1% for the cases plotted in Fig. 2 (compare to a 90% relative change in the electron temperature). At later stages, the sensitivity to the phonon temperature becomes dominant such that for the 1% change of $T_{ph}$, a 0.1% change of the permittivity is obtained. This implies on much higher sensitivity of the permittivity to the phonon temperature, and a permittivity thermoderivative of $\sim 0.1/K$; such values are comparable to those found in the ellipsometry measurements described in Ref. [54].

It is interesting to note that the permittivity change occurs mostly due to the change in the imaginary part of the permittivity. This is similar to the theoretical and experimental observations in the CW case reported in Ref. [25,48,55]. For Ag, the changes of the real part of the permittivity play a non-negligible role.\textsuperscript{48}

Overall, the above results show that the permittivity change depends on the interference patten (i.e., on the period) even though the absorbed power density are the same for all cases. More generally, the nonlinearity at short times (and to a lesser extent, also in the picosecond time scale) changes with the spatial frequency contents of the illumination. This implies that in addition to the temporal nonlocal nature of the metal response (see Ref. [52]), the nonlinear thermo-optic response of metals has also a spatial nonlocal aspect, an aspect which may contribute to the multiplicity of values assigned for the nonlinearity across the literature (see, e.g., Ref. [44,45]).
Discussion

Our simulations show that, quite peculiarly, the heat diffusion in metals determines the grating erasure time only in a rather narrow regime of timescales in which $0.5\text{ps} \sim 1/\Gamma NT \ll \tau_{diff} \ll 1/\Gamma Te \sim 1.25\text{ps}$. Out of this regime, the grating erasure time is dominated by spatially-local but temporally-nonlocal effects ($\Gamma NT$ and $\Gamma Te$, respectively) and the diffusion determines directly the build up time of the heat in the minimum illumination points for short diffusion times and mediates the heating level of the strongly illuminated regimes. While
this result is expected for micron-scale periods (weak diffusion), it might be less intuitive for sub-micron periods (strong diffusion). It originates from the relatively long (mildly sub-picosecond) thermalization times of metals.\textsuperscript{19,20,46,56} This also means that heat diffusion might be important over a wider range of parameters for materials with nearly-instantaneous thermalization such as graphite or graphene\textsuperscript{53,57} but less important in a semiconductor such as GaAs for which the thermalization is even slower.\textsuperscript{58} This exemplifies that there are quite a few aspects of ultrafast heat diffusion which deserve further research. This is especially true in the several picosecond time-scales where significant acoustic dynamics occurs.\textsuperscript{59–61}

The results shown in this manuscript are robust to the various approximations made in our model. Specifically, the phonon diffusion does not change the dynamics described above (indeed, it is a much weaker effect compared with the electron-induced phonon diffusion); there is also only weak sensitivity to the pump duration and to the metal layer thickness. Indeed, previous studies (e.g., Ref. [31,32,40] showed that accounting for the finite thickness gives rise only to modest quantitative changes in the temperature dynamics. Yet, our result might be sensitive to anisotropy in the diffusion and to structuring.\textsuperscript{32,43}

Our analysis, and especially the identification of the nonlocal nature of the permittivity change, is also essential for understanding the nonlinear thermo-optic response of metals. In particular, it shows that the nonlinear thermo-optic response depends not only on the pump and probe wavelengths, the illumination intensity and duration (see e.g. Ref. [44]), but also on the spatial distribution of the pump, namely, its spatial (in)homogeneity. Our work will hopefully also motivate similar studies of heat dynamics in additional high mobility materials such as graphene, and other 2D materials.\textsuperscript{62,63} From the applied physics perspective, our results show that since the grating contrast does not disappear on a picosecond time scale, TBGs applications based on metals are expected to work only for very short diffusion times (strong diffusion, nonlocal case), so that they might be less attractive compared to transient dynamics of free carriers in semiconductors.\textsuperscript{34}

Finally, while the eTTM handles correctly the temperature dynamics at all times, the
diffusion of the non-thermal electrons (i.e., within the first half picosecond or so, before complete thermalization) was neglected. In this regime, heat transport might even be ballistic. While some papers discuss this regime (see e.g., Ref. [1,6]), they introduce ballistic transport phenomenologically, sometimes even within the eTTM formulation. However, it is clear that a proper model for ballistic transport has to be derived from the Boltzmann equation, something that to the best of our knowledge, was not done so far. Experimentally, there have been some studies of vertical heat transport based on a configuration in which the pump and probe are incident on opposing interfaces of a metal film (see Ref. [26,28]), however, for lateral heat transport, there has been no evidence for ballistic diffusion even when using femtosecond temporal resolution, see Ref. [31]. Convincing models and measurements of heat transport in this regime will also be useful for resolving arguments regarding charge transfer on the femtosecond scale which are relevant for non-thermal (the so-called “hot”) electron studies, and their application for photocatalysis and photodetection\(^4,5,64\) and should eventually be combined with models of additional transport effects associated with interband transitions (see e.g., Ref. [65,66]).

**Acknowledgements.** The authors thank A. Block, P. Y. Chen, S. Sarkar and I. W. Un for many useful discussions. Y. S. was partially supported by Israel Science Foundation (ISF) Grant No. 899/16.

**References**


(42) Ou, J.-Y.; Plum, E.; Zhang, J.; Zheludev, N. I. Giant Nonlinearity of an Optically

Podolskiy, V. A.; Zayats, A. V. Designed ultrafast optical nonlinearity in a plas-
monic nanorod metamaterial enhanced by nonlocality. *Nature Nanotechnology* 2011,
6, 107111.

(44) Boyd, R. W.; Shi, Z.; Leon, I. D. The third-order nonlinear optical susceptibility of

748.

(46) Saavedra, J. R. M.; Asenjo-Garcia, A.; de Abajo, F. J. G. Hot-Electron Dynamics and

(47) Sivan, Y.; Chu, S.-W. Nonlinear plasmonics at high temperatures. *Nanophotonics* 2017,
6, 317–328.

(48) Gurwich, I.; Sivan, Y. A metal nanosphere under intense continuous wave illumination
- a unique case of non-perturbative nonlinear nanophotonics. *Phys. Rev. E* 2017, 96,
012212.

(49) Masia, F.; Langbein, W.; Borri, P. Measurement of the dynamics of plasmons inside
individual gold nanoparticles using a femtosecond phase-resolved microscope. *Phys.
Rev. B* 2012, 85, 235403.


59, 1460–1463.


