One-dimensional thin-film phonon transport with generation

A. Bulusu

Interdisciplinary Graduate Program in Materials Science
Vanderbilt University, Nashville, TN, USA

D.G. Walker *

Department of Mechanical Engineering
Vanderbilt University, Nashville, TN, USA

Abstract

The Boltzmann transport equation is often used for non-continuum transport when the mean free path of phonons is of the order of device sizes. One particular application involves heat generation in electronic devices. In a highly scaled MOSFET, for example, the majority of the heat is produced in a localized region immediately below the gate on the drain side. The size of this generation region is often smaller than the mean free path of phonons, which suggests the generation Knudsen number is large and non-continuum models are appropriate. Using a one-dimensional BTE and diffusion equation, a comparison between the continuum and non-continuum models is made. The focus of this comparative study is the behavior of each model for various Knudsen numbers for the device size and generation region. Results suggest that non-continuum distributions are similar to continuum distributions except at boundaries where the jump condition results in deviations from continuum distributions. Furthermore, the peak energy in a device predicted using the noncontinuum formulation is always less than that of the continuum model regardless of generation Knudsen number, which is in contrast to other prevailing studies.

Key words: Boltzmann transport, phonon, heat generation, hot spot

1 INTRODUCTION

Use of non-continuum models for thermal transport of highly scaled devices is usually justified by comparing the mean free path of phonons to characteristic device...
dimensions. The ratio of mean free path (l) to device size (L) is often termed the thermal Knudsen number, which is defined as Kn = l/L. If Kn ≪ 1, then continuum models such as Fourier’s law and the diffusion equation are adequate to model thermal transport. For devices where Kn is near or above unity, non-continuum effects dominate, and transport requires non-equilibrium models.

The argument for using non-continuum solutions to study heat generation and dissipation in microelectronic devices is similar to the argument for transport in scaled devices. If the generation region is small compared to the mean free path of phonons then non-continuum models are required [1]. This is often referred to as the hot-spot problem. In these cases, the generation Knudsen number Kn_g = l/L_g where L_g is the characteristic size of the generation region is often introduced. It has been suggested that the localization of the generation (large Kn_g) results in an increase in local energy [2,3]. However, if we consider that transport is largely ballistic in the region of generation for small generation regions, then the non-equilibrium phonons will travel beyond the generation region before a scattering event occurs. The energy deposited as a result of the localized generation then occurs over a volume larger than the actual generation region (for Kn_g > 1).

Non-continuum effects such as those found in large Knudsen number devices are very clearly seen by examining a one-dimensional film with an imposed temperature difference (see Figure 1). The temperature distribution in such devices with large Kn exhibits a jump at the boundaries and a smaller slope compared to the continuum limit. Many researchers use this case to validate more complex simulations, and so these results appear in numerous publications [4–8]. In fact, this canonical plot is repeated in the results section as a validation for the non-continuum formulation in the present work.

The pressure to reduce the size of microelectronic devices has largely driven the trend toward nanoscale science and technology. In fact, most nanoscale fabrication and characterization capabilities were developed originally for microelectronic devices. Consequently, microelectronic devices and materials are particularly interesting to study in terms of thermal transport [9]. The majority of thermal modeling of microelectronic materials has been relegated to thermal conductivity prediction of reduced-scale structures [10,11] or detailed examination of dispersion effects [12]. Likewise, thermal experiments on nanoscale devices involve mostly thermal conductivity measurements [13–15]. Nevertheless more effort is being devoted to actual transport including generation of thermal energy at reduced scales recently [16, e.g.]. Microelectronics devices provide ideal systems to study because thin-film devices such as SOI MOSFETs provide reduced scales with highly localized heating [17–19]. This particular application is significant because the continued scaling of practical CMOS devices is limited not by manufacturing technology but by heat removal of production technologies (i.e. air cooling) [20]. In the present work, localized heat generation in MOSFET type structures is used as motivation for comparing the BTE with continuum models with generation.
Highly-scaled devices with generation have been studied theoretically before, and evidence of large increases in the local temperature have been shown. Both Narumanchi et al. [7] and Yang et al. [16] show nearly a 200% predicted increase in the temperature rise between continuum and non-continuum models. These results lend credence to the hot-spot argument. However, the question remains unresolved as identified by Sinha et al. [21] who predict a 13% increase in local temperature due to small scale generation. All of these works attempt to explain the local increase in terms of phonon dispersion with varying degrees of complexity. However, to the best of our knowledge, measurements have not shown any unwarranted increase in local temperature due to the hot-spot size. The present work represents an attempt to understand the non-continuum generation problem in terms of a reduced order model with a limited number of parameters without the need for advanced dispersion relations.

A non-continuum analytic model of a one-dimensional system with generation is developed to help understand how non-continuum effects manifest themselves in devices with generation. The current work investigates the difference between non-continuum and continuum models to identify where continuum models break down. Furthermore, although non-continuum models have been used to predict non-continuum energy distributions, an analytic solution to the Boltzmann equation has not been provided that includes the generation of phonons. Therefore, this work can be used as a benchmark for more complex non-continuum models.

2 THEORY

2.1 Non-continuum

The derivation of the steady-state blackbody emissive power distribution for various Knudsen numbers (Kn) without generation is well known [22]. However, the development is repeated here with emphasis on the non-homogeneous generation term because this solution is central to the discussion. The intensity formalism that follows can be found in many radiation texts [23, e.g.] due to the similarity between phonon and photon transport [4]. In fact, Boltzmann transport for phonons is often called the equation of phonon radiative transport (EPRT) to reinforce the similarity to the equation of radiative transport (ERT). In contrast to the development of most radiation expressions, the generation in the present case is an additional non-homogeneity compared to the equilibrium intensity as will be shown.

The steady-state Boltzmann transport equation for phonons is written in terms of an intensity using the relaxation time approximation as

\[ \frac{\mu v}{\tau} \frac{dI}{dx} + \frac{I - I_0}{\tau} = g^* \quad \Rightarrow \quad \frac{\mu}{\tau} \frac{dI}{dx} + I - I_0 = \tau g^* = g, \tag{1} \]
where the coordinate \( x \) is non-dimensionalized with the quantity \( v \tau \) such that the real spatial coordinate \( x^* = v \tau x \). \( \mu = \cos \theta \) is the direction cosine, and \( I_0 \) is the equilibrium intensity. In the foregoing analysis the magnitude of the phonon group velocity \( v \) and scattering rate \( 1/\tau \) are assumed constant. This approximation is equivalent to the gray assumption, so the intensity is independent of frequency. The combined parameter \( l = v \tau \) is the mean free path of phonons. The generation rate \( g \) is a source term that represents an increase in local intensity resulting from electron phonon scattering, for example. Usually this term is not present in the ERT because radiation originating from the bulk is usually considered equilibrium in nature. The form of \( g \) is initially treated as arbitrary, but will have specific known forms for subsequent analysis.

The solution to the governing equation is given as a combination of positively moving phonons and negatively moving phonons as

\[
I^+(x, \mu) = C_1 \exp\left(-\frac{x}{\mu}\right) + \int_0^x \left[I_0(x') + \frac{g(x')}{\mu}\right] \exp\left(-\frac{x-x'}{\mu}\right) dx' \quad \mu > 0, \tag{2}
\]

\[
I^-(x, \mu) = C_2 \exp\left(-\frac{x-L}{\mu}\right) + \int_L^x \left[I_0(x') + \frac{g(x')}{\mu}\right] \exp\left(-\frac{x-x'}{\mu}\right) dx' \quad \mu < 0. \tag{3}
\]

For now, the integration constants \( (C_1 \text{ and } C_2) \) will be left as unknown constants but will be determined from emissive power at the boundaries of the film. Note that these constants are independent of the generation or the intensity distribution within the medium. This is a result of a thermalizing boundary often called a black body because it absorbs all incident energy and emits based on its equilibrium temperature only. The prescribed boundaries are therefore, actually flux boundary conditions, not temperature boundaries. Because the boundary is in equilibrium, the emissive power is uniquely determined by the temperature. Therefore, the boundary is often considered a specified temperature boundary, yet the constants represent an energy flux. Note that the generation \( g(x) \) is an arbitrary function of \( x \) as is the equilibrium intensity \( I_0(x) \), and both are treated as a non-homogeneity in the equation. Both quantities are energy fluxes per solid angle and represent the energy transported through a unit area. The two terms, however, represent different mechanisms and will be treated independently at a later step.

The heat flux is obtained from the intensity as an integration over the hemisphere yielding

\[
J(x) = J^+(x) + J^-(x) = 2\pi \int_0^1 \mu I^+(x, \mu) d\mu + 2\pi \int_{-1}^1 \mu I^-(x, \mu) d\mu. \tag{4}
\]

Each term on the right hand side can be expanded by plugging equations 2 and 3
into equation 4 such that

\[ J^+(x) = 2\pi C_1 \int_0^1 \mu \exp \left( -\frac{x}{\mu} \right) d\mu + 2\pi \int_0^x \left[ I_0(x') + g(x') \right] \int_0^1 \exp \left( -\frac{x-x'}{\mu} \right) d\mu \, dx', \]

\[ = 2\pi C_1 E_3(x) + 2\pi \int_0^x \left[ I_0(x') + g(x') \right] E_2(x-x') \, dx', \quad \text{and} \quad (5) \]

\[ J^-(x) = 2\pi C_2 \int_0^{-1} \mu \exp \left( -\frac{x-L}{\mu} \right) d\mu + 2\pi \int_L^x \left[ I_0(x') + g(x') \right] \int_0^{-1} \exp \left( -\frac{x-x'}{\mu} \right) d\mu \, dx', \]

\[ = 2\pi C_2 E_3(L-x) + 2\pi \int_L^x \left[ I_0(x') + g(x') \right] E_2(x-x') \, dx'. \quad (6) \]

The exponential integral \( E_n(x) \), along with identities and properties involving this special function can be found in Ref. [24]. The total heat flux is the sum of the heat flux components in each direction. In the absence of generation, the heat flux is constant across the domain such that energy is conserved. Therefore, we can write \( dJ(x)/dx = 0 \). With generation, the heat flux must still satisfy energy conservation, which is written as

\[ \frac{dJ(x)}{dx} = g(x). \quad (7) \]

This represents a departure in the derivation from the canonical no generation case. The condition is satisfied by summing equations 5 and 6 and differentiating.

\[ C_1 E_2(x) + \int_0^x \left[ I_0(x') + g(x') \right] E_1(x-x') \, dx' - 2[I_0(x) + g(x)] + C_2 E_2(L-x) - \int_L^x \left[ I_0(x') + g(x') \right] E_1(x'-x) \, dx' = -\frac{g(x)}{2\pi}. \quad (8) \]

The foregoing expression is a non-homogeneous integral equation in \( I_0(x) \) if we assume \( g(x) \) is known. Although we can not generate a closed-form solution, we can use any number of numerical integration approaches to approximate the solution.

2.2 Continuum

The Boltzmann solution for \( I_0(x) \) (equation 8) should recover the continuum solution for large domains. The one-dimensional heat diffusion equation provides a temperature distribution for continuum systems that can be compared to the non-continuum results. The normalized diffusion equation with generation is given as

\[ \frac{d^2T}{dx^2} = -G(x), \quad (9) \]

The solution depends on the functional form of the generation term \( G(x) \). In the case of no generation, the solution reduces to a linear distribution, and the integration constants are determined to satisfy the boundary conditions. For specified temperature boundaries, the solution is unique. For heat flux conditions on both
sides, the solution is not unique and an additional constraint must be imposed on
the boundaries. In the case of no generation, both boundaries must have the same
prescribed heat flux to balance energy. Even then, the temperature distribution can
only be determined relative to an integration constant. This feature is different than
the non-continuum case and highlights the difficulty in comparing the two solution
approaches.

For the case of uniform generation, the solution is obtained via direct integration,
where the boundaries conditions are specified temperature of zero.

\[ T(x) = \frac{L^2 G x}{2L} \left( 1 - \frac{x}{L} \right) \]  \hspace{1cm} (10)

where the first two terms represent the linear no generation component for left
and right boundary temperatures of \( B_1 \) and \( B_2 \) respectively. The generation term
results in a parabolic temperature distribution. Again, for heat flux boundaries, the
system must balance energy and the solution can only be determined relative to an
unknown integration constant.

Because we are interested in comparing results of the two models for various \( Kn_g \),
the continuum solution for a Dirac delta represents the limiting case of infinite \( Kn_g \).
If non-continuum solutions are warranted for large \( Kn_g \), then the Dirac delta rep-
resents the worst case scenario. The solution for homogeneous temperature bound-
daries is the Green’s function.

\[ T(x) = \begin{cases} 
G x (L - x')/L & \text{for } x < x', \\
G x' (L - x)/L & \text{for } x > x',
\end{cases} \]  \hspace{1cm} (11)

where \( x' \) is the location of the delta source, and \( G \) is the magnitude of the energy
generation.

2.3 Model comparison

When comparing the continuum solution to the non-continuum solution we need to
be cognizant of the following issues.

(1) The coordinate \( x \) was defined in terms of the mean free path. We are using
this same normalized coordinate for the continuum solution even though this
quantity does not have any real physical significance in the continuum limit.
However, we intend to keep this structure because we want to compare the two
models despite the presumed lack of applicability of the continuum model at
large \( Kn \). Note that the continuum solution is independent of \( Kn \).

(2) The definition of “temperature” in the two cases is fundamentally different. In
general, we regard equilibrium energy and temperature as related, so the real
question is whether the definition of equilibrium energy is equivalent for both systems. For non-continuum situations, the quantity $I_0$ represents an average energy of the left moving phonons and the right moving phonons emitted from their respective boundaries and the medium. It is not the equilibrium energy of the phonon system at any point since equilibrium does not exist. In the continuum limit, though, $I_0$ reduces to the continuum equilibrium energy. This difference between the definitions, though, is exactly what we intend to exploit to identify non-continuum effects.

(3) The boundary designation in each model is also not necessarily equivalent. In both cases, the boundaries are considered to be thermalizing boundaries. In the continuum model this means that the temperature is specified explicitly. In the non-continuum model, we specify the equilibrium (black body) emissive power at a boundary emitted into the system.

(4) The two models do not depend on the same material properties. The continuum model requires the thermal conductivity, whereas the non-continuum model requires the phonon mean free path. While these two parameters are phenomenologically related, an exact relation is difficult to obtain and dependent on a number of other factors [25]. However, we can eliminate the dependence of these parameters through suitable normalization.

3 RESULTS

The steady solution is shown in Figure 1 to demonstrate that the derivation and calculation can recover the expected behavior of transport through a thin film without generation. The jump condition at the boundary is due to the lack of interaction between oppositely moving phonons. Because scattering is limited for systems whose dimensions are of the order of the mean free path or smaller (large Kn), the phonons emitted from the boundary travel ballistically across the domain where they are thermalized at the opposite boundary. The intensity plotted in Figure 1 is simply the average energy of the left and right moving phonons. As the device becomes larger (small Knudsen numbers) more scattering causes the phonons from each boundary to equilibrate. In the continuum limit, both phonon systems are in equilibrium everywhere and the continuum solution is recovered.

For both generation cases, the boundary conditions are fixed at zero so the effects of the generation can be studied. For the non-continuum simulations, the boundaries impose an intensity supplied to the film from the contacts; in the continuum simulations, the temperature is held at a prescribed value. Even though both are set to zero, the boundaries are not necessarily equivalent.

To understand the energy distribution resulting from generation and the effects of a reduced generation region compared to the mean free path of phonons, two limiting cases are considered. In the first case, a uniform generation that spans the
entire domain is considered. In the second case, a delta function of heat generation, which represents the smallest possible generation region, is considered. The uniform generation case would model something like an extremely thin resistor where Joule heating is responsible for the thermal generation. This case, however, does not mimic the hot-spot problem because the generation region is considered to be the same size as the device domain.

For uniform generation, the resulting intensity distributions for various Kn are shown in Figure 2. The shape of each distribution is similar to the parabolic distribution of temperature found from the continuum solution. In fact a fit of the non-continuum data with a parabola result in an RMS deviation from the parabola of 0.3%. The results also demonstrate that the magnitude of the distribution grows with larger systems (small Kn). This also mirrors the behavior of a continuum system because more energy is introduced into the system as it becomes larger. Despite the zero boundary conditions, we still see a jump condition at the contacts. The jump is a result of ballistic phonons traveling to the thermalizing boundaries before they are equilibrated. Consequently, phonons produced by the generation near the boundary are not in equilibrium with the lattice. Hence the average energy shows a jump. From these observations, we might surmise that a continuum model could recover the behavior of large Knudsen number system except for the jump condition. However, the curvature of the distribution is not the same as that of a continuum model and the jump condition dominates over the energy distribution for larger Knudsen numbers. Figure 3 shows the same solution with a uniform generation that is scaled by the Knudsen number. Therefore each distribution should have approximately the same amount of energy in the system.

The non-continuum solution approaches that of the continuum solution for large Knudsen numbers, as expected, both in terms of the magnitude of the energy and in the reducing jump condition. We note that in equilibrium, the temperature and the average energy are equivalent, which is not so in non-equilibrium. Furthermore, for small Knudsen numbers, the jump condition obviates any attempt to match the non-continuum with the continuum solutions. The magnitude of the jump condition does not necessarily grow with Knudsen number. Realize that in Figure 3 the value of the generation is scaled by the Knudsen number, so the jump is actually a small percentage of the generation energy.

The scaled generation case is also interesting because the total amount of energy deposited into each case is the same despite the difference in device size. Therefore the heat flux at the boundary must be identical. The heat flux for each distribution is identical and matches the continuum solution. Figure 4 shows the components of the flux for various Knudsen numbers. In the case of large Knudsen numbers the generated phonons travel ballistically to the contacts. Therefore, the heat flux is derived from the generated phonons. For small Knudsen numbers, on the other hand, the generated phonons scatter, which results in an increase in the lattice energy. Emission from the lattice is the primary component in the heat flux for large de-
vices (small Knudsen numbers). The sum of the two components is linear from -0.5 at $xKn = 0$ to 0.5 at $xKn = 1$. This is precisely the continuum result and required to balance energy.

The pulse generation case models a situation more closely aligned with a hot-spot problem. In fact, the generation Knudsen number is infinite for this case. A Dirac delta pulse is located at the middle of the domain. Figure 5 shows non-continuum results for various device Knudsen numbers. In all cases, the distribution is similar to the non-continuum solution—linear between the boundary and the pulse. In fact, the continuum model matches the results well for all Knudsen numbers except for the jump condition. The distribution recovers the continuum solution for small Knudsen numbers as expected. Note that the non-continuum distribution produces a solution whose maximum energy is lower than that of the continuum solution. This feature can be understood in terms of the no-generation case. The pulse acts as a thermalizing boundary. Therefore, the solution between the real contact and the pulse looks as if it were between two contacts. The jump condition that appears at the contact prevails at the pulse location as well. This analysis suggests that peak energies of the non-continuum solution will always be less than the energy of the continuum solution.

For large Knudsen numbers the ballistic phonons being generated at the pulse do not contribute appreciably to the distribution except to increase the jump condition. By analogy a comparison between the no-generation ballistic case (Figure 1) and the ballistic pulse case is provided. Phonons are generated at the midpoint of the domain and travel ballistically toward the contacts where they are thermalized much like the phonons originating from the contact in the no-generation case. The background lattice has essentially zero energy, and the average is the equilibrium energy, which is nearly flat like the no-generation case.

As a final note, we conducted a simulation that involved four pulses equally spaced in a domain with $Kn = 50$. This particular configuration could mimic that of a microelectronic device with a series of transistors with highly localized generation. The device size is large enough to be considered continuum, but the generation Knudsen number is infinite. The solution to both the continuum and non-continuum models were achieved with superposition of results from the individual pulse solutions. Again we see the maximum energy predicted from the non-continuum solution is less than that of the continuum model.

The results do not explore the effects of dispersion and detailed scattering mechanisms on the local temperature. While these effects might alter the predicted magnitude of temperature signature, they are not expected to provide a significant difference in the comparison between the continuum and non-continuum solutions because the effective thermal properties in the continuum model will also adjust accordingly. An additional effect that is not considered is transient effects. Previous efforts have shown that unsteady operation can have dramatic non-local time
effects, but spatial variations and extreme temperatures were not predicted [17].

4 CONCLUSIONS

Non-continuum models with generation are similar to continuum models in functional form for all Knudsen numbers. The difference occurs in the jump condition at the boundary. For large Knudsen numbers, the jump condition is larger than the overall difference between the minimum and maximum temperatures. However, where non-continuum effects are prominent, the near-ballistic transport removes the generated phonons to the thermalized boundaries without increasing the average local energy. The results suggest that non-continuum models are warranted only for devices of small size and not necessarily for devices with large generation Knudsen numbers. In other words, non-local effects are the result of thermalizing boundaries and highly localized generation itself does not warrant non-continuum solutions. Furthermore, the maximum temperatures predicted by non-continuum models are less, but comparable, to those of the continuum model.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Tim Fisher at Purdue for helpful conversations.

References


List of Figures

1 Steady-state, one-dimensional solution without generation for various Knudsen numbers compared to the continuum solution. 13

2 Steady-state, one-dimensional solution with uniform generation over the domain for various Knudsen numbers. The continuum solution (equation 10) is also show for $L = 1/Kn = 10$ and $G = 1$. 13

3 Steady-state one-dimensional non-continuum solution with uniform generation. The amount of generation is scaled by the Knudsen number such that each film has the same amount of energy deposited. 14

4 (a) Heat flux due to the generated phonons (b) Heat flux due to lattice emission. The sum is the total heat flux, which is linear from -0.5 at $xKn = 0$ to 0.5 at $xKn = 1$ 14

5 Steady-state, one-dimensional solution with a generation spike in the middle of the domain for various Knudsen numbers. 15

6 Distribution for a large device ($Kn = 0.02$) with several localized generation regions. 15
Fig. 1. Steady-state, one-dimensional solution without generation for various Knudsen numbers compared to the continuum solution.

Fig. 2. Steady-state, one-dimensional solution with uniform generation over the domain for various Knudsen numbers. The continuum solution (equation 10) is also show for $L = 1/Kn = 10$ and $G = 1$. 
Fig. 3. Steady-state one-dimensional non-continuum solution with uniform generation. The amount of generation is scaled by the Knudsen number such that each film has the same amount of energy deposited.

Fig. 4. (a) Heat flux due to the generated phonons (b) Heat flux due to lattice emission. The sum is the total heat flux, which is linear from -0.5 at $xKn = 0$ to 0.5 at $xKn = 1$. 
Fig. 5. Steady-state, one-dimensional solution with a generation spike in the middle of the domain for various Knudsen numbers.

Fig. 6. Distribution for a large device (Kn = 0.02) with several localized generation regions.