INSTITUTE OF PHYSICS PUBLISHING SEMICONDUCTOR SCIENCE AND TECHNOLOGY
Semicond. Sci. Technol. **17** (2002) 579–584 PII: S0268-1242(02)32810-4

Nonequilibrium transport in scaled high electron mobility transistors

K Kalna and A Asenov

Device Modelling Group, Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow G12 8LT, UK

Received 17 January 2002, in final form 25 March 2002
Published 17 May 2002
Online at stacks.iop.org/SST/17/579

Abstract
Nonequilibrium transport and carrier scattering in aggressively scaled high electron mobility transistors (HEMTs) are investigated using a Monte Carlo device simulator. The devices are scaled down to sub-100 nm dimensions with respect to gate lengths of 120, 90, 70, 50 and 30 nm in order to improve their performance. Two scaling approaches assuming constant supply voltages are compared: full scaling and lateral-only scaling. These two scaling approaches are applied to two types of HEMTs: pseudomorphic HEMTs (PHEMTs) with an InGaAs channel and HEMTs with a standard GaAs channel. Carrier transport in the fully scaled PHEMTs is highly ballistic in the device channel but the ballistic limit is never reached. Instead, the ballistic nature of the carrier transport begins to be suppressed at the 50 nm gate length due to enhanced scattering in the channel and backscattering from the drain recess region as a result of the high electric field in the shorter devices. The transport in the lateral-only scaled PHEMTs and in the standard HEMTs scaled within both the approaches is negatively affected by deconfinement and real space transfer.

1. Introduction

The direct current (dc) and radio frequency (rf) performance of high electron mobility transistors (HEMTs) can be further increased by scaling these devices to decanano dimensions. To achieve this goal device scaling to sub-100 nm gate lengths has to be done proportionally in both lateral and vertical directions in order to retain gate control over the carriers in the channel [1, 2]. When HEMTs optimized for the 100 nm gate length are scaled only in the lateral direction device performance deteriorates [3–6]. Because in a constant voltage scaling the reduction in device dimensions results in an increase of the electric field, the carrier transport in the channel beneath the gate is highly nonequilibrium and to a great extent ballistic [7, 8]. A high degree of ballisticity is desired in semiconductor devices to increase the operational speed [9, 10]. Nevertheless, the increasing fraction of carriers travelling ballistically will hit a physical limit in the high-field regions. Indeed, in principle, with the reduction of channel length, the probability for a carrier to cross the channel ballistically without suffering a collision should increase [11]. At the same time, however, with the increase in the electric field, a carrier can gain higher energy which increases its scattering probability.

We investigate the impact of HEMT scaling on nonequilibrium transport and the fraction of carriers travelling ballistically through the channel. Two approaches to the device scaling are considered, which are full scaling and lateral-only scaling in two types of HEMTs: pseudomorphic HEMTs (PHEMTs) with a low indium content channel and HEMTs with a GaAs channel. Only the fully scaled devices exhibit a continuous improvement in performance [1, 2]. However, we argue that the degree of ballistic transport of carriers through the channel of the fully scaled PHEMTs increases with gate length reduction down to 50 nm channel length and then saturates due to high scattering rates associated with the high electric field in the channel.

The carrier transport along the HEMT channel is investigated in section 2 for both scaling approaches. This investigation is based on monitoring the average carrier velocity in the vicinity of the gate. To better understand the character of carrier transport a field–momentum relaxation time is introduced in section 3. Finally, section 4 draws up conclusions of our investigations.

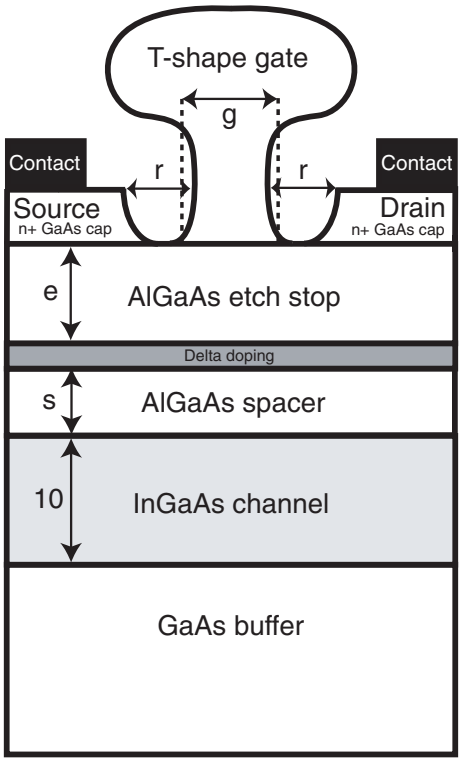

Figure 1. The cross-section scheme of HEMTs considered in this work. The channel is made of $In_{0.2}Ga_{0.8}As$ in a PHEMT or GaAs in a standard HEMT.

2. Transport in scaled HEMTs

We investigate the electron transport in a set of aggressively scaled HEMTs using a Monte Carlo (MC) device simulation. The study is based on an enhanced electron MC transport model embedded into our finite element simulator H2F [1]. This transport model includes the electron scattering with polar optical phonons, inter- and intra-valley optical phonons, non-polar optical phonons and acoustic phonons and ionized and neutral impurity scattering. In addition, alloy scattering and strain effects [12] are considered in the InGaAs channel. All scattering rates are calculated with a form factor F (the overlap integral) as

$$F(E, E') = \frac{(1+\alpha E)(1+\alpha' E') + 1/3\alpha E \alpha' E'}{(1+2\alpha E)(1+2\alpha' E')} \qquad (1)$$

proposed in [13] which allows us to extend the validity of the transport model for electric field up to 400 kV cm^{-1} [1]. In equation (1), an electron with an initial energy E scatters into the state with a final energy E'. α and α' are non-parabolicity parameters for the electron in the initial and final valleys, respectively.

A thoughtful calibration of the simulator has been performed against a real 120 nm gate length PHEMT [1]. The calibration requires the inclusion of additional, external to the MC simulation, contact resistances of the drain and source into the intrinsic *I–V* characteristics obtained directly from MC simulations using the procedure described in [14]. The generic cross section of the simulated PHEMT is illustrated in figure 1. The device has a T-shaped gate, a 30 nm Si-doped n+ GaAs cap layer with a doping concentration of 4×10^{18} cm^{-3}, an $Al_{0.3}$ $Ga_{0.7}As$ gate barrier layer, a Si-delta doped layer with a doping concentration of 7×10^{12} cm^{-2}, an $Al_{0.3}$ $Ga_{0.7}As$ spacer layer and a 10 nm channel. The whole structure is grown on top of a 50 nm GaAs buffer. The PHEMTs have an $In_xGa_{1-x}As$ channel with indium content $x = 0.2$ while the HEMTs have a GaAs channel which merges with the GaAs buffer.

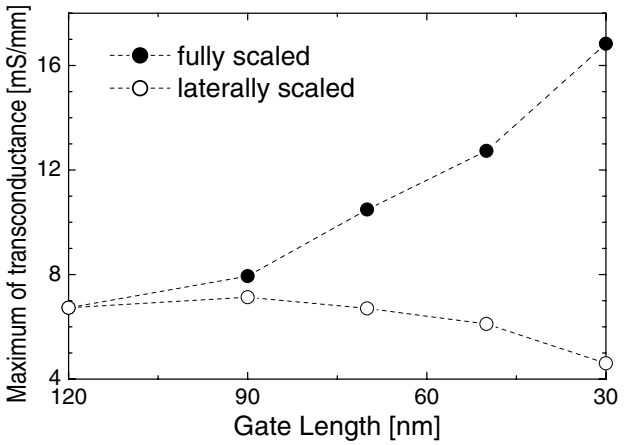

Figure 2. Maxima of transconductance as a function of the device gate length for fully scaled PHEMTs (full circles) compared with transconductance maxima of laterally scaled PHEMTs (open circles).

The constant voltage scaling of HEMTs follows two possible scenarios. The first approach is a full device scaling when both the lateral and vertical device dimensions are scaled down with respect to gate lengths of 90, 70, 50 and 30 nm. The second approach is lateral-only scaling which assumes that only the lateral dimensions of the devices are scaled while the vertical dimensions remain unchanged. Note that the devices are scaled in the lateral direction according to the respective gate lengths while in the vertical direction the scaling takes into account technological constraints [1]. Figure 2 illustrates that only the approach which fully scales the PHEMT delivers a continuous increase in transconductance [1]. The transconductance of the lateral-only scaled PHEMTs deteriorates after a marginal improvement at the 90 nm gate length.

The detailed study of the nonequilibrium transport in the scaled HEMTs requires monitoring of various aspects of the carrier dynamics in the device. The carrier dynamics are always investigated assuming intrinsic device conditions. The inclusion of an external resistance does not qualitatively change the carrier dynamics, the only additional effect being reduction of the drive current [1]. The basic physical quantity which is an indicator for the transport behaviour of carriers is their average velocity. We record the velocity of each carrier along the device channel during the time of an MC simulation (for a period of 10–20 ps after the initial transient) from which we may determine the average carrier velocity. Results of this study for the PHEMTs are presented in figures 3 and 4 where the average carrier velocity through the channel beneath the gate is compared for the two scaling scenarios considered in this paper. The shape of the velocity profile along the channel

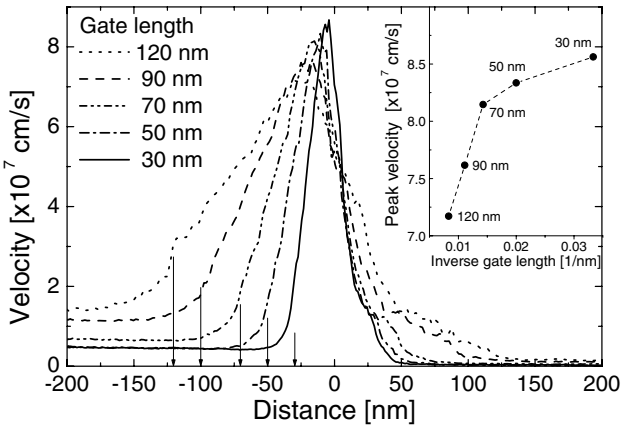

Figure 3. Average carrier velocity along the $In_{0.2}Ga_{0.8}As$ channel of the fully scaled PHEMTs. The end of the gate is always set at zero distance and the beginning of the gate is depicted by arrows for 120, 90, 70, 50, and 30 nm gate length devices, respectively. The inset shows the peak average velocity as a function of the inverse gate length.

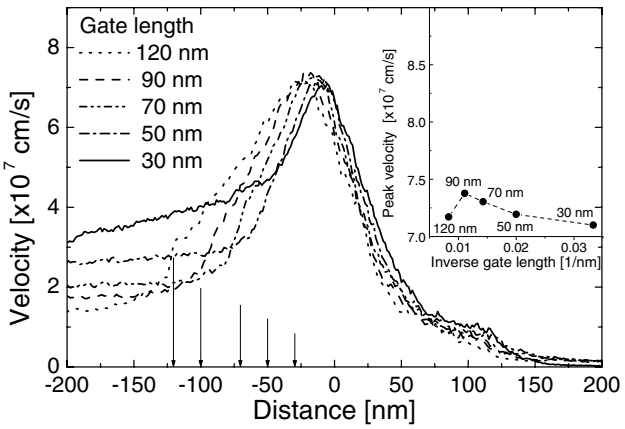

Figure 4. Average carrier velocity along the $In_{0.2}Ga_{0.8}As$ channel of the laterally scaled PHEMTs. The inset again shows the peak average velocity as a function of the inverse gate length.

shows velocities which are much higher than the saturation velocity thus indicating that the carrier transport is highly nonequilibrium. With the PHEMT scaling the peak of the velocity increases and at the same time becomes narrower. If we compare the average velocities at the beginning of the gate (at the source side of the device) indicated by arrows in figure 3 for the set of fully scaled PHEMTs, we can see that the velocity at this point decreases with scaling. This is opposite to the behaviour observed in scaled MOSFETs where the average velocity increases at the beginning of the gate consistent with the increase of peak velocity at the drain end [15]. The behaviour of the average carrier velocity in HEMTs is related to the geometry of the device which has recess regions on both sides of the gate and is distinctly different from MOSFET geometry. The presence of the recess region and the lateral component of the gate-fringing electric field [16] results in acceleration of the carriers entering the region. This effect is more pronounced in those devices with a large gate length and proportionally large recess regions.

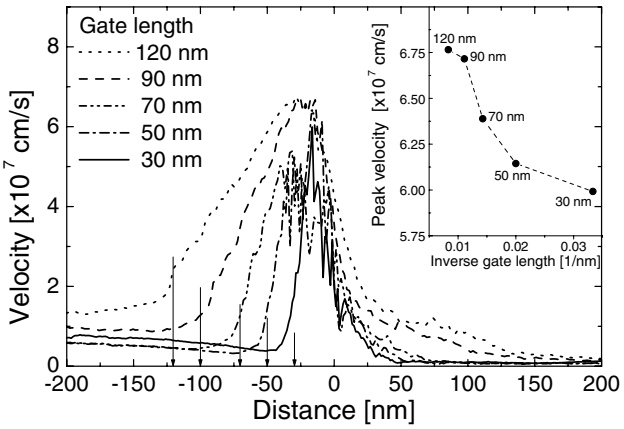

Figure 5. Average carrier velocity along the GaAs channel of the fully scaled HEMTs. The inset picks out the maximum velocity versus inverse gate length.

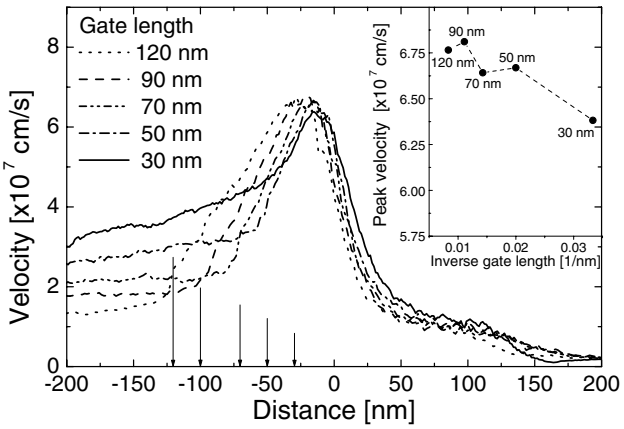

Figure 6. Average carrier velocity along the GaAs channel of the laterally scaled HEMTs. The inset indicates peak velocity versus inverse gate length.

The peak carrier velocity in all investigated devices is much higher than the bulk saturation velocity. A sharp drop in the velocity is observed when carriers reach the extremely high field recess region at the drain end of the gate. The maximum velocity in figure 3 increases sharply when the PHEMT is fully scaled from 120 nm to 70 nm and starts to saturate with the further scaling of devices to gate lengths of 50 nm and 30 nm. When the PHEMT is scaled only in lateral dimensions, the average carrier velocity in figure 4 increases only for the 90 nm gate length device. For gate lengths less than 90 nm the average velocity steadily decreases which indicates a loss of gate control over the device channel that softens the lateral field distribution. During the lateral-only scaling the area controlled by the gate becomes smaller while the relative distance between the gate and the channel remains the same. The electric field cannot effectively accelerate and keep the carriers in the channel and ballisticity declines. Consequently, the performance of the laterally scaled PHEMTs with gate lengths less than 120 nm deteriorates [1].

To highlight in full the benefit of the InGaAs channel in the PHEMT structures we have also simulated scaled HEMTs with GaAs channels [17, 18]. The average carrier velocity along the channel is plotted in figures 5 and 6 for the fully and

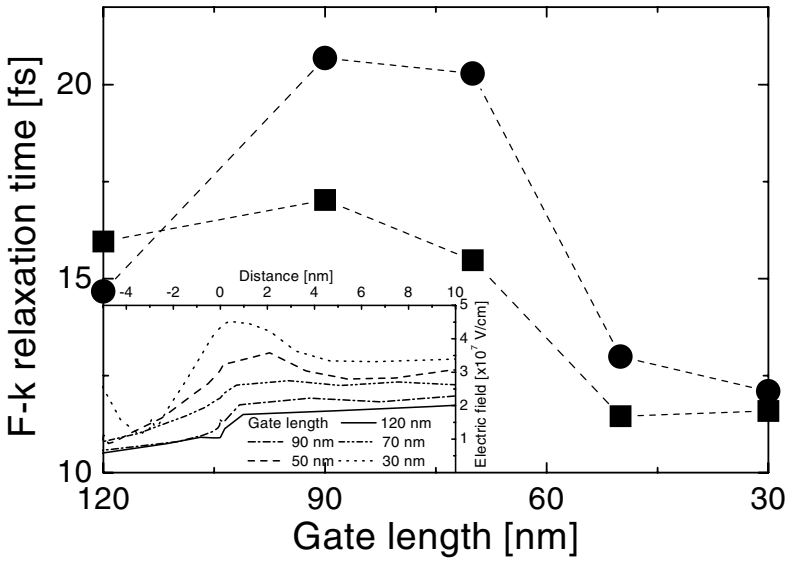

Figure 7. Field–momentum relaxation time as defined by equation (2) as a function of the gate length for the fully scaled devices. Circles denote PHEMTs and squares denote standard GaAs HEMTs. The inset shows the electric fields along the channel beneath the gate in a set of fully scaled PHEMTs. The zero distance is always set at the end of the gate.

laterally scaled HEMTs respectively. It is clear that the GaAs channel devices are not amenable to scaling and cannot offer improvement in device performance mainly on account of poor confinement. This is so even in the fully scaled HEMTs with a GaAs channel where the average velocity drops sharply when this device is scaled from a gate length of 120 nm to 30 nm. The situation is much worse for the laterally scaled HEMTs with the GaAs channel. Here, the average carrier velocity in figure 6 only slightly increases at 90 nm gate length and then starts decreasing.

3. Field–momentum scattering rate

The gate length of the decanano PHEMTs becomes comparable to the inelastic mean-free path of the carriers. Hence, electrons travelling through the gate region should have a high probability of passing through this region ballistically. To study the ballistic transport in the scaled devices we monitor particles in the gate-controlled channel region and then calculate a field–momentum (F–k) relaxation time as a reciprocal of $\Gamma_{\rm fm}$

$$\Gamma_{\rm fm} = \frac{e}{\hbar}\frac{|\boldsymbol{F}|}{|\boldsymbol{k}|}, \qquad (2)$$

where $\boldsymbol{F}$ is the electric field vector at the particle position and $\boldsymbol{k}$ is the particle wave vector. This relaxation time represents the time during which the absolute particle momentum is relaxed due to the effect of the electric field at the particle position. The F–k relaxation rate, $\Gamma_{\rm fm}$, is averaged over the number of particles passing through the selected region to give the mean F–k relaxation rate.

The inverse of $\Gamma_{\rm fm}$, which is the mean F–k relaxation time, can be compared for devices with a different layer structure but with the same geometry in order to assess typical features of carrier transport. When the F–k relaxation time increases a large number of carriers can travel ballistically suffering less scattering on average. On the other hand, a decrease of this relaxation time clearly indicates that carriers undergo many scattering events in a selected region, e.g. the device channel. Using the F–k relaxation time as one of the device characteristic parameters may help in assessing the degradation of the ballistic transport due to the enhanced carrier scattering which is expected as a result of the constant voltage scaling approach.

We investigate the F–k relaxation time in the gate-channel region of HEMTs. Although the vertical extent of the gate-channel region is simple to define geometrically, the lateral position is determined solely by the electrostatic potential along the channel. We employ the data from figures 3–6 and define the starting point to be the position along the channel at which velocity has increased by 5% above the mean velocity starting from the buffered source region. Similarly, the end point is defined as the position where the velocity is 5% higher than the mean velocity at the drain end.

Results of this investigation are presented in figures 7 and 8. The F–k relaxation time of the laterally scaled devices in figures 8 is always larger than the F–k relaxation time of the fully scaled ones because the electric field in the channel of the fully scaled devices abruptly increases in the scaling process (see the inset of figure 7) reaching much higher values compared to the lateral-only scaled devices (see the inset of figure 8).

Firstly, we focus on the F–k relaxation time of the fully scaled devices shown in figure 7. The values of $\Gamma_{\rm fm}$ for HEMTs with GaAs channel are always lower than those of PHEMTs which signals a slower carrier transport in GaAs-HEMTs. The faster transport in the InGaAs channel can be understood bearing in mind the smaller electron effective mass in the $In_{0.2}Ga_{0.8}As$ channel and the larger Γ–L valley separation as well as the better confinement in the structure with the AlGaAs/InGaAs interface. The mean F–k relaxation

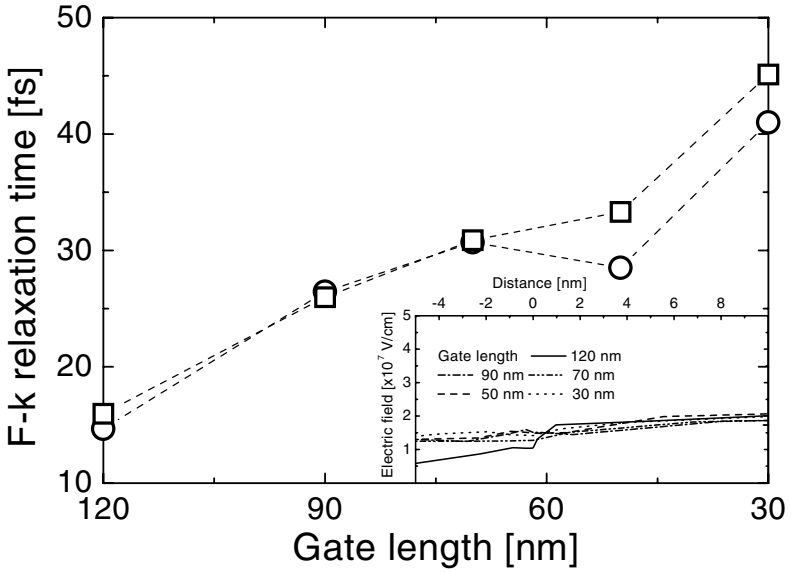

Figure 8. Field–momentum relaxation time as a function of the gate length for lateral-only scaled devices. Open circles show results for PHEMTs and open squares represent standard GaAs-HEMTs. The inset compares the electric fields along the channel beneath the gate in the lateral-only scaled PHEMTs. The end of the gate is again at zero distance.

time of the fully scaled devices increases with gate scaling only down to the 90 nm gate length in accordance with the behaviour of the average velocities in the channel and then drops down at 50 nm and saturates at the 30 nm device. This can be explained as follows: a huge fringing electric field [16] surrounds the region beneath the gate around the gate corners. The impact of this gate-fringing field on particles increases with reduction of the gate length and recess region in the scaling process. The high gate-fringing field in the drain recess region sharply increases the energy of the electrons which increases the probability for a Γ–L valley transition in devices with gate lengths less than 90 nm. This transition increases the effective electron mass and reduces the average velocity of the carriers and also backscatters slow particles in the channel region of the devices [19]. As a result, the mean *F–k* relaxation time starts dropping rapidly and then saturates as the gate length is scaled from 50 nm to 30 nm. The saturation of the *F–k* relaxation time occurs when the field particle acceleration and energy losses due to the increased scattering become balanced. Scattering reduces the fraction of carriers travelling ballistically through the device and adversely affects device performance. The main mechanisms which dominate scattering in the PHEMTs are the electron scattering with polar optical phonons and the alloy scattering. Below the 70 nm gate length PHEMT alloy scattering is replaced by intervalley scattering from the Γ to the L valley and from the L to the X valley, respectively. The alloy scattering is always a main scattering mechanism in the X valley.

The behaviour of the *F–k* relaxation time for the lateral-only scaled devices in figure 8 is somehow different. The *F–k* relaxation time consistently increases with increasing mean value of the electron wave vector in the channel due to less scattering. Nevertheless, although the ballisticity of the transport improves during the lateral scaling, device performance declines because the gate has already lost control over the carriers in the channel (the electric fields in PHEMTs shown in the inset of figure 8 are almost identical). We can see that the *F–k* relaxation time of the laterally scaled PHEMTs is smaller at 50 nm and 70 nm gate lengths than the relaxation time of GaAs-HEMTs. This demonstrates more scattering in the $In_{0.2}Ga_{0.8}As$ channel due to the presence of the alloy and intervalley scatterings, which can take place at lower electron energies than in GaAs because of a smaller bandgap.

4. Conclusions

We have studied nonequilibrium transport in PHEMTs and GaAs-HEMTs scaled to sub-100 nm dimensions using MC device simulations. Two scaling approaches have been considered. The fully scaled structures are scaled in both the lateral and vertical dimensions with respect to gate lengths of 120, 90, 70, 50 and 30 nm. The laterally scaled structures are scaled only in the lateral dimension while their gate-to-channel distance always remains the same. Only the fully scaled PHEMTs show an improvement in the device performance as a result of scaling. This is associated in part with the increased average electron velocity through the channel due to the increasing electric field. Furthermore, the *F–k* relaxation time indicates that the electron transport in the fully scaled PHEMTs is more ballistic in devices with channel length down to 50 nm and then the degree of ballisticity starts declining. This limitation of ballisticity in the 50 nm and 30 nm PHEMTs is induced by strong scattering at high electric fields and backscattering [19] which comes mainly from the interaction with polar optical phonons and intervalley phonons. The performance of the laterally scaled PHEMTs deteriorates due to the loss of gate control over the channel. The electron transport in the GaAs-HEMTs is slower due to two factors: the

larger electron effective masses which slow down transport in the device channel and the poorer electron confinement at the AlGaAs/GaAs interface which also leads to an enhanced real space energy transfer from the channel.

Acknowledgments

This work has been supported by EPSRC under grant no GR/M93383. KK would like to thank R C W Wilkins for a critical reading of the manuscript.

References

[1] Kalna K, Roy S, Asenov A, Elgaid K and Thayne I 2002 *Solid-State Electron*. **46** 631
[2] Kalna K, Roy S, Asenov A, Elgaid K and Thayne I 2000 *Proc. ESSDERC 2000* ed W A Lane *et al* (Cork: Frontier Group) p 156
Kalna K, Asenov A, Elgaid K and Thayne I 2001 *VLSI Design* **13** 435
[3] Zhou J R and Ferry D K 1992 *IEEE Trans. Electron Devices* **39** 473
[4] Dollfus P, Bru C and Hesto P 1992 *J. Appl. Phys*. **73** 804
[5] Kizilyalli I C, Artaki M, Shah N J and Chandra A 1993 *IEEE Trans. Electron Devices* **40** 235
[6] Fay P, Stevens K, Elliot J and Pan N 2000 *IEEE Electron Device Lett*. **21** 141
[7] Zhou J R and Ferry D K 1992 *Semicond. Sci. Technol*. **7** B456
[8] Han J 1999 *J. Korean Phys. Soc*. **34** S322
[9] Taur Y and Ning T H 1998 *Fundamentals of Modern VLSI Devices* (Cambridge: Cambridge University Press) p 283
[10] Timp G *et al* 1999 *IEDM Technical Digest* (New York: IEEE) p 55
[11] Datta S, Assad F and Lundstrom M 1998 *Superlattices Microstruct*. **23** 771
[12] Köpf C, Kosina H and Selberherr S 1997 *Solid-State Electron*. **41** 1139
[13] Matz D 1968 *Phys. Rev.* **168** 843
[14] Babiker S, Asenov A, Cameron N and Beaumont S P 1996 *IEEE Trans. Electron Devices* **43** 2032
[15] Bude J D 2000 *Proc. SISPAD'00 (Seattle, USA)* ed J Faricelli and P Leon (New York: IEEE) p 23
[16] Han J and Ferry D K 1999 *Solid-State Electron*. **43** 335
[17] Delagebeaudeuf D and Linh N T 1982 *IEEE Trans. Electron Devices* **29** 955
[18] Morkoç H and Ünlü H 1987 *Semiconductors and Semimetals* vol 24 ed R Dingle (London: Academic) p 135
[19] Lundstrom M, Ren Z and Datta S 2000 *Proc. SISPAD'00 (Seattle, USA)* ed J Faricelli and P Leon (New York: IEEE) p 1