Aust. J. Phys., 1993, **46**, 447–64

Charge Fluctuations in High-Electron-Mobility Transistors: A Review*

F. Green

Division of Radiophysics, CSIRO,
P.O. Box 76, Epping, N.S.W. 2121, Australia.

Abstract

The physics of high-electron-mobility transistors (HEMTs) plays a central role in contemporary design for millimetre-wave communications. HEMTs are the early fruits in a harvest of increasingly radical devices whose structural features are measured in nanometres. The operating principles of these devices are richly varied, and almost always far from classical. One of the tasks for device physics is to understand fluctuation phenomena, or noise: the control of charge fluctuations is basic to high performance, yet the description of these processes remains incomplete if not obscure. This paper reviews some aspects of charge-transport noise that affect HEMT operation.

1. Introduction

In the last twenty years, revolutionary methods of materials fabrication at the atomic scale have been developed, with ramifications which are still not easy to envisage (Capasso 1990*a,b*; Corcoran 1990; Special Section, *Science* 1991). The prototype of these techniques, and one of the best-established for the GaAs family of semiconductors (the so-called III–V compounds), is molecular-beam epitaxy, or MBE (Foxon 1988; Wu *et al.* 1985). MBE makes possible the growth—atomic layer by atomic layer—of very pure, very uniform, precisely doped semiconductor composites, with exquisite control of the underlying band structure. At least in the world of electronics, epitaxial technologies such as MBE and its products represent the first manufacturing systems to rely on the design and forming of materials at a truly atomic scale.

The desirable qualities of bulk GaAs as a substrate for high-frequency circuit applications have long been known (Brodsky 1990; Frensley 1987; Goronkin *et al.* 1985). The GaAs band structure in particular, with its small electronic effective mass and its direct gap, overcomes the restrictions of earlier semiconductor technologies over a huge frequency band, stretching from direct-current to the visible. While the difficult processing chemistry of the III—V family militates against their large-scale adoption as *bulk* semiconductors, that disadvantage is now far outweighed by the unique electronic properties embodied in epitaxially grown structures.

* Paper presented at the Gordon Godfrey Workshop, University of New South Wales, Sydney, 20-21 July 1992.

0004-9506/93/030447$05.00

The high-electron-mobility transistor, or HEMT, is already a bread-and-butter product of current MBE materials technology. (Two other acronyms, 'MODFET' and 'TEGFET', are also encountered in the literature.) HEMTs have inherently fast response times and low noise, which makes them ideal for millimetre-wave integrated circuit requirements (Archer 1992; Morkoç 1991; Pearton and Shah 1990). An obvious qualitative difference between the HEMT and its predecessor, the 'classical' bulk field-effect transistor, is the *quantum confinement* of majority carriers to a thin sheet within the semiconductor structure. The quasi-two-dimensional carrier population, free to move within a near-perfect crystalline matrix, is the key to remarkable improvements in signal gain, current density and quiet operation.

Current-fluctuation effects are central to all of these properties. Some of these are easily understood within linear-response theory, but other fluctuation phenomena are less tractable. In particular, nonequilibrium noise poses significant theoretical challenges, both descriptive and predictive. This paper examines a few of the basic physical issues which motivate device-noise theory. In Section 2 the structure and operation of a HEMT are first reviewed. Simple examples of fluctuation-dominated behaviour are then discussed, with numerical illustrations. Next, Section 3 outlines the recent nonlinear fluctuation theory of Stanton and Wilkins (1987*a,b*). Their idealised but exact approach helps to identify at least some of the complicated noise physics which can arise when carriers in GaAs-like conduction bands are subjected to high fields. Finally, Section 4 summarises the importance of understanding fluctuations in HEMTs as a prelude to understanding the next generation of devices.

2. Static Fluctuations: Charge Transfer

High-mobility transistors are fabricated on single-crystal GaAs wafers by epitaxy, lithography, and selective etching. Fig. 1 shows the basic anatomy of a HEMT. For applications to millimetre-wave integrated circuits, the HEMT's planar layout is (with one exception) of relatively generous dimensions, namely fractions of a millimetre; the really delicate *nanometre*-scale structure in these devices is found in the epitaxially grown layers just below the surface. The single exceptionally narrow surface feature is the current-control electrode, or gate, whose length in the direction of current flow must be less than $0 \cdot 2\,\mu$m so that carrier transit times are short enough to cope with signal frequencies up to 100 GHz (Pearton and Shah 1990). We will turn to the issues of charge transport in the following Section.

The first goal of HEMT design is to ensure a healthy population of mobile carriers within the quiescent device. The most important property of the carriers is that they are *strongly confined* by quantum effects to a thin ($\sim$15 nm) sheet buried within the epitaxial structure. Normally they are free to move only along this thin layer and not across it, like air bubbles trapped under an ice sheet. In effect, the carriers exist in a two-dimensional world. As in the familiar pn junction, the physical reason for charge segregation in a HEMT is the need to balance the Fermi level throughout a structure which is spatially nonuniform; the process is shown in Fig. 2. In heterojunction devices, the spatial inhomogeneity

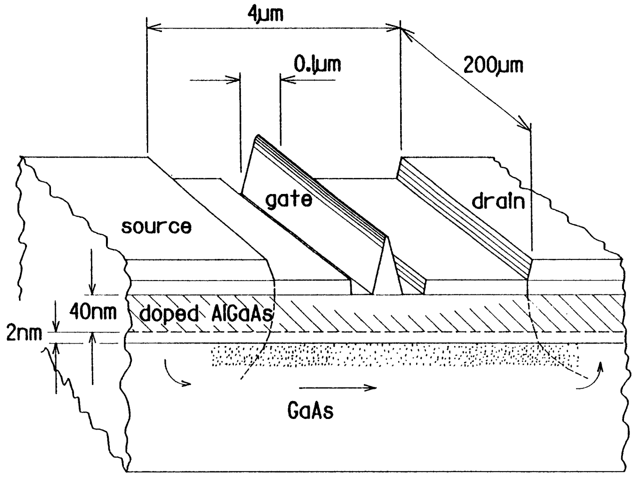

Fig. 1. Anatomy of a HEMT. Metallic current-injection (source) and current-collection (drain) pads are separated by an etched or recessed region, with a narrow control electrode (gate) to suppress or enhance conductivity via an applied biasing field. In the transverse direction, a high-purity GaAs substrate is overlaid by an epitaxially grown AlGaAs layer doped with Si. The abrupt AlGaAs/GaAs band-gap discontinuity promotes transfer and quantum confinement of donor electrons within a thin layer in the underlying GaAs. The trapped carriers remain highly mobile along the plane. Except for the gate length, planar dimensions are at the micrometre scale, transverse dimensions 100 times smaller.

is evident not just in the doping profile, but also in the local band structure of the host lattice. It is MBE which makes possible the large and sudden structural nonuniformities needed for a successful HEMT design.

In practical high-mobility transistors, the majority carriers are electrons. Optimal configurations are those for which the largest possible number of electrons transfers from the Si donors on the high-band AlGaAs side to the undoped, low-band GaAs side of the heterojunction (Fig. 2*b*). To determine the equilibrium Fermi level and the fraction of electrons transferring from the AlGaAs into the conductive sheet on the GaAs side, one solves Poisson's equation for the local electrostatic potential induced in the neighbourhood of the junction. Poisson's equation is coupled self-consistently to Schrödinger's equation for the electrons, trapped by their own potential at the heterojunction itself.

The coupled equations may be solved at various levels of sophistication. Having done this, one knows (i) the total electron density confined within the active channel in the GaAs, (ii) the strength of the self-confinement, and (iii) (by linear response analysis) the sensitivity to changes in the control voltage applied to the gate.

The sensitivity, defined as the differential response to the gate voltage, is governed by the static *density fluctuations* in the active channel beneath the gate. This is not entirely surprising; the fluctuation–dissipation theorem (Martin 1968)

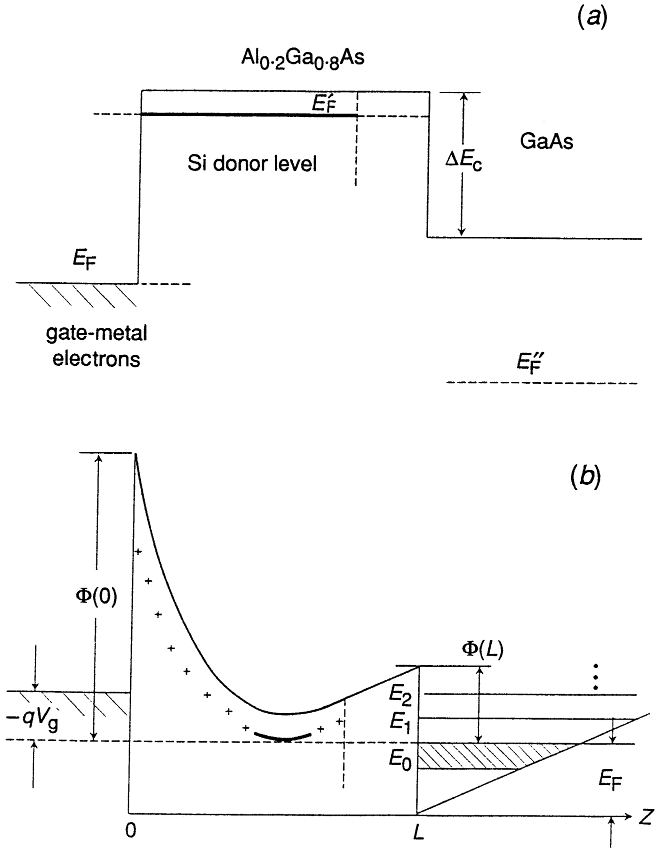

Fig. 2. Typical heterojunction band structure. (*a*) Conceptual alignment of conduction bands prior to formation of the confined electron population. Fermi level must become uniform throughout the inhomogeneous structure. (*b*) Actual band alignment; Fermi level uniform. Donor-bound electrons transfer either sacrificially to the gate-metal conduction band, or usefully to the quantum well at the AlGaAs/GaAs interface. The undoped AlGaAs layer next to the well boosts mobility by distancing carriers from ionised Si scatterers. The gate potential $\Phi(0)$, partly screened by intervening ionised donors, provides charge control. A signal applied at the gate displaces the local Fermi level and induces re-equilibration of the heterojunction.

asserts that there is a one-to-one correspondence between density fluctuations at equilibrium and the system's response to a small external disturbance.

Assume that the self-consistent problem has been solved for a given bias voltage V_g applied at the gate terminal. This means that we know the following parameters: the Fermi level E_F; the local electrostatic potential $\Phi(z)$ as a function of depth z in the Si-doped AlGaAs layer; and finally the area density $n_s(E_F)$ of electrons trapped in the thin conductive region. The situation is shown in Fig. 2*b*.

At the gate-metal/AlGaAs boundary ($z = 0$) we have the relation

$$\Phi(0) = \Phi_{\mathrm{S}} - qV_{\mathrm{g}}\,, \tag{1}$$

where Φ_{S} is the Schottky barrier energy (the work function for crossing the gate-metal/AlGaAs interface) and $-qV_{\mathrm{g}}$ is the electronic potential due to the external bias. At the AlGaAs/GaAs interface ($z = L$) we have the boundary condition

$$\Phi(L) + E_{\mathrm{F}} = \Delta E_{\mathrm{c}}\,, \tag{2}$$

where ΔE_{c} is the difference between AlGaAs and GaAs conduction-band levels. This is one of the parameters fixed by the choice of alloy composition, and is typically $\sim 0{\cdot}25$ eV.

From equations (1) and (2) we can determine the change in E_{F} with respect to a small change in the bias voltage:

$$\frac{\delta E_{\mathrm{F}}}{\delta V_{\mathrm{g}}} = q\frac{\delta\Phi(L)}{\delta\Phi(0)}\,. \tag{3}$$

In principle the energy levels $\{E_j(n_{\mathrm{s}})\}_{j=0}^{\infty}$ in the quantum well (bound states and free) are known. Therefore the thermodynamics of the channel population are also known. The dependence of the energy levels on a thermodynamic variable, namely the density n_{s} itself, follows from the self-generated nature of the trapping potential.

The next task is to express the electron density as a function of the Fermi energy and to obtain its derivative. For a planar device with active area A, the transferred carrier population forms an open Fermi system with partition function $Z(A, T, E_{\mathrm{F}})$ (Pathria 1972). A little algebra allows us to write

$$\begin{aligned} n_{\mathrm{s}} &\equiv \frac{k_{\mathrm{B}}T}{A}\frac{\partial \ln Z}{\partial E_{\mathrm{F}}} \\ &= D^* \sum_{j=0}^{\infty} k_{\mathrm{B}}T \ln\{1 + \exp([E_{\mathrm{F}} - E_j(n_{\mathrm{s}})]/k_{\mathrm{B}}T)\}\,. \end{aligned} \tag{4}$$

The parameter $D^* = m^*/\pi\hbar^2$ is the density of states for a two-dimensional conduction band with effective mass m^*. For each level $E_j(n_{\mathrm{s}})$ quantised in the well *perpendicular* to the interface, there exists a band of freely conducting states *along* the interface plane. Parameter D^* counts these states of planar motion, per unit 2D bandwidth per unit sample area. The jth term in the sum gives the bandwidth actually occupied at the jth level.

We now vary both sides of equation (4):

$$\delta n_{\mathrm{s}} = D^* \sum_{j=0}^{\infty} P_j\,\delta E_{\mathrm{F}} - D^* \sum_{j=0}^{\infty} P_j \frac{\mathrm{d}E_j}{\mathrm{d}n_{\mathrm{s}}}\delta n_{\mathrm{s}}\,, \tag{5}$$

with $P_j = \{1 + \exp[E_j(n_s) - E_F]/k_B T\}^{-1}$ being the relative probability that an electron belongs to the jth band in the quantum well. Two competing effects are thinly concealed within equation (5): thermal fluctuations in the electron population, and (as mentioned above) the self-consistency of electron trapping at the heterojunction. The latter effect is seen in the appearance of δn_s in the second term on the right-hand side of the equation.

To display the fluctuation structure of equation (5), we need only associate the first term on its right-hand side with the standard definition (Pathria 1972) of the mean-squared fluctuation $\langle (\Delta N)^2 \rangle \equiv \langle (N - \langle N \rangle)^2 \rangle$ in the total particle number $\langle N \rangle = n_s A$, held within the active area under the gate:

$$\begin{aligned}\frac{\langle \Delta N)^2 \rangle}{A} &\equiv \frac{k_B T}{A} \frac{\partial \langle N \rangle}{\partial E_F} \\ &= k_B T \frac{\partial n_s}{\partial E_F} \\ &= k_B T D^* \sum_{j=0}^{\infty} P_j \,. \end{aligned} \tag{6}$$

Combining equations (5) and (6), we have

$$k_B T \frac{\delta n_s}{\delta E_F} = \frac{\langle (\Delta N)^2 \rangle / A}{1 + D^* \sum_{j=0}^{\infty} (\mathrm{d}E_j / \mathrm{d}n_s) P_j} \,. \tag{7}$$

Lastly, we connect this relation to the measurable sensitivity of the channel current to the gate-control voltage. In the limit of a small driving field $\mathcal{E}$ between drain and source electrodes, the drain-to-source current I_{ds} which crosses the gate region scales in a simple way with the carrier density n_s, the device width W, and the mobility μ:

$$I_{ds} = q n_s W \mu \mathcal{E} \,, \tag{8}$$

so that the transconductance (sensitivity) g_m is

$$g_m \equiv \frac{\delta I_{ds}}{\delta V_g} = q W \mu \mathcal{E} \frac{\delta n_s}{\delta V_g} \tag{9}$$

or, in a more convenient logarithmic form,

$$\frac{g_m}{I_{ds}} = \frac{q}{n_s} \left(\frac{\delta n_s}{\delta E_F} \frac{\delta \Phi(L)}{\delta \Phi(0)} \right) . \tag{10}$$

In this last equation, the two calculable factors which determine the transconductance have been teased out for emphasis. The first factor describes the effects of the carrier fluctuations in the GaAs channel, while the second factor describes the *Debye screening* of the gate potential, due to the partially ionised Si donors within the AlGaAs layer. Screening attenuates the applied potential $\Phi(0)$, so that the potential actually felt by the 2D carriers at the heterojunction, $\Phi(L)$, is reduced in strength.

The form of equation (7) makes clear how density fluctuations in the electron population directly affect the response of a HEMT. This must be so—quite generally—for any field-effect device, be it novel or not so novel (recall the fluctuation-dissipation theorem). However, in heterojunction structures, the self-consistent quantisation of the conducting states causes the denominator of equation (7) to exert a strong renormalising effect on the observable fluctuations, which otherwise would behave classically. The point is illustrated by comparing the calculated behaviour of various conceptual HEMT structures. For present purposes, the Poisson problem for the AlGaAs doping region (Fig. 2*b*) has been solved within a local Thomas–Fermi scheme, while the self-consistent potential for the quantum well at the AlGaAs/GaAs heterojunction has been approximated to lowest order by a ramp of constant slope. This simple approach suffices to illustrate the basic physics.

Figs 3, 4 and 5 show three instances of static response behaviour in which density fluctuations and quantisation of the carrier states act in combination to determine the measurable outcome. Each set of figures presents the response and related quantities as functions of the total voltage at the gate pad $[-V_{bias} = \Phi(0)/q]$. Part *a* of each figure shows the transconductance g_{m} and its two governing factors, $\delta n_{\mathrm{s}}/\delta E_{\mathrm{F}}$ and $\delta\Phi(L)/\delta\Phi(0)$, while part *b* shows details of the density fluctuations.

In the first example, Fig. 3, quantisation of the level structure for the confined electrons has been suppressed artificially: $\{E_j(n_{\mathrm{s}})\}_{j=0}^{\infty}$ is replaced with its continuum analogue. We look first at Fig. 3*a*. Its most notable feature is the pronounced maximum in the transconductance. This contrasts with g_{m} profiles for conventional field-effect transistors having homogeneous bulk band structures, where both the carrier densities and their fluctuations have a classical form which decays monotonically as the gate voltage is made more negative. In heterojunctions, the peaked form of g_{m} arises because the carrier population in the GaAs, though initially large ($\sim 10^{12}\ \mathrm{cm}^{-2}$), is screened from the gate voltage by the partially ionised Si donors in the intervening AlGaAs layer. As more electrons are stripped from the donors, the screening length for the charge distribution in the AlGaAs layer becomes comparable to the AlGaAs thickness. The carrier population in the GaAs is no longer isolated from the gate potential, and its Fermi level begins to drop rapidly.

It can be seen, then, that the first distinguishing property of response in HEMTs is the characteristic transconductance peak. This depends in the first place on the physics of charge depletion in the donor layer, not explicitly on quantisation. With current MBE technology, it is not too hard to obtain transconductances on the order of $1000\ \mathrm{mS\,mm}^{-1}$, exceeding by several factors the best results for bulk GaAs devices.

The solid line in Fig. 3*b* shows the bare fluctuation $\langle(\Delta N)^2\rangle/\langle N\rangle$ [equation (6)], normalised to the mean carrier population. The dot–dashed line shows the channel density n_{s} in units of $D^* k_{\mathrm{B}} T$ [see equation (4)]. It can be shown that a classical population confined by a triangular potential has $n_{\mathrm{s}} \propto \exp(E_{\mathrm{F}}/2k_{\mathrm{B}}T)$ after integrating over the spectral 'continuum'; there is no explicit quantisation correction as seen in equation (7). Consequently, for our (artificial) classical

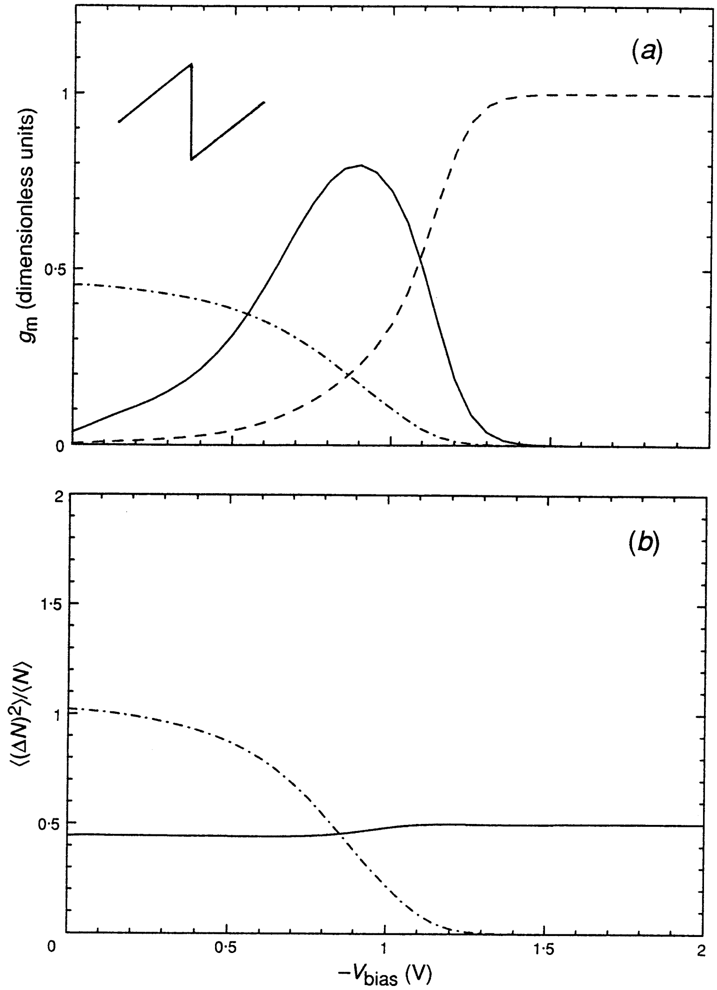

Fig. 3. Static HEMT response as a function of gate voltage, carrier quantisation suppressed. Inset shows shape of AlGaAs/GaAs well. (*a*) Solid line: dimensionless transconductance $D^{*-1}(\delta n_s/\delta\Phi(0))$ (see Section 2), magnified twentyfold. Dot–dashed line: dimensionless fluctuation factor $D^{*-1}(\delta n_s/\delta E_F)$. Dashed line: Debye attenuation factor $\delta\Phi(L)/\delta\Phi(0)$. Note the distinctive transconductance peak when Debye screening no longer shields the carrier population from the gate potential. (*b*) Channel density and its fluctuations in dimensionless units. Solid line displays fractional fluctuation. Dot–dashed line: channel density n_s in units of $D^* k_B T$. Beyond the 'pinchoff' voltage, 1·3 V, channel is depleted and fractional fluctuation goes classically to 1/2. Details of epitaxial structure: $Al_{0\cdot2}Ga_{0\cdot8}As$/GaAs with 20 nm undoped AlGaAs below the gate metal, 20 nm AlGaAs Si-doped at 10^{18} cm^{-3}, 2 nm undoped AlGaAs, then GaAs. Temperature is 300 K.

system, $\langle(\Delta N)^2\rangle/\langle N\rangle = \frac{1}{2}$. (The slight deviation from this value at small V_{bias} is due to Fermi–Dirac statistics.) This behaviour establishes a benchmark for the more realistic examples which we now discuss.

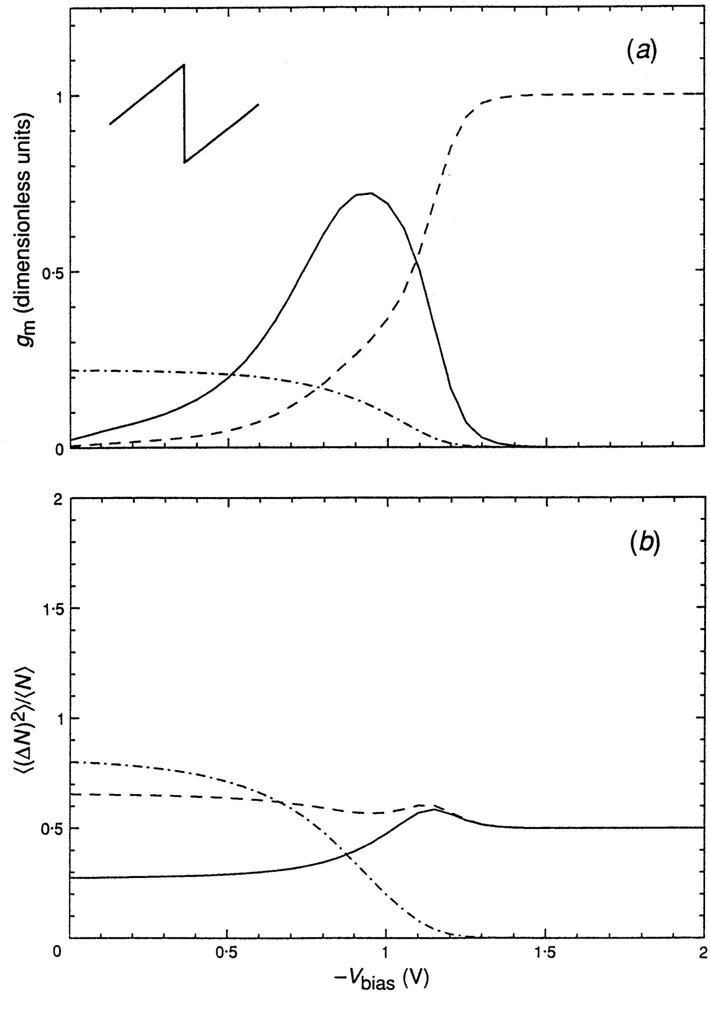

Fig. 4. Static HEMT response for AlGaAs/GaAs heterojunction, quantisation included. Inset shows shape of AlGaAs/GaAs well. Epitaxial parameters and graphical legends as for Fig. 3. (*a*) Quantisation depresses fluctuation factor, but Debye factor is enhanced. Overall transconductance is not greatly reduced. (*b*) Solid line shows self-consistently renormalised fractional fluctuations [see equation (7)]; dashed line shows bare fluctuations. At pinchoff, fractional fluctuation goes over to classical limit as in Fig. 3*b*.

Fig. 4*a* shows g_m and its two factors for a standard AlGaAs/GaAs HEMT structure, with bound-state quantisation now properly included. We note immediately that the fluctuation factor (dot–dashed line) is strongly suppressed in comparison with Fig. 3*a*. The transconductance, however, is not too much degraded because the Debye screening factor $\delta\Phi(L)/\delta\Phi(0)$ (dashed line) increases faster than before. Fig. 4*b* is more interesting; here the bare fluctuation term is shown by the dashed line, while the solid line now shows its dressed counterpart [equation (7)]. Recalling that the difference between these forms exists only when there are discrete bound-state levels in the spectrum $\{E_j(n_s)\}_{j=0}^{\infty}$, the conceptual and quantitative importance of the latter is clear. As the gate potential becomes strong enough to deplete the active GaAs channel, the bound states again merge with the continuum. Beyond $1 \cdot 3$ V the density becomes asymptotically classical, and we regain the high-potential behaviour seen in Fig. 3*a* (at least within the present simplified treatment).

The power of epitaxial technology springs from the variety and flexibility in control of band-structure parameters. Our last example, Fig. 5, shows just one option for enhancing the transconductance, namely tailoring the density fluctuations by setting up a different quantum well structure. In the present case a 15 nm layer of InGaAs alloy is interposed between the underlying GaAs substrate and the AlGaAs layer. The AlGaAs specification is unchanged. InGaAs has a *smaller* band gap than GaAs. An AlGaAs/InGaAs/GaAs composite thus has a wider conduction-band discontinuity, and quantisation of the carrier states becomes even stronger than for the AlGaAs/GaAs configuration in Fig. 4. The result can be seen in Fig. 5*a*, where the maximum in g_m is not only higher but sustained over a wider voltage range. Comparison with Fig. 4*a* shows that this is due in part to a ~50% enhancement of the fluctuation parameter $\delta n_s/\delta E_F$. In Fig. 5*b* we see that the saturated carrier population of the 'pseudomorphic' InGaAs well is twice that for GaAs. Looking at the relative density fluctuations—bare and dressed—in the same figure, we see that initially they behave similarly in InGaAs as they do in GaAs. This is because the energy-level spectrum is dominated by its density dependence when the density is high; the band jump at the InGaAs/GaAs boundary is small compared with the self-consistent potential. The fluctuations in each well act quite differently, however, as the bias voltage depletes the channel. The self-consistent potential quickly dies and the InGaAs well reverts to its original rectangular shape, so that its bound-state spectrum does not collapse with the density. Because the density dependence is rapidly lost, the asymptotic limit for the fluctuations in InGaAs is $\langle (\Delta N)^2 \rangle / \langle N \rangle \sim 1$ rather than $\frac{1}{2}$ as in the GaAs well, where the wedge-shaped potential depends more persistently on the density. The practical importance of the pseudomorphic approach is that not only does the overall carrier density surpass by a healthy factor that in a standard GaAs heterojunction, but the response is also enhanced by stronger fluctuations throughout the whole density range. As a bonus the electron mobility is improved as well, thanks to the lower effective electron mass in InGaAs.

Another way of enhancing sensitivity is to grow donor-layer structures with very strong spatial modulation of the doping profile; this is a technique for tailoring the screening function $\delta\Phi(L)/\delta\Phi(0)$, the other determining factor for g_m. Systematic evaluation of this design path, in combination with InGaAs wells, is central to current research but outside the scope of this review. Nevertheless, the examples presented above encapsulate basic physical principles of MBE-based device design. We now look briefly at the more dramatic, and problematic, dynamical properties.

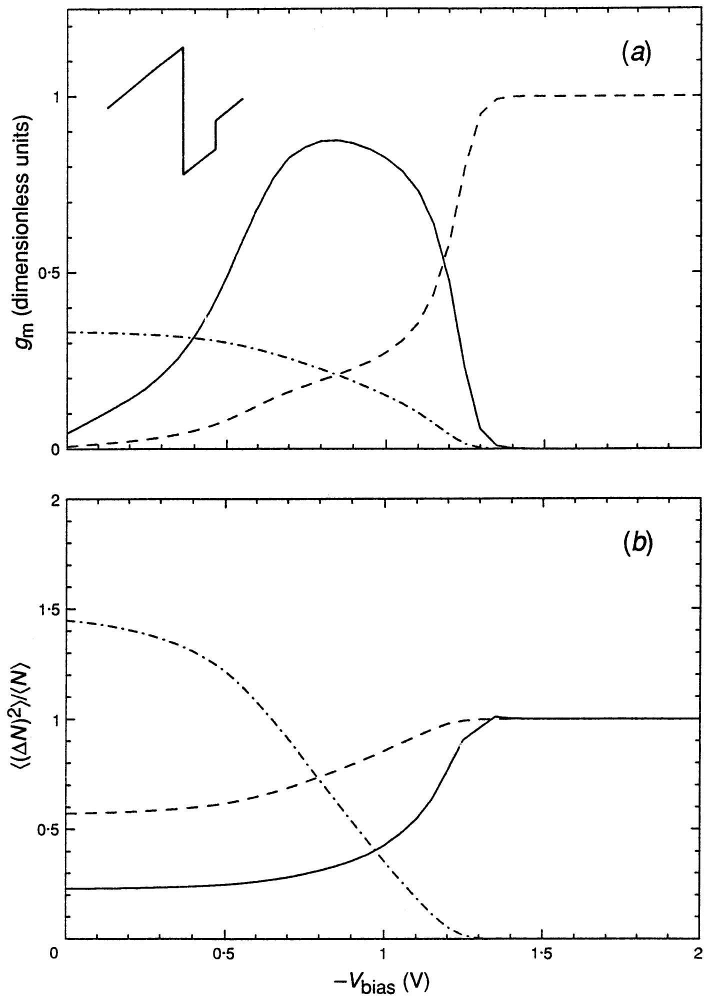

Fig. 5. Static HEMT response for pseudomorphic heterojunction. Inset shows shape of AlGaAs/GaAs/InGaAs well. AlGaAs structure as for Fig. 4. Quantum well is 15 nm $In_{0\cdot15}Ga_{0\cdot85}As$. (*a*) Transconductance peak is higher and wider as stronger confinement boosts channel density and fluctuations. (*b*) In the pinchoff limit fractional fluctuations go to 1, not $\frac{1}{2}$, because InGaAs bound-state spectrum does not collapse with density.

3. Dynamic Fluctuations: High-Field Transport

HEMTs are three-terminal devices. In the previous Section we relegated the drain-to-source driving field $\mathcal{E}$ to a small influence, in order to simplify the discussion of the gate-control transconductance. In working devices, potentials of a few volts are typically applied from drain to source, eliciting high current densities (up to 1 A per mm of channel width), and inducing voltage gradients as high as 100 kV cm^{-1} in the region underlying the narrow gate. Highly localised gradients of this magnitude are inaccessible to direct measurement; their existence must be inferred from elaborate device simulations. But the very problem of setting up a reliable transport model consists in capturing the nature and consequences of these strong spatial and temporal transients. The fact that one is dealing with strongly coupled electron and phonon populations far from equilibrium is one reason why 'hot-electron' physics requires powerful tools from the theoretical repertoire (Frensley 1990; Jacoboni and Lugli 1989; Reggiani 1985), and why it can be theoretically exciting. Another reason is that heterojunction devices behave classically in the plane and quantally in the transverse direction (although this neat decoupling is blurred at high fields); thus HEMTs provide one platform for examining phenomena at the border between macro- and mesoscopic physics.

A vast established body of work exists in the field of nonequilibrium device physics (for a succinct technical review see Stern 1987). While much crucial design information comes from careful Monte Carlo modelling (see e.g. Jacoboni and Lugli 1989), it remains necessary to find a relatively uncluttered understanding of what goes on in these devices. This is especially true of fluctuations and noise at high fields. Therefore, as always, there is an important role for idealised but well-controlled models of high-field processes. We now examine one such approach (Stanton and Wilkins 1987*a,b*).

The recent work of Stanton and Wilkins looks at nonequilibrium current noise by generating rigorous solutions for the time-dependent correlation functions of the Boltzmann equation for III–V conduction bands. They extend the usual relaxation-time phenomenology for collisional effects. Within this phenomenology they follow the exact consequences for current noise at arbitrarily large fields $\mathcal{E}$. It is a semiclassical, dynamic, nonequilibrium single-particle theory of *velocity fluctuations.* Their work is complementary to the quantum, static, equilibrium many-particle description of density fluctuations reviewed in Section 2. The merits of the Stanton–Wilkins study are that it is clear, tractable, and still general enough to simulate the major processes in hot-electron transport. Three effects are reviewed here: field-induced electron heating, nonparabolic conduction bands, and inter-valley scattering. All examples are taken from the two references by Stanton and Wilkins.

Fig. 6 shows the noise-power characteristic and velocity-field curve for bulk GaAs as functions of uniform electric field in a constant-relaxation-time approximation—the simplest assumption. The velocity-field relation corresponds to $v = \mu\mathcal{E}$, with the relaxation time τ_0 related to the mobility μ by $\tau_0 = m^*\mu/q$. At larger field values we note from Fig. 6 how the noise power rapidly exceeds its quiescent equilibrium value (defined by the Nyquist formula). The threshold field for this rise is $\sim$5 kV cm^{-1} in typical GaAs material.

One can describe this noise behaviour as an increase in effective temperature, recalling that noise is a measure of temperature in the familiar equilibrium picture. The accelerated electrons pick up energy from the driving field faster than relaxation processes can dissipate the excess to the surroundings. The carriers' velocity distribution function becomes skewed towards higher values and broadens. Since the broadening is essentially a measure of velocity fluctuations in the hot-electron population, this leads to an enhanced noise characteristic.

Within this model of nonequilibrium transport, electrons in a HEMT channel would display a rising temperature profile as they traversed the high-field region

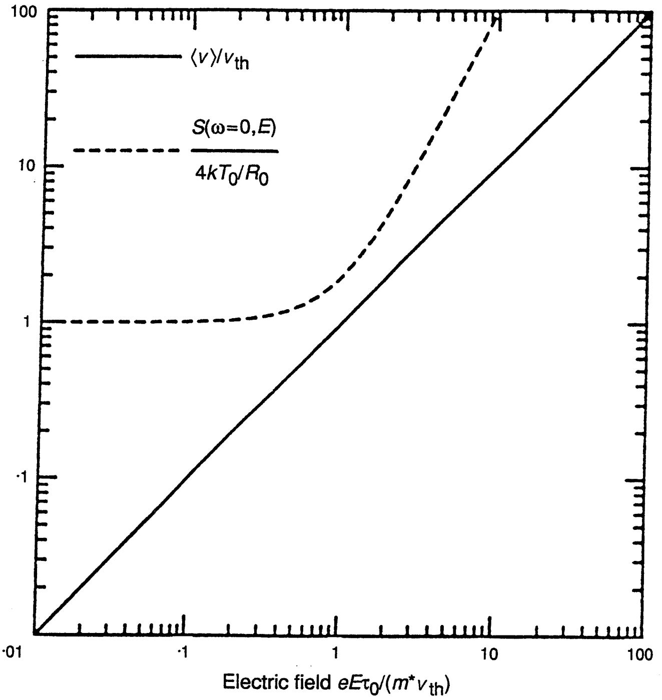

Fig. 6. Nonequilibrium electron heating: electron drift velocity and zero-frequency noise power in constant-relaxation-time approximation, taken from Stanton and Wilkins (1987*a*). Velocity is in units of classical thermal velocity. Noise power is in units of the equilibrium Nyquist value.

just under the gate. One might then expect more, not less, current noise when the device was driven hard. The fact that HEMTs are so remarkably quiet in operation shows that things are not as simple as that. The second example, Fig. 7, displays the effects of a *nonparabolic band* on velocity fluctuations. Away from the centre of the Brillouin zone, GaAs-like conduction bands depart from their quadratic energy dispersion to a form better approximated by (Stanton and Wilkins 1987*a*)

$$\epsilon(k) = \frac{1}{2\alpha}[(1 + 2\alpha\hbar^2k^2/m^*)^{\frac{1}{2}} - 1]\,. \tag{11}$$

The nonparabolicity parameter α is in the range $0{\cdot}5$–$1{\cdot}0$ eV^{-1}, and causes the energy dispersion to be asymptotically linear for large electron wavevectors k. Fig. 7*a* shows that the electron drift velocity in such a band saturates in the high-field limit. With a constant relaxation time as before, the noise power (Fig. 7*b*) now behaves very differently, with strong suppression at large

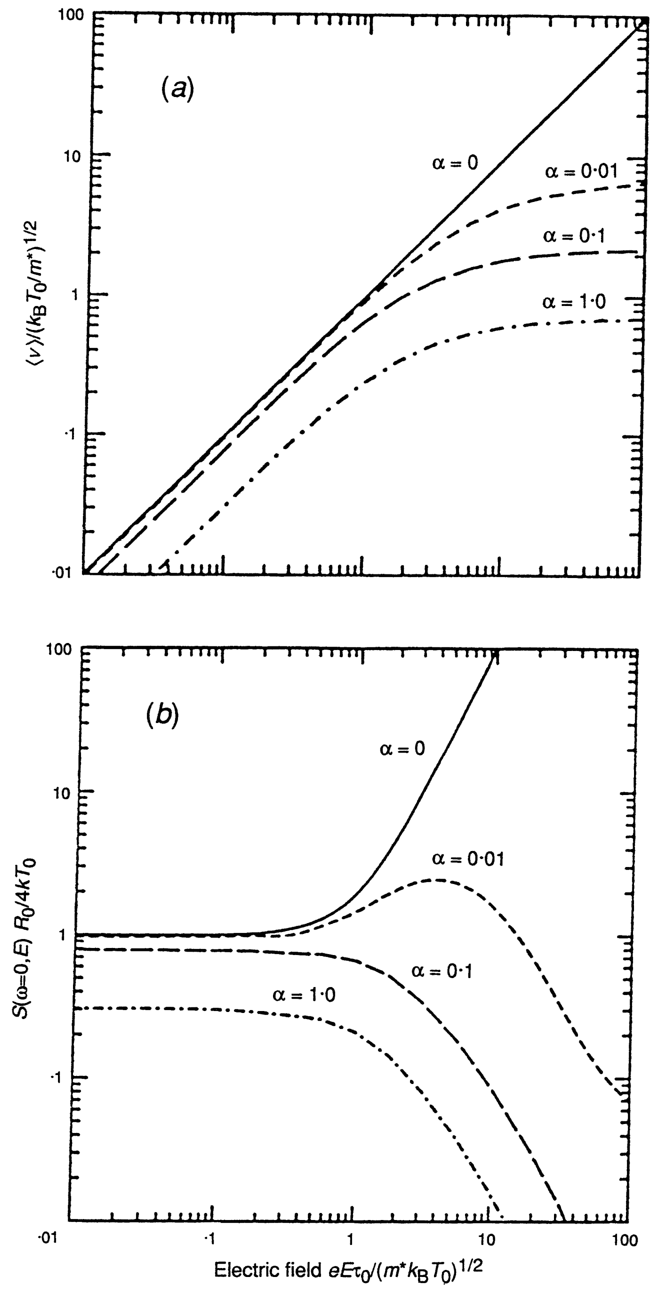

Fig. 7. Nonparabolic-band effects on drift and noise, from Stanton and Wilkins (1987*a*); where α is the nonparabolicity parameter [see equation (1)]. (*a*) Drift velocity saturates at high fields because conduction-band group velocity is constant at large electron momenta. (*b*) Velocity fluctuations are strongly suppressed since group velocity is the same for all high-momentum states in the band.

fields. The reason for this surprising behaviour is that the linear dependence of the nonparabolic band energy $\epsilon(k)$ at large k leads to a weak k-dependence of the group velocity. It follows that deviations from the average drift velocity are necessarily small in the large-momentum region, where the group velocity is virtually a constant and where—under the influence of a large field—one finds the largest contribution to ensemble averages.

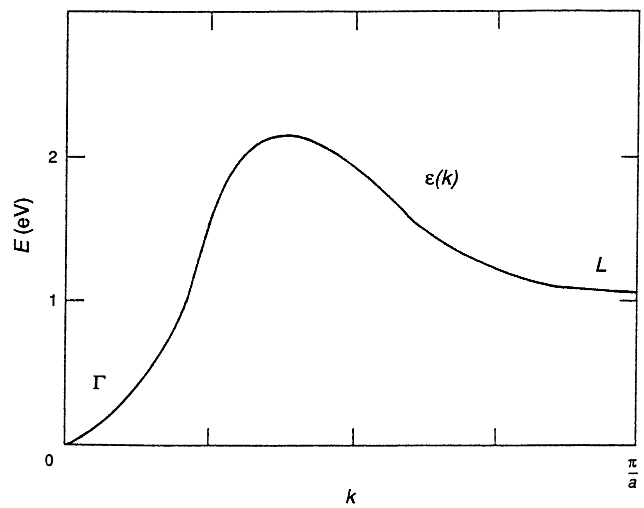

Fig. 8. Schematic plot of energy dispersion in the GaAs conduction band. At small momenta, electrons in the central Γ valley have small effective mass. Near the zone boundary, electrons in side valleys (e.g. L) are ~1 eV above Γ and have a much larger effective mass. The Γ-to-L transitions are dominant at high fields, leading to velocity overshoot.

We have seen two counteracting effects on nonequilibrium carrier noise in an idealised system driven by large potential gradients. The final example from Stanton and Wilkins addresses the important phenomenon of *inter-valley scattering.* Fig. 8 shows the general shape of a GaAs-like conduction band. We see that a large enough transfer of energy and momentum—easily reached in the active region of a HEMT—will take an electron from the central Γ valley to a peripheral L valley (with a nonparabolic domain in between). The L-valley effective mass is almost an order of magnitude greater than the Γ-valley mass, so that an electron thus displaced in the Brillouin zone suddenly acquires a much smaller group velocity. This accounts for the characteristic overshoot in the velocity-field curve, Fig. 9*a*: beyond the threshold field at which a Γ-to-L transition can take place, the carrier drift velocity falls dramatically. Most electrons crossing the active region of an operating HEMT are driven into this state.

Once promoted to the L valley, electrons have a short lifetime (~10^{-14} s) and shed their excess energy by optical-phonon emission or otherwise. The overall relaxation time is sensitive to the details of inter-valley scattering channels, selection rules, reduced dimensionality (as in a HEMT), etc. The most interesting point here (Stanton and Wilkins 1987*b*) is that the velocity-field relation, Fig. 9*a*,

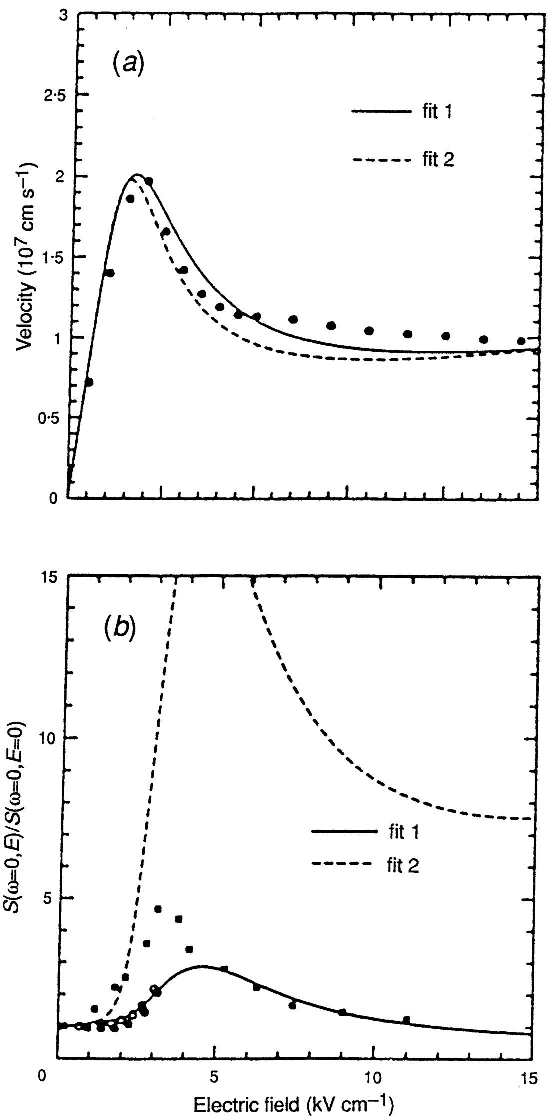

Fig. 9. Effects of inter-valley scattering, from Stanton and Wilkins (1987*b*). (*a*) Drift velocity shows overshoot when driving field imparts enough energy to promote an electron to a side valley with much reduced mobility. Inter-valley scattering rate does not much affect the v–$\mathcal{E}$ relation. (*b*) Noise power is very sensitive to scattering rate. Dots are experimental data; see above reference for details.

is not at all sensitive to changes in the scattering rate, while by contrast the noise-versus-field relation, Fig. 9*b*, is extremely sensitive to the scattering rate.

For the microscopic description of high-field noise performance, the evidence of Fig. 9 carries a wider and more powerful message: it constrains any attempt to link drift velocity (a steady-state quantity) to noise (a fluctuation) by a too-simple phenomenology, guided perhaps by linear-response theorems which do not apply at large fields. The Stanton and Wilkins result dramatically restates the principle that there is no valid way to go from knowledge of an *average* property alone, to knowledge of *random deviations* from that average. Even in linear-response theory one must first calculate the fluctuations and then look for a simpler connection with the steady-state average. It is sometimes tempting to turn this principle on its head, given all the additional demands of practical HEMT-based circuit design; but convenient rules of thumb may totally misrepresent important physical processes.

4. Summary: Issues in Device Modelling

In Section 2 we saw that the process of understanding and optimising the static response of heterojunction devices is intimately connected with density fluctuations. This is an area in which systematic theoretical improvements are reasonably straightforward; it is, however, only the lowest in a progression of theoretical plateaux. In Section 3 we visited the next level, that of dynamical fluctuations for high-field transport, where the degree of complexity is high (even in a semiclassical context) and cannot be ignored safely without misrepresenting important effects.

A thorny and recurrent issue in this branch of device physics is the valid description of populations in open quantum systems (e.g. confined carriers in a HEMT) when they couple strongly to their environment. In particular, it remains difficult in practice to marry the one-dimensional quantum-well description of a HEMT to two-dimensional current flow in the plane. In actual operating conditions, local inhomogeneities develop in the Fermi-level, velocity, and field profiles within the active channel. These non-uniformities are on a length scale of tens of nanometres, comparable to the dimensions of the epitaxial structure, so that it becomes impossible to resolve the total picture neatly into a confinement part and a transport part. Even in heavily computational (e.g. Monte Carlo) approaches, there are conceptual difficulties in specifying boundary conditions, good quantum numbers, and so on.

Beyond the problems of mesoscale systems in high fields lie the novel questions posed by yet smaller-scale systems of a few elementary units (Capasso 1990*b*), in configurations too delicate to be treated in the usual thermodynamic limit. Instances are quantum wires, quantum dots and molecular entities such as C_{60}. These are certainly open systems but more strongly quantised than the electrons in a HEMT; extended *quantum-coherence effects* provide a new physical setting and a new class of operating principles. The next generation of nano-fabricated devices can still benefit from HEMT theory to the extent that it provides exposure to the technicalities of irreversibility, strong coupling to heat baths, intense external gradients, very short transients (important for high-frequency phenomena), and also in some degree to the fuzzy intersection between classical and quantum regimes.

Finally, a tenable view of epitaxially grown heterojunction technology is that much is already known and much has been attempted towards practical solutions to design problems for low-dimensional devices, but that much remains to be done if we are ever to gather a large and valuable menagerie of special-purpose tools into a consistent taxonomy.

References

Archer, J. W. (1992). *In* '1992 Asia–Pacific Microwave Conference Proceedings', pp. 271–2 (IEEE: Piscataway).
Brodsky, M. H. (1990). *Sci. Am.* **262**(2), 56.
Capasso, F. (1990*a*). *Phys. Today* **43**(2), 22.
Capasso, F. (1990*b*). *In* 'Physics of Quantum Electron Devices' (Ed. F. Capasso), pp. 1 ff. (Springer: Berlin).
Corcoran, E. (1990). *Sci. Am.* **263**(5), 74.
Foxon, C. T. (1988). *In* 'Interfaces, Quantum Wells, and Superlattices' (Eds C. R. Leavens and R. Taylor), pp. 11 ff. (Plenum: New York).
Frensley, W. R. (1987). *Sci. Am.* **257**(2), 68.
Frensley, W. R. (1990). *Rev. Mod. Phys.* **62**, 745.
Goronkin, H., Grondin, R. O., and Ferry, D. K. (1985). *In* 'Gallium Arsenide Technology' (Ed. D. K. Ferry), pp. 155 ff. (Howard Sams: Indianapolis).
Jacoboni, C., and Lugli, P. (1989). 'The Monte Carlo Method for Semiconductor Device Simulation' (Springer: Vienna).
Martin, P. C. (1968). *In* 'Many-Body Physics' (Les Houches Summer School Proceedings 1967, Eds C. DeWitt and R. Balian), pp. 58 ff. (Gordon and Breach: New York).
Morkoç, H. (1991). *Circuits and Devices* **7**(6), 15.
Pathria, R. K. (1972). 'Statistical Mechanics', pp. 98 ff. (Pergamon: Oxford).
Pearton, S. J., and Shah, N. J. (1990). *In* 'High-Speed Semiconductor Devices' (Ed. S. M. Sze), pp. 283 ff. (Wiley: New York).
Reggiani, L. (1985). *In* 'Hot-Electron Transport in Semiconductors' (Ed. L. Reggiani), pp. 7 ff. (Springer: Berlin).
Special Section, 'Engineering a Small World' (1991). *Science* **254**, 1300.
Stanton, C. J., and Wilkins, J. W. (1987*a*). *Phys. Rev.* B **35**, 9722.
Stanton, C. J., and Wilkins, J. W. (1987*b*). *Phys. Rev.* B **36**, 1686.
Stern, F. (1987). *In* 'Physics and Applications of Quantum Wells and Superlattices' (Eds E. E. Mendez and K. von Klitzing), pp. 133 ff. (Plenum: New York).
Wu, C. T., Hendel, R. H., and Dingle, R. (1985). *In* 'Gallium Arsenide Technology' (Ed. D. K. Ferry), pp. 107 ff. (Howard Sams: Indianapolis).

Manuscript received 14 December 1992, accepted 24 March 1993