HIGH-SPEED SEMICONDUCTOR DEVICES
IN COMPUTER TECHNOLOGY

Juras POŽELA and Vida JUCIENĖ

ICSC - World Laboratory, Lithuanian Branch
Semiconductor Physics Institute
2600 Vilnius, A.Goštauto St.11, Lithuania

Abstract. The main approaches to improve the computer performance by
enhancing the operating speed of transistors, used in modern computer tech-
nology, are reviewed. The qualitative enhance of the transistor operating speed
could be realized by using semiconductor materials with the higher than in silicon
electron mobility (such as GeAs, InGaAs) and by decreasing the size of transis-
tors until nanometric dimensions. That opens the way to design computers with
the operating speed larger than 10 billion operations per second.

Key words: computer technology, integrated circuits, logic gates, semi-
conductor devices and materials, high-speed transistors.

1. Introduction. High-speed semiconductor transistors are
one of the basic elements of modern computer technology. They are
used in many specific combinations in all main computer compo-
nents and their parameters determine the computer performance.
Progress in microelectronic semiconductor technology is the back-
ground to develop computer technology.

This paper is a review of the present-day physical and techno-
logical approaches to improve the parameters of transistors which
determine the computer performance. The main parameters of
transistors used in computers are: the operating speed, the power
consumption for unit logic switch, the size or scale of integration
of transistors in integrated circuits (IC).

The transistor operating speed is characterized by the thresh-
old frequency for unity current gain \( f_T \), by the maximum frequency
of oscillation (unity power gain) \( f_{\text{max}} \) and by the switching delay
time \( \tau_D \).
The power consumption $P_D$ for switching from one logic state to another is connected with the scale of integration of individual transistors. The power consumption in IC is equal to $W_{IC} = P_D N$, where $N$ is the number of simultaneously working transistors. The weaker $P_D$ the greater is allowable degree of scale of integration in IC. The joint parameter which determines the quality of a transistor in IC is the power-delay product $P_D \tau_D$, i.e., a switching energy.

The minimal size of a transistor could be characterized by the length $\lambda_G$ of minimal electrode (gate or emitter). The smaller size of a discrete transistor means the shorter distance between them in ICs and the shorter delay time for the transmission of the logic signal from one element to another. In the contemporaneous ICs that time is less than 2 ps.

![Graph showing the reduction of the dimension $\lambda_G$ and switching energy per gate $P_D \tau_D$ in FETs.](image)

**Fig. 1.** Reduction of the dimension $\lambda_G$ and switching energy per gate $P_D \tau_D$ in FETs.

Due to the progress in microelectronic technology the size and the power consumption of the transistors, used in computer ICs, rapidly decrease every year (Fig. 1). That gives the possibility to
create the complete computer on a single monolithic semiconductor crystal (e.g., transducer). Transducers are the basis of the progressive way in computer technology: computers with parallel computation. In this paper we shall see that advances in creation of semiconductor devices based on the high-speed transistors open the qualitatively new prospects in development of computer technology.

In the paper we shall describe, firstly, the progress in design and technology of usual bipolar (BT) and field-effect (FET) transistors, and then we shall analyze the new approaches in transistor physics and technology: hot electron and resonant tunneling transistors. In conclusion the maximum achieved operating speed parameters for different types of transistors are summarized:

2. Bipolar transistors. Schematic diagram of a bipolar transistor (BT) is presented in Fig. 2. The current and power gain of a BT is determined by an efficiency of minority-carrier transport from the emitter to the collector. The efficiency of the transport is characterized by the transfer coefficient

$$\alpha = \frac{\partial I_C}{\partial I_E} |_{V_{CB}}$$

(1)

where \(I_C = I_{nC} + I_{pC}\) is the collector current and \(I_E = I_{nE} + I_{pE}\) is the emitter current (the subscripts \(n\) and \(p\) refer to electrons and holes, respectively), \(V_{CB}\) is the collector-base voltage. The current gain of a BT in the circuit with a common emitter is equal to

$$\beta = \frac{\alpha}{1 - \alpha}.$$  

(2)

When the transport efficiency \(\alpha \approx 1\), the coefficient \(\beta\) reaches high values. The BT has the largest current gain compared with other types of transistors.

The operating speed of a BT is determined by the transit time \(\tau_{EC}\) of minority carriers from the emitter to the collector, as well as by the time of charge accumulation on outer and inner capacitances \(\tau_{RC}\). The unity current gain cutoff frequency

$$f_T \approx \frac{1}{2\pi\tau_T},$$  

(3)

where \(\tau_T \approx \tau_{EC} + \tau_{RC}\).
Fig. 2. Schematic drawing of energy band structure of a bipolar transistor (a) and its cross-sectional schematic view (b). E – emitter, B – base, C – collector, $\epsilon_C$ and $\epsilon_V$ – energies at the bottom of the conduction and the top of the valence bands, respectively, $\epsilon_F$ – Fermi level.

The maximum frequency of oscillation (unity power gain)

$$f_{max} \approx \frac{1}{2} \left[ \frac{\alpha \tau}{2 \pi (R_B C_C)_{eff}} \right]^{1/2},$$

(4)

is determined by the transfer coefficient $\alpha$, the transit time $\tau_T$ (3), and the charging time of capacitances $C_C$, connected with the collector junction, which is charged through the resistance of the base.
layer along the collector junction. The values $f_T$ and $f_{\text{max}}$ depend considerably on BT geometric dimensions. A BT is a vertical device as the charge carriers flow from the emitter towards the collector, crossing the horizontal $n$ and $p$ layers of the semiconductor structure (Fig. 2b). The reduce of the thickness of the base and the collector layers leads to the decrease of the carrier transit time $t_{BC}$ from the emitter to the collector and to the enhancement of $f_T$.

Modern technological methods allow to obtain vertical structures with thickness down to a single-atom layer. Technological progress is obvious there. Diffusion doping and ion implantation technology had been worked out in the sixties. That changed the alloyed one by 1970 allowing to develop BTs with the base depth of about $10^{-5}$ cm, which assured the high speed of a transistor at the level of 1 GHz. In 1970 liquid-phase epitaxy was created and developed, with the help of which the layers with the interface sharpness less than $10^{-6}$ cm were obtained. At last, in the eighties with the help of two powerful but rather complex methods, i.e., MBE (molecular-beam epitaxy) and MCVD (metal-organic chemical vapour deposition), the layers with interface sharpness less than $10^{-7}$ cm were obtained. At present both these technologies and methods to diagnose a semiconductor layer composition as well as doping of such thin layers have been developed.

On the other hand the reduction of the base thickness leads to the increase of horizontal base resistance and the charge time of the collector capacitance $R_{BC}C$. Therefore, the change of vertical thickness of layers calls for finding an optimum variant (see, Pożela, 1989).

How to reduce the thickness of the base and simultaneously to reduce the base resistivity is the main contradictory problem of the enhancement of BT operating speed. There are two approaches used to solve the problem. The first one is to minimize the transistor horizontal size. Because of the reduction of the horizontal base length there decrease the base resistance as well as the collector and other inner transistor capacitances. The reduction of the emitter size is limited too as the power of a BT is lowered and outer
resistances are increased. The size of the emitter and the collector in the contemporary BT are around 1–3 μm. More radical is the second approach – to enhance conductivity of the base layer by high impurity doping. In the homojunction silicon BT the doping of the base can not be large because of the decrease of the transfer coefficient α due to the increase of the ratio of the hole current $I_p$ to the electron current $I_n$ through the emitter. This ratio is equal to

$$\frac{I_p}{I_n} \approx \frac{N_A v_p}{N_D v_e},$$

(5)

where $N_A$ and $N_D$ are the concentrations of ionized acceptors in the base and donors in the emitter, respectively, $V_p$ is the diffusion hole velocity and $V_e$ is the corresponding velocity of the electrons injected through the barrier into the base. The ratio decreases with the increase of the base $N_A$ by comparison with the emitter, and the difference in doping must be great. This requirement is in conflict with the necessity for low base resistance, i.e., its high doping. That is the main limitation of enhancement of the operating speed of the homojunction Si transistors (Požela, 1989).

To overcome these contradictions the heterojunction between the wide-bandgap emitter and the narrow-bandgap base can be used (Fig. 3). In this case the ratio of hole current to electron one is reduced by $\exp\left(-\Delta \varepsilon/k_BT\right)$ times, where $\Delta \varepsilon$ is the supplemental energy barrier to holes, $k_B$ is the Boltzmann constant, $T$ is the absolute temperature (Kroemer, 1982). $\Delta \varepsilon$ ranges up to several tenth electronvolt and, therefore, the decrease of this ratio due to the $p$–$n$ heterojunction reaches several orders. That allows to decrease $R_B$ by a high doping of the base region. Besides, the electrons injected into the base from the heteroemitter have high initial velocity given by a heterobarrier. In heterojunction transistors there is a possibility to form a smooth graded bandgap structure allowing to create the built-in field for minority charge carriers. Finally, the heterojunction can be used to create two-dimensional high-conductance channel in the base at the collector.

Thus, one can solve the problem to increase the operating speed of BTs by using heterostructures as well as minimizing the
Fig. 3. Schematic drawing of energy band structure of the $p-n$ heterojunction (a) and cross-sectional view of a heterostructure bipolar transistor (b). $AlGaAs$ – wide-bandgap semiconductor, $GaAs$ – narrow-bandgap one.

sizes of the emitter and other transistor components.

In homojunction $Si$ BTs with the submicron emitter a high performance has been attained: $\tau_D$ ranges up to $\sim 30 - 100 \text{ ps}$, and $f_T$ up to $20 \text{ GHz}$. That appears to be the maximum possible operating speed of $Si$ BTs, because in the case of the emitter submicron sizes there occur a number of parasitic effects connected with the emitter vicinity and the ones decreasing transistor parameters.

There is a number of technological methods helping to improve the high-frequency parameters of $Si$ BTs. First of all, those are the
polysilicon contacts with emitter and base widely used in present-day bipolar silicon technology. That improves both the switching speed and the packing density of Si BTs in ICs as well as increases the gain. Due to higher values of BT current gain ($\beta_0 \approx 10^4$) the computer logic gates based on BTs have the higher operating speed by comparison with the gates based on FETs, though at higher power consumption per gate.

There are developed and fabricated heterojunction BTs (HBTs) based on GaAs with AlGaAs emitter regions. The characteristic parameters of the doping in the layers of HBT structure are as follows: $5 \cdot 10^{17} \text{cm}^{-3}$ in the emitter ($n-Al_{0.26}Ga_{0.74}As$); $10^{19} - 10^{20} \text{cm}^{-3}$ in the base ($n-GaAs$); $5 \cdot 10^{16} \text{cm}^{-3}$ in the collector ($n-GaAs$); $5 \cdot 10^{18} \text{cm}^{-3}$ in the subcollector ($n^+ - GaAs$). A high doping level of the base and a thick collector ensures a great decrease in $R_{BC}$ sharply increasing high-speed performance of the HBT. A typical size of the emitter area for high-speed HBTs is of $(1-2 \mu m) \times (2-10 \mu m)$ and is obtained by conventional lithography techniques. The high operating speed in AlGaAs/GaAs HBTs is thrice that in Si BTs.

The record values of high-speed parameters for AlGaAs/GaAs HBTs range up to the following values: $\tau_D \approx 5 - 15 \text{ ps at } P_D \approx 1 - 10 \text{ mW, } f_T \approx 170 \text{ GHz, } f_{\text{max}} = 218 \text{ GHz}$ (Chen et al., 1990). The current gain reaches 50000, transconductance $\approx 10^4 \text{ mS/mm}$ of the emitter length, when its width equals 1.2 $\mu m$ (Chang and Asbeck, 1990). The calculated switching times of HBTs reach 1-3 ps (Hu et al., 1989).

AlGaAs/GaAs HBTs are used effectively in digital circuits. There is produced a four bit universal up down counter for use in a frequency synthesizer. The circuit has 380 transistors and operates at a frequency of 2.86 GHz, which is twice the speed of the fastest existing counter. There are fabricated divide-by-four and divide-by-eight prescalers operating at frequencies up to 10 GHz. The HBT gate array which can be configured to give up to 144 equivalent logic gates and data converters with the sample rate (6 bit) 7.5 GSPS are manufactured. All these circuits are fabricated on a
basis of a transistor with the emitter length of 3.0 μm. The HBT with the emitter of 1 μm is supposed to have $f_{\text{max}} = 118$ GHz. Divider circuits using such a transistor can operate at clock frequencies of 45 GHz.

In recent years, intense efforts have been made to develop high-speed ICs operating at 10 Gbit/s for future optical transmission systems. These ICs have been mainly fabricated using the Si bipolar, GaAs MESFET and AlGaAs/GaAs HBT technologies. To achieve speeds far in excess of 10 Gbit/s, the AlGaAs/GaAs HBT is the most promising device among them because it has both the high transconductance of bipolar devices and the high electron mobility of compound materials. An ultrahigh-speed exclusive-OR/NOR(XOR)IC has been developed for future optical transmission systems (Ichino, et al., 1991). The IC was fabricated by AlGaAs/GaAs HBT technology. It operates up to, at least, 20 Gbit/s with extremely low rise/fall times of 22/14 ps.

Lately, besides the AlGaAs/GaAs heterostructure, there has been growing interest in the possibility of developing HBT, based also on the heterostructures with InGaAs base and InP, GaAs or AlGaAs emitters. The search for other structures is mainly connected with the requirements to integrate HBTs with optoelectron devices operating at different wavelengths (Pożela, 1989).

In recent years BTs with polysilicon emitter and SiGe base are developed. The value of $f_T = 75$ GHz for these transistors is much larger than for Si BTs (Patton et al., 1990). The SiGe BTs are compatible with usual Si integrated circuits.

3. Field-effect transistors. The field-effect transistor (FET) is a horizontal unipolar device (Fig. 4). The current between source and drain is regulated by the voltage on the third electrode – gate.

Qualitatively the operating speed of FETs is determined by charge time of the largest capacitance in the transistor (gate capacitance $C_G$) through the smallest resistivities, which are determined by the transconductance ($g_m = \partial I_D/\partial V_G$) and the differential conductivity of channel ($g_{ch} = \partial I_D/\partial V_D$). These transistor parameters
Fig. 4. Cross sectional view of a field-effect transistor. 1 – channel, 2 – depleted layer, S – source, G – gate, D – drain.

determine the cutoff frequency

\[ f_T = \frac{g_m}{2\pi C_G} \]  

(6)

and the intrinsic maximum frequency of oscillations

\[ f_{\text{max}} = \frac{f_T}{2} \left( \frac{g_m}{g_{ch}} \right)^{1/2} \]  

(7)

The transconductance of a FET has a maximum value in the region of drain current saturation (Shur, 1987; Pożela, 1989). The current saturation takes place, when the voltages applied to the gate and the drain are larger than the threshold voltage, at which the channel depletion depth at the drain side is compared with the channel depth \( a_{ch} \). In these conditions the maximum transconductance is equal to the differential conductivity of the fully open channel \( g_{ch}(v_d \rightarrow 0) \):

\[ g_{\text{max}} = 2q \frac{\mu n_s}{\lambda_G} w, \]  

(8)

where \( w \) is the gate width, \( \lambda_G \) is its length, \( \mu \) is the mobility of carriers, and \( n_s \) is the surface charge carrier density in the channel.

At homogeneous doping \( n_s = Na_{ch} \), where \( N \) is the number of ionized donors or acceptors in the channel. Equation (8) reveals the ways of achieving high-speed operation of FETs, i.e., the increasing carrier density \( n_s \) as well as their mobility \( \mu \) in the channel and the decreasing gate length \( \lambda_G \). Following from what has been said
we shall consider how the transistor operating speed is increased, firstly, due to its scaling, and later on, due to the choice of materials with high mobility and drift velocity of charge carriers.

One can well understand that the decrease of transistor dimensions leads to the electric field increase, what in turn leads to the saturation of the carrier drift velocity ($\nu_d = \nu_{sat}$). At the velocity saturation we have

$$f_T = \frac{\nu_{sat}}{2\pi\lambda_G} = \frac{1}{2\pi\tau_p},$$

where $\tau_p = \lambda_G / \nu_{sat}$ is the physical transit time of charge carriers through the gate.

The decrease in the gate length gives the increase in $f_T$.

Dimensions of horizontal structures are determined by lithography potentialities. When lithography is carried out by visible light having a wavelength 0.4–0.8 μm, then diffractional spread of the lighted region border restricts a possibility of obtaining a figure with lines thinner than 2–3 μm. Widely used ultraviolet photolithography (wave length is less than 0.3 μm) allows to obtain lines of 1–2 μm in length. X-ray and especially electronic lithography due to small wave length ($\lambda_e = h(\hbar\varepsilon_e)^{1/2} \approx 0,01$ nm, when $\varepsilon_e \approx 10^3$ eV) allows to reduce the diffractional spread down to a few hundredths of a micron.

The usual chemical etching techniques fail when an exposed figure with element dimensions less than a micron is to be etched. However, "dry" plasmachemical and ion etching techniques, which differ in their selectivity and anisotropy, yield the etching of regions down to 0,1 μm with an accuracy up to 0,05 μm.

It is unlikely that in the immediate future the horizontal dimensions of transistor circuit elements will be less than $10^{-1}$ – $10^{-2}$ μm, therefore, electron lithography and plasma ion etching possibilities do not essentially restrict these dimensions. Moreover, it is worth noting that electron beam lithography is still an expensive method, and the main thing is that it is of a very low productivity what is not allowable for mass production. Therefore, at present the search for high-productivity methods to obtain sub-micron circuit elements is continued. It should be pointed out that
ion beam lithography is one of the newest and evidently promising method in microelectronics technology. It combines the capabilities for exposing a resist to a resolution of better than 0.1 μm, its etching and, finally, using the ion beam to dope microregions of the same size (Moreau, 1988; Valiev, 1990).

In recent years along with the above discussed lithography techniques to form horizontal submicron structures the self-alignment techniques are used. To form submicron structures the anisotropy effect of etching rate with respect to crystallographic directions is also picked up.

Based on the methods of self-alignment, selective and anisotropic etching silicon MOSFETs with the gate length less than 0.3 μm are produced.

Modern technology does not limit the minimum dimensions of transistor. The limitation of miniaturization is arisen due to physical effects. Main of them are short-channel effects, the suppression of which is the main problem in high-speed FET design.

There are two short-channel effects. The first one, i.e., the parasitic short-channel effect is an abrupt increase of a leakage current lateral to the channel. The leakage is caused by the overlap of the depletion layers of the drain and source contacts. The leakage current is proportional to $\exp(-\lambda_G/L_D)$, where $L_D$ is the Debye screening length, and it abruptly increases, when $\lambda_G \sim L_D$. To eliminate the leakage the various potential barriers, e.g., $p-n$ junction between the channel and the substrate and between the drain and the source, are built in. In general, high potential barriers development on both sides of the channel is also useful to suppress hot electron emission to deep traps on the substrate and in insulated gate layer. The effect of hot-electron emission is responsible for the degradation of FETs (Quisse et al., 1990).

The second effect is the deformation of the potential distribution under the gate, when the channel length $\lambda_G$ appears to be comparable and smaller than the channel depth $a_{ch}$. This effect designated as the edge effect leads to the threshold voltage value $V_{th}$ sensitivity for small changes of transistor doping geometric pa-
rameters and to the inadmissible for IC scattering of values $V_{th}$ on
the semiconductor chip. Since the threshold channel cutoff voltage
shift is proportional to ratio $a_{ch}/\lambda_G$, then to avoid this effect at the
transistor scaling this ratio is held constant. Another way of sup­
pressing this effect is $a_{ch}$ minimization down to the formation of the
channel with two-dimensional electron gas at the heterojunction.

The increase of the carrier density $n_s$ in channel gives the in­
crease of $g_m$ and transistor operating speed (8). The doping level
and channel thickness display a natural limit – the breakdown of
the channel. Practically the maximum value $n_s$ is given by the
barrier height $\varphi_{ph}$. In the GaAs channel $n_s$ achieves $0.9 \times 10^{12} \text{cm}^{-2}$.

In silicon MOSFETs with submicron gates the high operating
speed with time delay under 30 ps at $P_{TD} < 40 \text{ fJ}$ is achieved (see,
e.g., Požela and Jucienė, 1985). However, the region $P_{TD} > 100 \text{ fJ}$
should be considered as the open-up region for silicon FETs. On
the basis of silicon CMOSFET the static RAMs with the number
of devices per chip up to $10^6$, with address access time up to 10
ns and a power dissipation of $5 \times 10^{-2} \text{ W}$ per chip are developed.
Practically all the modern computers use silicon ICs with micron
and submicron gates.

A further increase in operating speed of computers requires
devices with higher real-time digital signal processors. This can
be realized instead of Si using materials with higher mobility and
maximum drift velocity of charge carriers. Such materials are GaAs,
InP, InGaAs.

The GaAs MESFET with active layers formed by ion implan­
tation is a promising approach for realizing ultrahigh-speed LSI
(large scale integrated) circuits. The multifunction self-aligned gate
(MSAG) process technology (Bahl et al., 1990) and self-aligned
implantation for $n^+$-layer suppressing As outdiffusion technology
(SAINT) (Enoki et al., 1989) have proved to be capable of man­
ufacturing high-performance, highly reliable and reproductive de­
vices.

High-speed MESFETs obtained by SAINT process allowed to
realize 5.9 ps/gate operation ring oscillators, 26.8 GHz operation
High-speed semiconductor devices

Limits to the ultimate frequency performance which can be realized with GaAs MESFETs have been projected (Golio J.M. and Golio J.R.J., 1991). These predictions were based on the reported from 1966 to 1988 data performance of 137 devices with micron and submicron gates. The least-squares fit lines to the reported data may be given for the MESFET performance approximately by the equations:

\[ f_{\text{max}} \approx 40/\lambda_G \quad \text{(GHz)}, \]
\[ f_T \approx 10/\lambda_G \quad \text{(GHz)}, \]
\[ g_m \approx 60/\lambda_G \quad \text{(mS/mm)}, \]

where \( \lambda_G \) is in \( \mu \)m and \( \lambda_G \approx 0.1 - 10 \mu \)m. The gate-source capacitance has remained at approximately 1 pF/mm. The data projected at \( \lambda_G = 0.1 \mu \)m: \( f_{\text{max}} = 300 - 1000 \) GHz, \( f_T = 80 - 200 \) GHz and \( g_m = 300 - 1000 \) mS/mm. The noise temperature at 12 GHz for the submicron devices is given approximately by the equation

\[ T_s \approx 50 \lambda_p (K). \]

Mobility and maximum drift velocity are even more greater for \( \text{In}_{0.53}\text{Ga}_{0.47}\text{As} \) in comparison to GaAs. The saturation velocity evaluated from the drain current of \( \text{InGaAs} \) FETs has the value \( v_{\text{sat}} = 2.95 \times 10^7 \) cm/s, what is by the factor of 2 higher than that in GaAs. High value of maximum drift velocity is achieved due to enlargement of the energy separation between the \( \Gamma \)- and \( L \)-valleys with the increasing in percentage in the \( \text{InGaAs} \) alloy. \( \text{InGaAs} \) FETs with a gate length of 0.25 \( \mu \)m exhibit \( f_T = 126 \) GHz and \( f_{\text{max}} = 232 \) GHz (Feng et al., 1991).

The mobility and drift velocity of charge carriers in the FET channel depend on purity of the semiconductor.

The reduction of impurity scattering center number in the crystal yields a distinct increase in mobility. Therefore, the maximum purity of the material is determined by the maximum mobility. However, to increase the transistor conductance the high level of doping in the active area by impurities is a must. This discrepancy
between necessity of doping and attainment of high mobility can be removed in heterostructural transistors by selective (or modulated) as well as δ doping. If in a heterostructure, e.g., AlGaAs/GaAs, we dope by donors only the wide-bandgap AlGaAs layers, then with rather small layer thickness most of the electrons transit from them into the undoped layers of a more narrow-bandgap GaAs (Fig. 5). The electrons resulting from the impurity centers appear to be separated from them in a space and as a consequence of a screening effect of the ones the electron mobility in pure GaAs layers is rather high.

![Diagram of AlGaAs/GaAs heterojunction](image)

**Fig. 5.** Structure of the conduction band for an AlGaAs/GaAs heterojunction in a MODFET near the gate. $\varepsilon_{F0}$ – Fermi level at equilibrium.

A narrow channel with high charge carrier mobility is formed by the presence of selective (or modulated) doping of a heterojunction. Since the channel thickness in such a junction is less than the electron de Broglie wavelength, then the electron gas in the channel appears to be two-dimensional with discrete levels of energy. The channel thickness is of tens of nanometers, what is much less than its submicron length.

The FET with modulation doped heterostructure (MODFET)
is also called HEMT (high electron mobility transistor), TEGFET (two-dimensional electron gas transistor), SDHT (selectively doped heterojunction transistor). We shall use the MODFET name.

The unity current gain frequency $f_T$ in $AlGaAs/GaAs$ MODFETs with the gates $\lambda_G = 0.2 - 0.4 \, \mu m$ reaches 70–80 GHz, and in the heterojunction FETs with planar-doped gate $f_T = 80 - 120$ GHz, $f_{\text{max}} = 200 - 230$ GHz, when $\lambda_G = 0.2 - 0.1 \, \mu m$ (Morkoc, 1991).

Conduction band potential step in the $Al_{0.3}Ga_{0.7}As/GaAs$ heterojunction is of 0.24 V. That restricts GaAs electron sheet density to $(0.8 - 0.9) \times 10^{12} \, cm^{-2}$. The efficient drift velocity in the GaAs channel is of $1.2 \times 10^7 \, cm/s$. The addition of In into GaAs leads to an increase of conduction band potential step. Therefore, electron sheet density reaches $(1.6 - 1.7) \times 10^{12} \, cm^{-2}$ in the $In_{0.15}Ga_{0.85}As$ channel and $(2.4 - 2.5) \times 10^{12} \, cm^{-2}$ in the $In_{0.25}Ga_{0.75}As$ channel. In the latter case the conduction band potential step at the heterojunction reaches 0.44 V.

The record values $f_{\text{max}}$ for the $GaAs/InGaAs$ FET with $0.15 \times 150 \, \mu m$ gate reaches 152 GHz. The device has yielded 1.6 db noise figure at 60 GHz. A similar transistor with $\lambda_G = 0.08 \, \mu m$ has $f_{\text{max}}$ equal to 270 GHz.

The MODFET with $Al_{0.48}In_{0.52}As$ barrier and $Ga_{0.47}In_{0.53}As$ channel lattice-matched to $InP$ substrate has the highest frequency performance. The conduction band potential step is 0.52 V allowing electron sheet density up to $3 \times 10^{12} \, cm^{-2}$. The mobility in such a channel is found to be high, and the effective drift velocity reaches $2.4 \times 10^7 \, cm/s$. The device has 0.8 db noise figure at 60 GHz for $\lambda_G = 0.15 \, \mu m$ and $g_m = 880 \, mS/mm$, $f_T = 250$ GHz, $f_{\text{max}} = 450$ GHz at room temperature (Ho et al., 1991). A low noise level (1.4 db at 93 GHz) has been obtained in an analogous MODFET (Chao et al., 1990). The latter one has the great value $f_{\text{max}} = 405$ GHz. At present these transistors in spite of the existing problems concerning the leakage and parasitic charge in $GaInAs$ channel have the highest frequency response and the lowest noise figures.

The latest advances in the fabricating technology of the above listed high-speed transistors allow to develop high-speed LSIC
Fig. 6. Delay time $\tau_D$ versus power dissipation per gate $P_D$ for various types of transistors.

Based on them. The low values of $P_D\tau_D$ make them promising for VLSI (very large scale integrated) circuits because of their high operating speed and low power dissipation. At present static RAMs based on GaAs MODFETs with the number of devices above one hundred thousand are produced. Although the switching time of a discrete heterostructure FET is under 10 ps, the average propagation delay time per gate in IC as a result of operation performance delay in the logic gate and on the connecting metal tracks occurs to be greater by the order of magnitude. A complete signal delay in MODFET and MESFET ICs is of 50 – 200 ps/gate when the power being used is below 5 – 200 mW/gate. The processors with high operating speed above 5 GHz, with clock rate of more than 1 GHz, memory devices up to 64 K with select access time below 1 ns as well as analog-digital converters (8 bit) with more than...
Fig. 7. Address access time versus power dissipation of the static RAM compared with Si MOS, bipolar, GaAs MESFET and GaAs MODFET static RAMs.

$10^9$ selections/s are developed (Bursky, 1988). The maximum high speed regions obtained experimentally for discrete transistors are shown in Fig. 6 (Pożela, 1989). Fig. 7 compares the address access time and power dissipation of the static RAM (1–16 K) to Si MOS, bipolar, GaAs MESFET and GaAs/AlGaAs MODFET static RAMs (Abe et al., 1985).

4. Hot electron transistors. Several novel semiconductor device structures different from FETs and BTs have been proposed in order to increase the operating speed of the device. In many of them hot electron injection over a barrier or by tunneling through it is used. These devices have not yet received as wide application as computer components. However, they open new avenues for a
J. Poiela and V. Jucierl,j

557

qualitative increase of computer operating speed. Practically in this Section we shall discuss physical principles of the operation of new devices experimentally realized in research laboratories. We consider two groups of such devices: (1) ballistic and hot electron transistors; (2) static induction and permeable base transistors.

Ballistic electron transistor (BET), just as an ordinary bipolar transistor, has emitter, base and collector, but in contrast to a BT, a BET is a monopolar transistor. The electrons in a BET ballistically pass through a high doped base over the deep potential well.

In ballistic transistors the energy of an electron in the emitter is higher than that in the base. As a consequence the starting velocity of an injected electron in the base is very high. The velocity is determined by the barrier height $\varphi_{EB}$:

$$v_d = \sqrt{2\varphi_{EB}/m}.$$  

In the case of AlGaAs/GaAs heterojunction, emitter $\varphi_{EB} = 0.3$ V and the injected electron velocity will be $9.2 \cdot 10^7$ cm/s. In homogeneous GaAs crystal the maximum achieved drift velocity is only $2 \cdot 10^7$ cm/s.

After the injection the hot electron ballistically flies through the base. As a result of the scattering the mean free path of the ballistic electron injected with high drift velocity is short: $(1-3) \cdot 10^{-3}$ cm. The above mentioned means that thickness of the base of ballistic transistor have to be around $10^{-5}$ cm. The base is high doped ($\sim 10^{18}$ cm$^{-3}$), its resistivity is lower than 500 $\Omega$ on square and the time to charge the collector capacity ($R_{BC}$) is less than 1 ps. That opens the possibilities to design the ballistic transistors with $\tau_D < 1$ ps.

The ballistic electron transistors are developed with GaAs, InAs, InGaAs and Si base. The parameters of the realized BETs with planar-doped barriers and with hot electron tunneling through the barrier are as follows: $\tau_D \approx 1$ ps and $f_{\text{max}} \approx 15$ GHz.

The main disadvantage of the ballistic hot electron transistors is a sophisticated MBE technology of their fabrication.
In the real space electron transfer transistors due to nonhomogeneous heating of electron gas by electric field the real space electron transfer takes place. These effects are named the electrogradient ones (Dienys and Požela, 1971). The electrogradient effects can be used to create two types of high-speed transistors: (1) the transistors with negative resistivity in channel (NERFET) and (2) the transistors with charge injection of hot carriers through the barrier (CHINT).

The NERFET is the conventional GaAs FET structure but there are made additional AlGaAs and GaAs layers under the channel. The NERFET includes two devices in one: a FET with negative resistivity (due to hot-electron emission over AlGaAs barriers into the lower layer of GaAs) in the channel and some analog of a vacuum triode with the channel as a heated cathode and anode on the other side of AlGaAs barrier. The same structure may be used as injector of hot electrons in the layer of GaAs under the barrier. This structure is named as CHINT. The CHINT is some analog of a hot-electron ballistic transistor. In this device the upper GaAs layer and AlGaAs barrier are represented as the emitter and the ohmic contact at the lower GaAs layer is a collector.

The potential application of the real-space transfer devices are logic and memory elements of computers (Požela, 1989; Luryi and Pinto, 1991).

Static induction and permeable base transistors (SIT, PBT) in construction are analogues to vacuum tube triode, in which a vacuum space is changed by a semiconductor. These transistors are also called analog transistors or vertical field-effect transistors. In SITs, as opposed to FETs, the space charge carrier depletion region extends in the whole space between the electrodes, and the electron flow from the cathode (source) to the anode (drain) is regulated by the potential barrier height in the grid windows (gate). The vertical SIT construction assures a small submicron value of a transit length from a source to a drain and a small gate-source resistance. The SIT structure has a number of advantages for high-speed operation over FETs and BTs. The evaluation of the upper limit of the
current-gain cutoff frequency $f_T$ for an "ideal" SIT yields the value of 780 GHz. High-speed SITs for analog and switching applications are developed. SITs are used as high-speed phototransistors.

In permeable base transistors (PBT) the gate as a metallic grid or thin film with holes is imbeded into a semiconductor. In the latter case the PBT represents a semiconductor-metal-semiconductor structure. In contrast to SITs, the semiconductor layers in PBTs are doped ($N = 10^{16} - 10^{17} \text{cm}^{-3}$) and the current is controlled by thickness of the depletion layer of the Schottky barrier in the grid windows.

The fabrication of stacked grid PBTs can integrate logic gate functions within a single three-dimensional device. Tungsten, tantall and milibden are used as a metallic grid in GaAs PBTs. In Si PBTs the metal of a grid can be changed by silicid ($CoSi_2$) yielding the good Schottky barrier and having the lattice constant distinguishing itself only by 1.2% from Si. $CoSi_2$ grid is produced by local implantation of $^{65}Co$ into Si (Schuppen et al., 1990). In all the cases the technology of fabricating of PBT is rather complicated and retards the wide application of PBTs in ICs, although the advantages of PBTs over FETs and BTs are clearly seen.

PBTs belong to the fastest known three-terminal semiconductor devices. The calculated values for Si PBTs: $f_T \approx 35 \text{ GHz}$ and $f_{\text{max}} \approx 100 \text{ GHz}$ turn out to be higher than for silicon BTs and FETs. The GaAs PBT made by using a tungsten grid with a 0.32 $\mu$m periodicity has achieved $f_{\text{max}}$ values above 200 GHz (Badoz et al., 1990). Simulation of PBTs with a grid characteristics suggests a possibility of obtaining short delay times $\tau_D = 1.7 - 6 \text{ ps}$ under the ultra low product $P_D\tau_D \approx 0.1 - 0.4 \text{ fJ}$.

5. Resonant-tunneling transistors. Present-day technology has made possible the fabrication of semiconductor structures with the sizes comparable to the de Broglie wavelength for an electron in semiconductor. When the active regions of the devices decrease down to the value smaller than, roughly speaking, the size of a free electron in crystal, one should give up a conception of an electron as a classic particle and pay attention to wave nature of
High-speed semiconductor devices

electrons. New quantum-mechanical properties of electron gas in semiconductors manifest themselves in the decrease of electron freedom degree or electron gas dimension from the three-dimensional (3D) state down to the two-dimensional (2D), the one-dimensional (1D) or even the zero-dimensional (0D) state. Peculiarities of optical and electron properties of electron gas with a low dimension assign multifunctional possibilities to a semiconductor device to use them as logic circuits (Capasso et al., 1989).

At present a lot of physical properties of electron gas with a low dimension in semiconductors have been revealed (see, e.g., Workbook EP2DS-9, 1991). They basically manifest themselves at low temperatures. However, a number of devices, diodes and transistors operating at 300 K, in which these properties are used to achieve a high-speed operation of devices and give them multiple valued possibilities, are already developed. A double barrier quantum well structure (DBQW) is the basic element in these devices. The DBQW structure is realized as a heterostructure, in which between two layers of wide-bandgap semiconductor (e.g., GaAlAs) forming high potential barriers there is located the semiconductor layer with smaller bandgap (e.g., GaAs) forming a potential well of a structure. Electron tunneling through the DBQW structure takes place when the electron energy outside the structure coincides with the one of the discrete levels in the well and breaks down in the absence of such a coincidence. Therefore, current-voltage characteristics of the DBQW diode are of the N-type character, as is in the case of conventional tunneling diodes. However, tunneling through the resonant energy levels in the DBQW structure is more efficient than through an ordinary barrier. The tunneling inertia is determined by the value of electron delay time at the excited level of a quantum well:

$$\tau_D = 2\hbar/\Gamma,$$

where $\Gamma = \varepsilon_n T_R$, $\varepsilon_n$ is the transmission resonance energy and $T_R$ is the transmission of the individual barriers.

The $\Gamma$ value depends exponentially on the thickness and height of the barriers and of the well. In GaAs/AlAs structures with the
well and barriers thickness of 10–100 nm, the $\Gamma$ value is of the order of 10–100 meV, what yields $\tau_D = 10 - 100$ fs.

Low inertia of the tunneling through the DBQW structures makes it possible to build them into transistor structures what gives them new functional possibilities preserving the high operating speed. At present the diodes based on the DBQW structure belong to the fastest known semiconductor devices. Solner et al. (1984) experimentally observed the signal detection up to 2.5 THz and the oscillation frequencies in excess of 400 GHz by the $\text{AlGaAs}/\text{GaAs}/\text{AlGaAs}$ DBQW diode.

Various transistor modifications with the built-in DBQW structures or superlattices based on $\text{AlInAs}/\text{InGaAs}$ were developed (Požela, 1989). Bipolar transistors with the DBQW structure built in the base as well as in the emitter and operating at 300 K were firstly demonstrated in 1986.

The DBQW structure is also used in FETs as a gate, a drain and a source and even as a channel. The main peculiarity of resonant-tunneling BT and FET is integration of element with negative resistance with a conventional transistor. All that makes a resonant-tunneling transistor to become a multi-functional device on the basis of which the logic gates with a smaller number of transistors are developed. The NOR logic gate using only a single resonant transistor, instead of 7 conventional FETs, is built. A lot of functional potentialities make it possible to build a number of sequentially situated DBQW structures in the transistor. The devices with repetitive breaks-down on the current-voltage characteristic can be used as cells with several steady states as frequency multipliers. An 11-bit parity generator on five successively switched $\text{AlInAs}/\text{GaInAs}$ DBQW structures is proposed. This diode replaces ten EXCLUSIVE-OR gates.

Recently the structures, in which electron gas dimension is even lower, are under close investigation. The structures with one-dimensional electron gas – quantum wires – have been realized. Such structures are realized in nanometric mesoscopic ones as well as by placing metallic and semiconductor lines on the structure.
with 2D electron gas.

What is more, that there are realized zero-dimensional structures with quantum dots by placing dot elements on the semiconductor layer with 2D gas. An abrupt suppression of electron scattering processes is assumed to occur in homogeneous mesoscopic structures, what should lead to a mobility increase in uniform quantum wires. Quantum wires are waveguides for electron waves. A quantum electron wave interference effect has been observed. This effect can be used as the basis of a new class of semiconductor devices - quantum devices with low-dimensional electron gas. Investigation of interaction between the 2D and 1D layers is rather promising. A further decrease in IC packing density and device size miniaturization will apparently lead in future to the change of IC architecture itself, in which the coupling between the circuits will be of a quantum coupled character and the elements themselves will be multifunctional quantum devices with a low dimension.

6. Conclusions. Table 1 shows the achieved maximum values of operating speed for different types of discrete transistors. A maximum value of the threshold frequency for FETs is 450 GHz and for BTs is 248 GHz. The transistors with \( f_{\text{max}} > 200 \) GHz could be used as a switch device with delay switching time less than 1 ps. The above mentioned means that the transistors developed during the last two years may be used to design of the computer ICs with operating speed larger than many milliard operations per second.

These the newest high-speed transistors are made not from silicon and, therefore, they are not compatible with the contemporaneous silicon computer base. But transistors based on GaAs, GaInAs, InP have many advantages in comparison with the silicon transistors, especially, in ten times faster operating speed. Therefore, these transistors will replace silicon ones in the new computer generation. One important advantage of new high-speed transistors based on direct-band semiconductors is their compatibility with optoelectronic devices (lasers, photosensors, fibre lines). That opens the way to design optoelectronic computers where the information carrier is light, and the communication between the computer com-
ponents is based on optical fibres.

**Table 1. Maximum experimental values of operating speed parameters for various types of transistors**

<table>
<thead>
<tr>
<th>Transistor types</th>
<th>$f_T$ (GHz)</th>
<th>$f_{max}$ (GHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaAs MESFETs ($\lambda = 0.1 - 0.3 , \mu m$)</td>
<td>126</td>
<td>115 - 120</td>
</tr>
<tr>
<td>AlGaAs/GaAs MODFETs ($\lambda = 0.1 - 0.3 , \mu m$)</td>
<td>80 - 120</td>
<td>200 - 230</td>
</tr>
<tr>
<td>M/InGaAs/GaAs, InP MODFETs ($\lambda = 0.25 , \mu m$)</td>
<td>126</td>
<td>232</td>
</tr>
<tr>
<td>AlGaAs/InGaAs/GaAs MODFETs ($\lambda = 0.1 - 0.3 , \mu m$)</td>
<td>150 - 250</td>
<td>270 - 400</td>
</tr>
<tr>
<td>AlInAs/InGaAs/InP MODFETs ($\lambda = 0.1 - 0.3 , \mu m$)</td>
<td>170 - 250</td>
<td>405 - 450</td>
</tr>
<tr>
<td>AlGaAs/GaAs HBTs ($\lambda = (1 - 2) \times 10 , \mu m$)</td>
<td>170</td>
<td>218</td>
</tr>
</tbody>
</table>

Resonant tunneling DBQW structures with low-dimensional electron gas are already used as modulators and converters for optical signals.

Among the quantum semiconductor devices the negative-conductivity DBQW diodes have the highest operating speed. The incorporation of DBQW diodes into transistor structures makes it possible to raise their operating speed and gives them additional functional capabilities because of the negative regions in their current-voltage characteristics.

Possibly, this class of devices opens up a whole new era in ultrahigh-speed electronics (with operating speed parameters $f \approx 1000 \, \text{GHz}$, $\tau_D < 0.01 \, \text{ps}$).

**REFERENCES**

High-speed semiconductor devices


Luryi, S., and M.Pinto (1992). Symmetry of the real-space transfer and collector-


Received June 1992
V. Jucienė graduated from Department of Physics and Mathematics of Vilnius University in 1960, received her Ph.D. of Physics and Mathematics in 1975. She is a scientist research worker at the Semiconductor Physics Institute, Vilnius, Lithuania.

J. Požela graduated from Moscow University, Physics Department in 1952. In 1953–1956 he specialized at the Semiconductor Physics Institute of the USSR Academy of Sciences where he received his Ph.D. of Physics and Mathematics (1956). In 1964 for his work on the hot electrons in semiconductors J. Požela received his Dr. of Sciences from the Ioffe Physical-Technical Institute. In 1964 he became a professor of Vilnius University. In 1968 he was elected as a member of the Lithuanian Academy of Sciences. Since 1984 he is a member of the USSR Academy of Sciences (after 1991 – Russian Academy of Sciences). His range of interest in science is concentrated on semiconductor physics and electronics. He is the author of five monographs and more than 200 scientific papers.