BIBLIOGRFIA

1. A. S n y d e r: *Surface mode coupling along a tapered dielectric rod.* IEEE Transactions Antennas and Propagation, 1965, vol. AP-13, p. 821
2. G. B i e r n s o n, A. S n y d e r: *A modelling of vision employing optical mode patterns for color discriminatoin.* IEEE Transactions Systems Science and Cybernetics, 1968, vol. SSC-4, nr 7, p. 173
3. A. S n y d e r, P. H a l l: *Unification of the electromagnetic effects in human retinal receptors with three pigment color vision.* Nature. 1969, vol. 223, nr 8, p. 526
4. A. S n y d e r: *Coupling of modes on cylindrical tapered dielectric rod.* Proceedings of IEEE, 1969, nr 57, p. 737
5. A. S n y d e r: *Coupling of modes on a tapered dielectric cylinder.* IEEE Transactions on Microwave Theory and Techniques, 1970, vol. MTT-18, nr 7, p. 383
6. A. S n y d e r: *Mode propagation in nonuniform cylindrical medium.* IEEE Transactions on Microwave Theory and Techniques, 1971, vol. MTT-19, nr 4, p. 402
7. A. S n y d e r: *Coupled-mode theory for optical fibers.* Journal of the Optical Society of America, 1972, vol. 62, nr 11, p. 1267
8. A. N e l s o n: *Coupling optical waveguides by tapers.* Applied Optics, 1975, vol. 14, nr 12, p. 3012
9. R. W i n n, J. H a r r i s: *Coupling from multimode to single-mode linear waveguides using horn-shaped structures.* IEEE Transactions on Microwave Theory and Techniques, 1975, vol. MTT-23, nr 1, p. 1267
10. A. M i l t o n, W. B u r n s: *Mode coupling in optical waveguide horns.* IEEE Journal of Quantum Electronics, 1977, vol. QE-13, nr 10, p. 828
11. N. A m i t a y, H. P r e s b y: *Optical fiber tapers — a novel approach to self-aligned beam expansion and single-mode hardware,* Journal of Lightwave Technology, 1987, vol. LT-5, nr 1, p. 70
12. D. M a r c u s e: *Mode conversion in optical fibers with monotonically increasing core radius,* Journal of Lightwave Technology, 1987, vol. LT-5, nr 1, p. 125
13. D. M a r c u s e: *Theory of dielectric optical waveguides.* Academic Press, New York and London 1974, p. 95–126
14. A. M a j e w s k i: *Teoria i projektowanie światłowodów.* Wydawnictwa Naukowo-Techniczne, Warszawa 1991, p. 195–196
15. M. S z u s t a k o w s k i: *Elementy techniki światłowodowej.* Wydawnictwa Naukowo-Techniczne, Warszawa 1992, p. 120–124

K. PERLICKI

OPTICAL PROPERTIES OF TAPER CYLINDRICAL DIELECTRIC ROD

S u m m a r y

Multiwavelength wave propagation along taper cylindrical dielectric rod is described in the form of a coupled mode theory. Power transfer between the HE_{11} and HE_{12} mode is investigated. Analysis of taper optical connectors is presented.

Key words: wave propagation, optical connectors, dielectric rods.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 2, ss. 203–255

Semiconductor $m-n-n^+$ diodes for frequency conversion at millimeter and submillimeter waves

MAREK T. FABER AND MIROSŁAW E. ADAMSKI

Instytut Podstaw Elektroniki, Politechnika Warszawska

Otrzymano 1995.02.15

Autoryzowano do druku 1995.03.20

A comprehensive review of operating principles and properties of $m-n-n^+$ semiconductor diodes used now-a-days in millimeter- and submillimeter-wave frequency converters is presented. On these bases a circuit model of the Schottky diode (i.e., $m-n-n^+$ structure) is derived. The model includes both the voltage and frequency dependence of the diode series impedance and incorporates voltage modulated thickness of the depletion layer, driving up to flat-band conditions, carrier velocity saturation effect, skin effect (mainly in the substrate), dielectric relaxation (mainly in the epilayer) and carrier scattering both in the epilayer and in the substrate. The model is adequate at frequencies up to about 1 THz.

Key words: $m-n-n^+$, diode. Schottky barrier, varactors, noise, circuit model, GaAs, microvawes, frequency conversion, noise temperature.

1. INTRODUCTION

Frequency conversion requires in practice either nonlinear current-voltage (varistor) or capacitance-voltage (varactor) characteristic of a device. Real devices exhibit both nonlinarities and, depending on technology used to make a device and a circuit employing it, either nonlinearity can play dominant role in a converter. Devices used in frequency conversion process must have strong nonlinearities, repeatable and stable electrical properties and be fast enough for operation in a given frequency range. They should also be small in size and weight, reliable, long lived and should provide easy integration with circuits. Simple power supply, efficiency, high acceptable voltage and power levels complement the requirements.

Only two types of devices are of practical significance and general use at present in these applications (however, new emerging devices may change this situation). These are metal-semiconductor junction diodes (non strictly correctly but usually called

Schottky-barrier or Schottky diodes) and semiconductor $p-n$ junction diodes. Historically, $p-n$ diodes were first used in mixers and frequency multipliers. Special version of $p-n$ junction diode, i.e. step-recovery (or snap-off) diodes were used if high multiplication ratio was required at relatively low microwave frequencies. $p-n$ junction devices are subject to minority carrier recombination and suffer from the diffusion charge-storage effects which limit their application to varactor mode, typically at microwave frquencies.

In these respects Schottky-barrier diodes are superior because they are majority carrier devices. Furthermore, metal-semiconductor junctions can be fabricated more precisely than $p-n$ junctions and excellent, repeatable characteristics can be achieved. High quality junctions, sometimes with diameters below one micron, have been successfully developed making mixers and frequency multipliers feasible even at frequencies above 1000 GHz. Therefore, the dominant devices used in frequency converters at millimeter and submillimeter waves are the metal-semiconductor junction diodes. $p-n$ junction diodes are still used at lower frequencies while step-recovery diodes have limited application in frequency synthesizers as "comb generators".

The Schottky-barrier diode is a two-terminal semiconductor device which utilizes nonlinear properties of a metal-semiconductor contact. The rectifying properties arise from the presence of an electrostatic barrier between the metal (the anode) and the semiconductor (the cathode). The barrier is created by the unequal work functions of the metal and semiconductor and conduction is mainly controlled by thermionic emission of majority carriers over the barrier. The Schottky diode is therefore a majority carrier device whose cut-off frequency is not limited by minority carrier effects.

Electrical properties of a Schottky diode are predominantly defined by the metal and semiconductor combination and the size and condition of the contacting surfaces. Operation at high frequencies requires low series resistance and low junction capacitance. This implies the use of a semiconductor with high carrier mobility and saturation velocity. Most of the Schottky diodes in use at present are realized on a silicon (Si) or gallium arsenide (GaAs) semiconductor. Carrier mobility is greater for n-type materials than for p-type. Hence, n-type semiconductors are used almost exclusively for Schottky diodes. Many metals can create a Schottky barrier on either silicon or GaAs semiconductors; common are platinum, gold, aluminum [1 – 3].

GaAs is superior to silicon for high-frequency applications because its electron mobility and saturation velocity are much higher than that of silicon. At cryogenic temperatures ($\leqslant 20$ K) only GaAs diodes can be used because of lack of carrier freeze-out, which occurs in silicon around 40 K [4]. Although performance achieved with GaAs diodes is clearly superior to that achieved with silicon, silicon diodes are still in use at lower frequencies because of their significantly lower price.

The schematic drawing of the construction of a modern Schottky diode is presented at Figure 1. Practical diodes are fabricated on a lightly doped (donor concentration N_{De}) thin epitaxial layer that is grown on a heavily doped (concentration $N_{Db} \gg N_{De}$) buffer (substrate) layer. This $m-n-n^+$ structure allows the

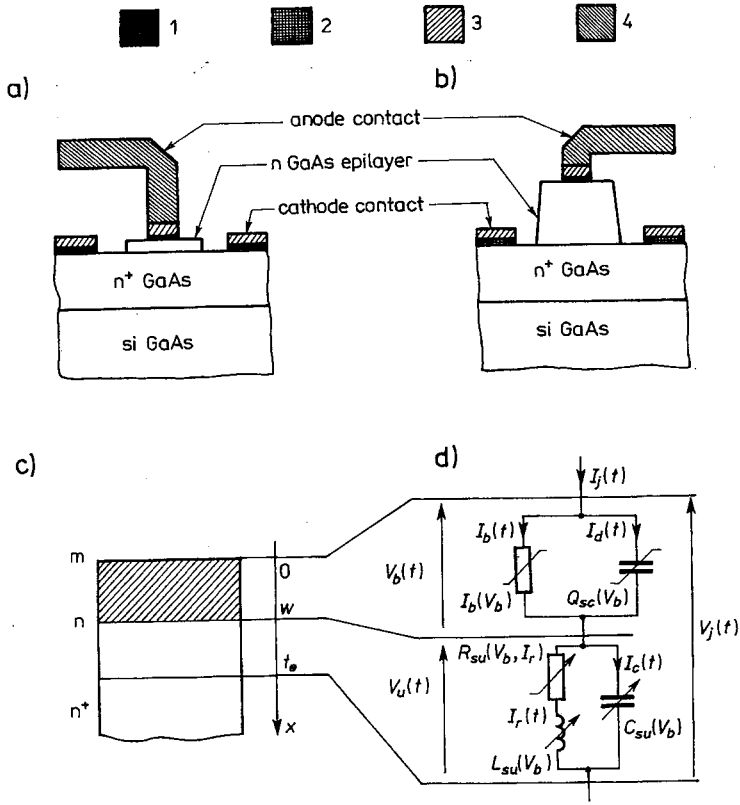


Fig. 1. Cross-sectional view of planar Schottky barrier diodes intended for use as a) varistor and b) varactor. c) one dimensional model of m–n–n$^+$ structure and d) quasistatic circuit model of the junction
1 – anode metal 2 – ohmic metal
3 – gold overlay 4 – electroplated gold

epitaxial layer to be used for the junction and the heavily doped region to minimize series resistive losses, i.e. real part of series impedance. The nonlinear junction shown at Figure 1.c is the rectifying metal-semiconductor contact and its direct vicinity within the epitaxial layer which has the thickness t_e. The surface of the metal defines the upper edge of the diode's junction ($x=0$) as shoiwn in Figure 1.c. The potential berrier below the anode is associated with a depletion region (layer). The depletion layer width $w(t)$, space-charge contained in this layer $Q_{sc}(t)$ and the barrier current $I_b(t)=SJ_b(t)$ (S is the junction area) are all dependent on the voltage drop $V_b(t)$ across the barrier and conditions at the barrier edge. These conditions are functions of the junction current $I_j(t)$ which is the sum of the barrier current $I_b(t)$ and the displacement current $I_d(t)=\mathrm{d}Q_{sc}/\mathrm{dt}$.

The maximum width (thickness) of the barrier w_{br} is determined by the breakdown voltage V_{br} at which reverse current increases rapidly due to avalanche multiplication of carriers. If the epitaxial layer thickness t_e is smaller than w_{br}, then the epilayer is punched through by the barrier at reverse bias $|V_b| < |V_{br}|$. The breakdown voltage

is then limited by the epitaxial layer thickness t_e. In Schottky diodes designed to be used as varactors, t_e is much larger and often even $t_e > w_{br}$, which allows the junction capacitance to follow $C(V_b)$ law up to the breakdown voltage.

The undepleted layer contains undepleted part of the epitaxial layer and/or thin layer of the buffer (substrate). The junction voltage $V_j(t)$ is the sum of the voltage drop $V_u(t)$ across the undepleted layer and the barrier voltage $V_b(t)$. The junction voltage differs from the voltage $V(t)$ applied to the diode because of the voltage drop across the buffer layer (substrate).

The operation of the Schottky diode involves two fundamental processes: firstly, formation of the surface electrostatic potential barrier and, secondly, transport of electrical carriers through the barrier and adjacent semiconductor regions (epitaxial layer and substrate). These will be presented in more detail in next sections.

2. FORMATION OF BARRIER

In order to explain formation of the potential barrier, let us examine the energy band structures of a metal and an *n*-type semiconductor shown at Figure 2. The work function, $q\Phi_m$ for the metal and $q\Phi_s$ for the semiconductor, is the difference between the Fermi level E_F for each material and the free-space energy level E_0. The work function is, therefore, the average energy required to remove an electron from the material. The electron affinity χ is defined such that when multiplied by the electron charge q, i.e. $q\chi$, equals the energy required to remove an electron from the bottom of conduction band to free space. χ is a constant for each material and must remain constant throughout it. However, the Fermi level of the semiconductor, and hence its work function, are dependent on doping density of the semiconductor. In equilibrium the energy levels are constant throughout each of the materials. Their Fermi levels are generally unequal indicating that the electrons in one material (in this case the metal) have less energy, on the average, then those in the semiconductor.

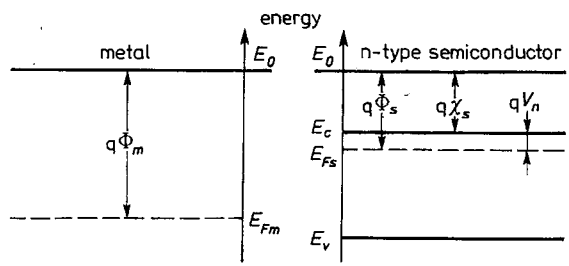


Fig. 2. Band structure of the metal and semiconductor before contact. E_0 is the free-space energy level, E_c is the bottom of the conduction band, and E_v is the top of the valence band. E_{Fm} and E_{Fs} are the Fermi levels in the metal and semiconductor, respectively

When such materials are joined, higher energy electrons of the semiconductor move spontaneously into the metal and collect on its surface. Electrons passing into

the metal leave behind ionized donor locations, which are positively charged. An electric field is created between these positive charges and negative surface charge of the electrons on the surface of the metal. This field eventually inhibits further electron flow into the metal, which happens when the Fermi levels of both materials are equalized.

If we consider that in equilibrium the Fermi levels for the semiconductor and metal must be constant throughout the system, the electron affinity must be constant, and that the free-space energy level must be continuous, then joining of the materials must result in bending valence and conduction bands of the semiconductor at the junction, as is shown in Figure 3.

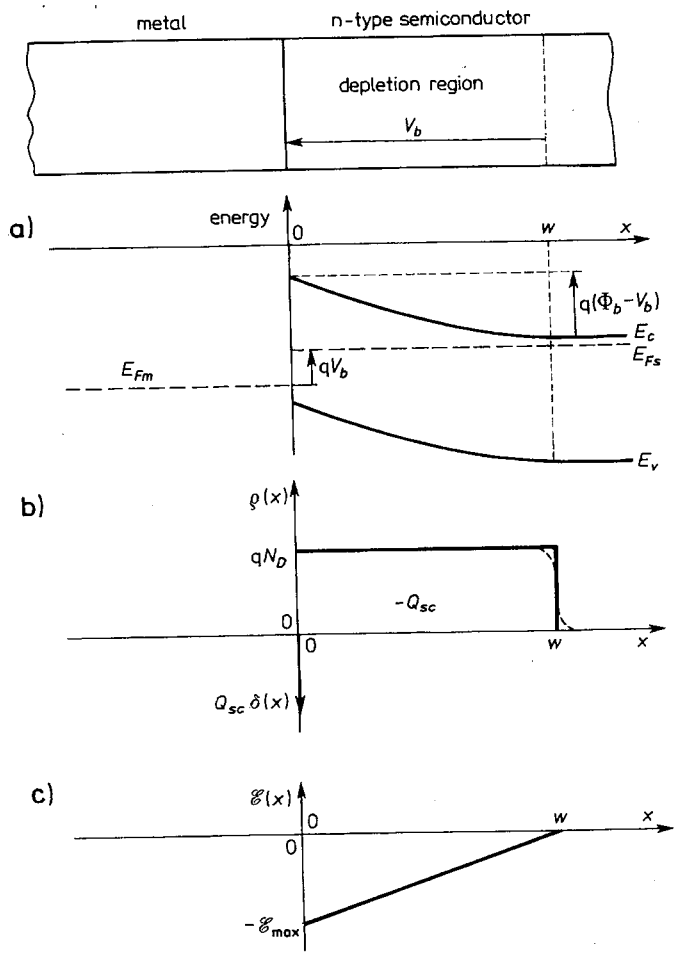


Fig. 3 a) Band structure of the Schottky junction; b) charge densities at the junction (the negative component is the surface electron concentration on the metal); c) electric field in the depletion region

Assuming that the region of the semiconductor where the bands are bent upwards is completely depleted of conduction electrons (what is known in the literature as "*depletion approximation*"), the space charge is due entirely to the uncompensated

donor ions. If these are uniformly distributed, there will be a uniform space charge in the *depletion region*, and the electric field strength will increase linearly from $-\mathscr{E}_{max}$ at the junction ($x=0$) to 0 at the edge of the depletion region ($x=w$). The magnitude of the electrostatic potential will decrease quadratically and the resulting potential barrier will be parabolic in shape. This is known as a "*Schottky barrier*" [1, 3]. Its height Φ_b is given by

$$\Phi_b=\Phi_m-\Phi_s+(E_c-E_F)/q=\Phi_m-\chi_s \tag{1}$$

where $\chi_s=\Phi_s-(E_c-E_F)/q$ is the electron affinity of the semiconductor.

In most practical metal-semiconductor contacts the ideal situation shown in Figure 3 is never reached, because there is usually a thin oxide layer, about 1-2 nm thick, on the surface of the semiconductor. A practical contact is, therefore, more like that shown at Figure 4. The additional barrier presented by the oxide *interfacial layer* is usually so thin that electrons can penetrate it by quantum-mechanical tunnelling.

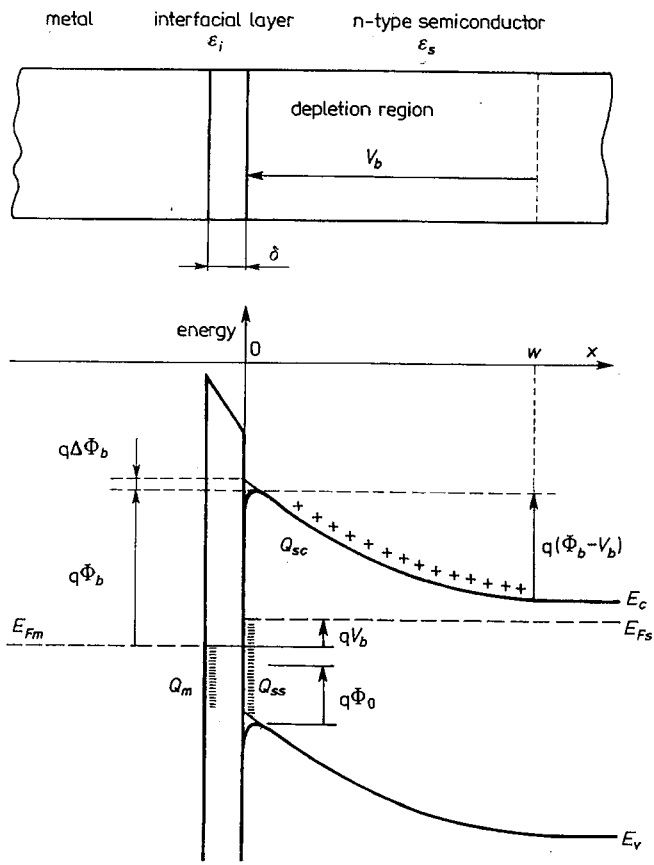


Fig. 4. Detailed energy-band diagram of a metal and *n*-type semiconductor contact with an interfacial layer. After: Sze S.M.: *Physics of semiconductor devices (2nd ed.)*, John Wiley and Sons, New York, 1981

Equation (1) assumes that the surface dipole layers do not change when the metal and the semiconductor are brought into contact. These surface dipole layers arise because at the surface of a solid the atoms have neighbors on one side only. This causes a distortion of the electron cloud belonging to the surface atoms, so that the centres of the positive and negative charge distributions do not coincide. The surface state theory [1, 3] predicts that an atom at the surface can either give up its electron, acting as a donor, or accept another, acting as an acceptor. The surface states are usually continuously distributed in energy within the forbidden bandgap, and are characterized by a "*neutral level*" potential Φ_0 (see Figure 4). If we consider that there is the thin oxide layer between the metal and the semiconductor, then the charge in the surface states together with its image charge on the surface of the metal will constitute a dipole layer. This dipole layer will alter the potential difference between the semiconductor and the metal and eqn. (1) has to be modified to:

$$\Phi_b = \gamma_b(\Phi_m - \chi_s) + (1-\gamma_b)(E_g/q - \Phi_0), \tag{2}$$

where

$$\gamma_b = \frac{\varepsilon_i}{\varepsilon_i + q\delta D_s}$$

and where E_g is the bandgap of the semiconductor, δ the thickness of the oxide layer, ε_i its total permittivity, and D_s is the energy density of surface states assumed to be uniformly distributed in energy within the bandgap.

Let us notice that if there are no surface states, $D_s = 0$ and $\gamma_b = 1$ and eqn. (2) gives $\Phi_b = \Phi_m - \chi_s$, which is the classical Schottky–Mott approximation of eqn. (1). If the density of states is very high, γ_b becomes very small and Φ_b approaches the value $E_g - \Phi_0$. In this case the barrier height is not dependent on the metal work function $q\Phi_m$, and is determined by the semiconductor and quality of its surface. This points out the importance of surface preparation of the semiconductor, e.g. [5–7].

Assuming that surface states influence only the barrier height, and under the abrupt charge density approximation (depletion approximation), i.e. $\rho \simeq qN_{De}$ for $x < w$ and $\rho \simeq 0$ for $x > w$, the width of the depletion region (layer) is given by

$$w = \left[\frac{2\varepsilon_s}{qN_{De}}(V_d - V_b - V_T)\right]^{\frac{1}{2}}, \tag{3}$$

where N_{De} is the donor doping density, ε_s is the total permittivity of the semiconductor, V_d is the so-called *diffusion* (or *build-in*) *voltage*, k is the Boltzman's constant and T is the temperature. The term $V_T = kT/q$ (≈ 26 mV at room temperature) arises from the presence of the transition region (at $x \approx w$) where the electron concentration falls from a value equal to N_{De} to a value negligible compared with N_{De}. V_b is a possible external voltage which, in general, may be applied to the barrier.

The strength of the electric field in the barrier changes linearly with distance x from the junction surface

$$|\mathscr{E}(x)| = \frac{qN_{De}}{\varepsilon_s}(w - x). \tag{4}$$

The diffusion voltage V_d differs from the barrier height Φ_b by a quantity V_n

$$\Phi_b = V_d + V_n \tag{5}$$

such that qV_n is the energy difference between the bottom of the conduction band and the Fermi level

$$V_n = \frac{kT}{q} F_{\frac{1}{2}}^{-1}\left(\frac{n\sqrt{\pi}}{2N_c}\right), \tag{6}$$

where $F_{\frac{1}{2}}^{-1}$ is the inverse Fermi function of order 1/2, n is the free-carrier concentration, and N_c is the effective density-of-states in the conduction band [8, 9]. V_n depends on temperature and semiconductor doping and, for example, for GaAs doped to $N_{De} = 3 \cdot 10^{16}\,\mathrm{cm}^{-3}$ $V_n = 75\,\mathrm{mV}$ at $T = 300\,\mathrm{K}$ and decreases to $V_n = 2\,\mathrm{mV}$ at $T = 20\,\mathrm{K}$ [10].

The space charge due to the ionized donors contained in the depletion layer (region) per unit area is then

$$Q_{sc} = \varepsilon_s \mid \mathscr{E}_{\max} \mid = qN_{De}w = [2q\varepsilon_s N_{De}(V_d - V_b - V_T)]^{\frac{1}{2}}. \tag{7}$$

Because there is no minority carrier storage in a Schottky diode, there is no diffusion capacitance either. Hence the capacitance of the junction is due only to the charge in the depletion layer and therefore

$$c_b \equiv \left|\frac{\partial Q_{sc}}{\partial V_b}\right| = \frac{\varepsilon_s}{w} = \left[\frac{q\varepsilon_s N_{De}}{2(V_d - V_b - V_T)}\right]^{\frac{1}{2}}. \tag{8}$$

The above expression can be written in the form

$$c_b = \frac{c_{b0}}{\left[1 - \dfrac{V_b}{V_d - V_T}\right]^{\frac{1}{2}}} = \frac{c_{b0}}{\left[1 - \dfrac{V_b}{\Phi_b - V_n - V_T}\right]^{\frac{1}{2}}}, \tag{9}$$

where c_{b0} is the barrier capacitance per unit area at zero bias $V_b = 0$, given by

$$c_{b0} = \left[\frac{q\varepsilon_s N_{De}}{2(V_d - V_T)}\right]^{\frac{1}{2}}. \tag{10}$$

When an electron in the semiconductor approaches the metal, a positive charge is induced on the metal surface. The force of attraction between the metal and the induced positive charge, so-called "*image force*", has the effect of reducing the barrier height by an amount that depends on the electric field in the semiconductor

$$\Delta\Phi = \left[\frac{q\mathscr{E}_{\max}}{4\pi\varepsilon_s}\right]^{\frac{1}{2}}. \tag{11}$$

The maximum value of the magnitude of the electric field strength occurs at $x \simeq 0$. Substituting eqn. (3) and (4) yields the "*image-force lowering*" of the Schottky barrier height

$$\Delta\Phi=\left[\frac{q^3N_{De}}{8\pi^2\varepsilon_s^3}(V_d-V_b-V_T)\right]^{\frac{1}{4}}. \tag{12}$$

From eqn. (2) considering eqn. (12), the height of the potential barrier is given by

$$\Phi_b=\gamma_b(\Phi_m-\chi_s)+(1-\gamma_b)(E_g/q-\Phi_0)-\Delta\Phi. \tag{13}$$

Thus, the height of the potential barrier is dependent on the junction bias voltage. For forward bias ($V_b>0$) the barrier height is slightly larger, while for reverse bias ($V_b<0$) slightly smaller than the barrier height at zero bias. Although the barrier lowering is small ($\approx$30 mV for typical GaAs diode), it does have a profound effect on current transport processes in metal-semiconductor systems, as discussed in next section.

The theory outlined above is based on many simplifying assumptions. It explains the formation of the barrier but it can not take into consideration many factors related to particular technological processes used in manufacturing Schottky diodes. The barrier height depends on the thickness and composition of the thin insulating layer at the metal-semiconductor interface, and this is bound to depend on the method of preparing the surface. As a result, the barrier height is nearly always determined by defects at the interface and for GaAs is almost independent on the metal. For most commonly used combinations, such as *n*-type GaAs and platinum, Φ_b usually is within 0.86 – 0.94 V, and 0.90 V in case of gold. For *n*-type silicon-platinum combination the barrier height is 0.90 V, and 0.80 V when gold is used as the metal [3, 11].

The barrier height is nearly independent of doping of a semiconductor. However, by introducing a thin layer ($\sim$100 Å or less) of semiconductor with a controllable number of dopants on a semiconductor surface (e.g., by ion implantation), the effective barrier height for a given metal-semiconductor contact can be varied. This layer allows additional control of the electric field strength in the semiconductor region adjacent to the interface. For example, by increasing the maximum field from 10^5 V/cm to 10^6 V/cm, the effective barrier height can generally be reduced by 0.2 V in silicon and over 0.3 V in gallium arsenide.

3. CURRENT-TRANSPORT MECHANISMS

Once the potential barrier has been formed as discussed in Section 1, the conduction properties of the junction are determined by the transport of carriers across the barrier. Under forward bias electrons can be transported through a metal-semiconductor junction by the following mechanisms [1]:

1) emission of electrons from the semiconductor over the top of the barrier into the metal;

2) quantum-mechanical tunnelling through the barrier;

3) recombination in the space-charge region;

4) recombination in the neutral region ("*hole injection*").

The inverse processes occur under reverse bias.

In most practical Schottky diodes mechanism 1) plays dominant role and determines the current/voltage characteristic. Such diodes are referred to as "*nearly ideal*". Processes 2), 3) and 4) cause departures from this ideal behaviour. Emission over the barrier and tunnelling through the barrier, shown schematically in Figure 5, are the most important processes and should be considered here. Hole injection is fully negligible for current densities below 10^4 A/m^2 and must be considered only in power rectifiers. Recombination in the depletion region (identical to the recombination process in a $p-n$ junction) in specific cases can cause small variations in I/V characteristics of GaAs diodes working in very low temperatures. These can be accounted for by introducing small changes in determining an effective temperature of the junction.

Emission over the Barrier Before electrons can be emitted over the barrier into the metal, they must first be transported from the interior of the semiconductor to the interface. Their motion through the depletion region of the semiconductor is governed by the mechanisms of diffusion and drift in the electric field of the barrier. When they arrive at the interface their emission into the metal is determined by the rate of transfer of electrons across the boundary between the metal and the semiconductor. These two processes are essentially in series and the current is determined predominantly by that process which sets lower limit to the flow of electrons.

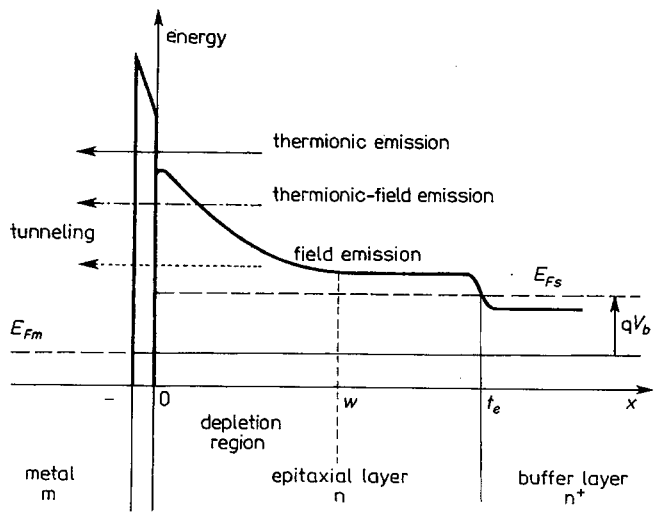


Fig. 5. Basic mechanisms of electron transport across a metal-semiconductor barrier at forward bias

The "*diffusion theory*" developed by Wagner and Schottky and Spenke [1] in the thirties assumes that current is limited by the diffusion and drift of electrons in the depletion region, and that the conduction electrons in the semiconductor immediately adjacent to the metal are in thermal equilibrium with those in the metal. For an n-type

semiconductor the current density J_b, as predicted by the diffusion theory, can be found from [1]

$$\frac{J_b}{kT\mu_n N_c}\int_0^w \exp\left(\frac{qE_c}{kT}\right)dx=\exp\left[\frac{q\zeta_n(w)}{kT}\right]-\exp\left[\frac{q\zeta_n(0)}{kT}\right], \tag{14}$$

where N_c is the effective density of states in the conduction band, qE_c is the energy of the bottom of the conduction band, μ_n the electron mobility, w is the width of the depletion region, and ζ_n is the quasi-Fermi level for electrons defined by

$$n=N_c\exp\left[\frac{-q(E_c-\zeta_n)}{kT}\right], \tag{15}$$

where n is the concentration of electrons in the n-type semiconductor.

The current/voltage relationship is completely determined if E_c is known as a function of x, and if the values of $\zeta_n(0)$ and $\zeta_n(w)$ can be specified for a particular value of applied bias. If we take the Fermi level in the metal as the zero energy level then $\zeta_n(w)$ is equal to the voltage V_b applied to the barrier, i.e. $\zeta_n(w)=V_b$. The assumption that the current flow is limited entirely by the processes of drift and diffusion in the depletion region is equivalent to $\zeta_n(0)=0$. Using the depletion approximation with a constant donor density N_{De} we have

$$E_c(x)=\Phi_b+\frac{qN_{De}}{2\varepsilon_s}(x^2-2wx), \tag{16}$$

where Φ_b is the barrier height and ε_s is the permittivity of the semiconductor. Substitution of $E_c(x)$ into eqn. (14) leads to Dawson's integral which has no analytical form and must be computed numerically. To obtain analytical J/V reltionship additional simplifying approximations are necessary — the simplest is to neglect the x^2 term in eqn. (16). This approximation is equivalent to assuming that the electric field strength $\mathscr{E}$ is constant and equal to its maximum value $\mathscr{E}_{max}$, given by eqns (3) and (4) for $x=0$, throughout the depletion region. Such assumption is justified only for $q(V_d-V_b)>4kT$, which excludes from consideration large values of forward bias. For the above conditions, the J/V relationship predicted by the diffusion theory is given by

$$\begin{aligned} J_b&=\left[qN_c\mu_n\mathscr{E}_{max}\exp\left(-\frac{q\Phi_b}{kT}\right)\right]\cdot\left[\exp\left(\frac{qV_b}{kT}\right)-1\right] \\ J_b&=J_{SD}\left[\exp\left(\frac{qV_b}{kT}\right)-1\right], \end{aligned} \tag{17}$$

where J_{SD} is the saturation current density, which is not constant but depends on the bias voltage because $\mathscr{E}_{max}$ is voltage dependent.

The "*thermionic emission theory*" developed by Bethe [3] in the forties in contrast to diffusion theory, assumes that the current is limited by the actual transfer of

electrons across the interface between the semiconductor and the metal, and that the effects of drift and diffusion in the depletion region are negligible (it is equivalent to assuming an infinite mobility). The above assumptions imply that quasi-Fermi level for electrons remains flat throughout the depletion region and coincides with the Fermi level in the bulk semiconductor. Thus, the electron concentration on the semiconductor side of the boundary is given by

$$n = N_c \exp\left[\frac{-q(\Phi_b - V_b)}{kT}\right] \tag{18}$$

The flux of these electrons across the interface into the metal is given by kinetic theory as $n\bar{v}/4$, where $\bar{v}$ is the average thermal velocity of electrons in the semiconductor. The flux in the average direction is independent of V_b because, neglecting image-force lowering, the barrier height Φ_b seen by electrons flowing from the metal remains unchanged by V_b. For zero bias ($V_b = 0$) the current from metal to semiconductor just balances the current from semiconductor to metal. The net current density is the difference of these two currents and therefore J_b is given by

$$J_b = \frac{qN_c\bar{v}}{4} \exp\left(\frac{-q\Phi_b}{kT}\right) \cdot \left[\exp\left(\frac{qV_b}{kT}\right) - 1\right]. \tag{19}$$

If we consider that $A^* = 4\pi m^* qk^2/h^3$ is the Richardson constant corresponding to the effective mass in the semiconductor m^*, then eqn. (19) can be written in the form

$$\begin{aligned} J_b &= \left[A^* T^2 \exp\left(-\frac{q\Phi_b}{kT}\right)\right] \cdot \left[\exp\left(\frac{qV_b}{kT}\right) - 1\right] \\ J_b &= J_{ST}\left[\exp\left(\frac{qV_b}{kT}\right) - 1\right]. \end{aligned} \tag{20}$$

The "*thermionic-emission/diffusion theory*" as a synthesis of the above theories was proposed by Crowell and Sze [3] in the sixties. Considering the two current limiting mechanisms to be in series, the theory has allowed the position of the quasi-Fermi level to be found at the interface which equalizes the current flowing due to each of the mechanisms. Introducing the concept of a "*recombination velocity*" $v_r = \bar{v}/4$ at the top of the barrier (which occurs within the semiconductor owing to the effect of the image force), and adapting the reasoning which led to eqn. (17), the thermionic-emission current is derived as

$$J_T = qN_c v_r \exp\left(\frac{-q\Phi_b}{kT}\right) \cdot \left\{\exp\left(\frac{q\zeta_n(0)}{kT}\right) - 1\right\}, \tag{21}$$

where $\zeta_n(0)$ is the position of the quasi-Fermi level at the interface.
Equation (14) can be used to give the current limited by drift and diffusion in the depletion region as

$$J_D=\frac{kT\mu_n N_c\left\{\exp\left(\frac{qV_b}{kT}\right)-\exp\left[\frac{q\zeta_n(0)}{kT}\right]\right\}}{\int_0^w \exp\left(\frac{qE_c}{kT}\right)dx}. \tag{22}$$

Because the thermionic-emission current must equal to current determined by drift and diffusion, $\zeta_n(0)$ can be eliminated between eqns. (21) and (22) and finally current density $J_b=J_T=J_D$ can be written as

$$J_b=\frac{qN_c v_r}{1+(v_r/v_d)}\exp\left(\frac{-q\Phi_b}{kT}\right)\cdot\left[\exp\left(\frac{qV_b}{kT}\right)-1\right], \tag{23}$$

where

$$v_d=\left[\frac{q}{\mu_n kT}\exp\left(\frac{-q\Phi_b}{kT}\right)\int_0^w \exp\left(\frac{qE_c}{kT}\right)dx\right]^{-1} \tag{24}$$

is an effective diffusion velocity associated with the transport of electrons from the edge of the depletion region at w to the potential energy maximum. The integral in the expression (24) can be expressed in terms of Dawson's integral. If one adopts the same approximation as was used in deriving eqn. (17), namely that the electric field is constant and equal to its maximum value $\mathscr{E}_{max}$ given by (4) for $x=0$ and with (3) (thus excluding large values of forward bias), eqn. (24) simplifies to $v_d=\mu_n\mathscr{E}_{max}$.

Crowell and Sze also take into account the effects of optical phonon scattering in the region between the top of the barrier and the metal, and of the quantum mechanical reflection of electrons which have sufficient energy to surmount the barrier. Their combined effect is to replace the Richardson constant A^* with A^{**}, i.e. except for high bias, J/V characteristic can be expressed as

$$J_b=\left[A^{**}T^2\exp\left(-\frac{q\Phi_b}{kT}\right)\right]\cdot\left[\exp\left(\frac{qV_b}{kT}\right)-1\right]$$
$$J_b=J_S\left[\exp\left(\frac{qV_b}{kT}\right)-1\right], \tag{25}$$

with modified Richardson constant in the form

$$A^{**}=\frac{f_p f_q A^*}{1+f_p f_q\frac{v_r}{v_d}}, \tag{26}$$

where f_p is the probability of an electron reaching the metal without being scattered by an optical phonon after having passed the top of the barrier, and f_q is the average transmission coefficient. At room temperature for high mobility semiconductors doped to $10^{16}-10^{17}$ cm^{-3} and in the electric field strength range $10^4-2\cdot10^5$ V/cm, A^{**}

remains essentially at a constant value which, as calculated in [3], is about 96 $Acm^{-2}K^{-2}$ for silicon and 4.4 $Acm^{-2}K^{-2}$ for GaAs.

Although the above ideas about phonon scattering and quantum-mechanical reflection are qualitatively correct, it is not clear how reliable they may be quantitatively and one must depend on experiments. Except for high bias, the value of $A^{**}=96$ $Acm^{-2}K^{-2}$ for silicon is confirmed experimentally to be accurate to within 2%. For gallium arsenide the commonly accepted value is from 8.2 $Acm^{-2}K^{-2}$ [10, 12] to 8.6 $Acm^{-2}K^{-2}$ [13]. The low value of the modified Richardson constant for GaAs implies that the knee of the J/V characteristic occurs at higher applied voltages for GaAs diodes.

In conclusion, it should be stated that there are good theoretical reasons, confirmed experimentally, for believing that in Schottky diodes made from fairly high-mobility semiconductors (as Si or GaAs) the forward current at room temperature is limited by thermionic emission and eqn. (25) well describes J/V relationship that the forward bias is not too large.

Tunnelling through the Barrier In a heavily doped semiconductor and/or operation at low temperatures the current flowing through the junction can be influenced or even determined by quantum mechanical tunnelling of electrons through the barrier. At low temperature and high donor density the potential barrier may be so thin that electrons with energies close to the Fermi energy in the semiconductor can easily tunnel through the barrier. This is known as *field emission*. If the temperature is raised electrons are excited to higher energies and thus they see thinner and lower barrier. The probability of tunnelling through the barrier increases very rapidly. On the other hand, the number of electrons having a particular energy decreases very rapidly with increasing energy. As a result, the current is stabilized at a certain value (which depends on temperature). This is known as *thermionic-field emission*. Further increase of temperature leads eventually to a point at which virtually all of the electrons have enough energy to go over the top of the barrier. The tunnelling is now negligible and so pure *thermionic emission* determines the current.

From the theory of field and thermionic-field emission developed by Padovani and Straton and by Crowell and Rideout [1, 3] in the sixties, it follows that field emission occurs only in degenerate semiconductors, and because of the very small effective mass, it shows up at lower concentrations in gallium arsenide than in most other semiconductors. For example, in GaAs doped to $N_D \approx 10^{17}$ cm^{-3} it shows up at $kT \approx 2$ meV [1, 6, 8, 14]. In silicon diodes tunnelling may be neglected in most practical cases.

Except for very low values of V_b the forward current-voltage relationship is given by [1]

$$J_b = J_S \exp\left(\frac{V_p}{E_0}\right), \tag{27}$$

where

$$E_0 = E_{00} \coth\left(\frac{qE_{00}}{kT}\right) \tag{28}$$

and
$$E_{00}=\frac{h}{4\pi}\left(\frac{N_{De}}{m^*\varepsilon_s}\right)^{\frac{1}{2}}. \tag{29}$$

Here $m^*=m_r m_0$ is the effective mass of electrons in the n-type semiconductor, $\varepsilon_s=\varepsilon_r\varepsilon_o$ its permittivity and h is the Planck's constant. The pre-exponential term J_S is a complicated function of the temperature, barrier height and semiconductor parameters.

At low temperatures ($kT/qE_{00}\ll 1$), $E_0\simeq E_{00}$ (field emission), and so the slope of the graph of $\ln J_b$ against V_b is independent of temperature. At high temperatures ($kT/qE_{00}\gg 1$), E_0 approaches the value kT/q, which corresponds to pure thermionic emission. For temperatures such that $kT\sim qE_{00}$, thermonic-field emission occurs and the slope of the $\ln J_b$ against V_b curve can be written as $q/\eta kT$, where $\eta=qE_0/kT=(qE_{00}/kT)\coth(qE_{00}/kT)$. Precise theoretical determination of the temperature ranges at which particular emission mechanisms dominate is rather very complicated, but they can be found experimentally by measuring the slope of the $\ln J_b(V_b)$ curve as a function of temperature. From such measurements it comes out that in Schottky diodes made from n-type GaAs doped to $N_D\approx 2\cdot 10^{17}$ cm^{-3} the tunnelling of electrons may be neglected (i.e. pure thermionic emission dominates) at temperatures above 100 K [8, 10, 24]. Lowering the doping to $N_D\approx 10^{16}$ cm^{-3} lowers this temperature to below 30 K.

4. GENERALIZED I/V RELATIONSHIP

Formal similarity of J/V relationship in case of pure thermionic emission (eqn. 23), and in case of field and thermionic-field emission (eqn. 25), induces us to write the J/V relationship in a generalized form which would be describing the junction at any temperature range and for various semiconductor doping.

Let us introduce the concept of an "*ideality factor*" η of the junction, defined such that $\eta=1$ for pure thermionic emission, and its increase would be due to any deviation from this ideal model

$$\eta\equiv\frac{q}{kT}\frac{\partial V_b}{\partial(\ln J_b)}. \tag{30}$$

With the above definition I/V characteristic of the junction in which current flows due to pure thermionic emission, for $I_b\gg I_S$ may be expressed in the form

$$I_b=SA^{**}T^2\exp\left(\frac{-q\Phi_b}{\eta kT}\right)\exp\left(\frac{qV_b}{\eta kT}\right)=SA^{**}T^2\exp\left(\frac{V_b-\Phi_b}{V_0}\right) \tag{31}$$

where S is the area of the junction and

$$V_0=\frac{k}{q}\eta T=8{,}617\cdot 10^{-5}\,\eta T \quad [\mathrm{V}] \tag{32}$$

is the "*slope parameter*" of the I/V characteristic. Small variations of η from 1 come from image-force lowering of the barrier and other minor effects. Let us point out that for pure thermionic emission, V_0 is not dependent on the junction current and its value decreases linearly with decreasing temperature [8, 15].

Deviations from the above model are caused primarily by tunnelling of electrons through the barrier. In general it can be written

$$V_0 = E_0 = E_{00} \coth\left(\frac{qE_{00}}{kT}\right) \tag{33}$$

and

$$\eta = \frac{q}{kT} E_{00} \coth\left(\frac{qE_{00}}{kT}\right), \tag{34}$$

where E_{00} is given by eqn. (29).

For pure thermionic emission, i.e. high temperature and/or low semiconductor doping, $\coth(qE_{00}/kT) \to kT/qE_{00}$ and $\eta = 1$ and $V_0 = kT/q$ depends linearly on temperature. For field emission, i.e low temperature and/or high semiconductor doping, $\coth(qE_{00}/kT) \to 1$ and $\eta = qE_{00}/kT$ and $V_0 = E_{00}$ is not dependent on temperature.

Another generalization of I/V relationship comes from the concept of an "*effective junction temperature*" θ which may be different from physical temperature T of the junction [10, 16, 24]

$$\theta \equiv \frac{qI_b}{k} \cdot \frac{dV_b}{dI_b} \tag{35}$$

For pure thermonic emission and $I_b \gg I_S$ the I/V characteristic is thus given by

$$I_b = SA^{**} \theta^2 \exp\left[\frac{q(V_b - \Phi_b)}{k\theta}\right]. \tag{36}$$

In order to preserve the form of the above characteristic it is necessary to relate θ to particular carrier-transport mechanism. So we have

for pure thermionic emission

$$\theta = \theta_T = \eta T \tag{37}$$

for field emission

$$\theta = \theta_F = \frac{qh}{4\pi k}\left(\frac{N_{De}}{\varepsilon_s m^*}\right)^{\frac{1}{2}} \tag{38}$$

and for thermionic-field emission

$$\theta = \theta_{TF} = \theta_F \coth\left(\frac{\theta_F}{T}\right). \tag{39}$$

In the above model field emission shows up as an increase of the effective temperature above the junction physical temperature which prevents the electron gas from being "*cooled*" below θ_F temperature. The only way to lower θ_F is lowering the semiconductor doping.

Let us notice that

$$\theta=\frac{qV_0}{k}=\eta T \tag{40}$$

and the two models are equivalent. Which one is used depends on an ease and convenience of interpreting the tunnelling of electrons through the potential barrier, either as a change in the I/V characteristic slope caused by the increase of ideality factor η, or as an increase of the effective temperature which makes it impossible to cool the electrons below θ_F temperature.

5. HIGH BIAS CONDITIONS

Increasing the operating frequency of the diode requires us to reduce junction capacitance, i.e. to reduce the junction area. Small area in most practical cases means higher current densities, and the instant value of the current density may be high enough to cause the depletion layer almost to vanish. For example, in diodes with junction diameters $d\approx 1$ μm used at higher millimeter waves, this occurs for a value of current ≈ 4 mA at room temperature and as low as ≈ 1.5 mA at cryogenic temperatures.

The commonly accepted current/voltage relationship, eqn. (25), predicts that $\log | J_b |$ varies linearly with V_b in the high forward direction, with a slope of $q/\eta kT=1/\eta V_T$. This expectation together with capacitance/voltage dependence of eqn. (9) lead to many confusions among engineers and researchers. Indeed, simplest inspection of eqn. (25) and eqn. (9) shows that things go disastrously wrong as V_b approaches V_d; the model fails to reflect the intuitive and reasonable expectation that there should be no voltage drop at a nonexistent barrier (i.e. junction resistance should go to zero), and that the capacitance of the barrier should not increase to infinity when the applied forward voltage is large enough to make the barrier vanish. (In fact, (7) predicts that Q_{SC} vanishes at $V_b=V_d-V_T$ but not at $V_b=V_d$).

To avoid discrepancies between calculated and measured junction behaviour some authors were forced to make nonphysical assumption, such as assuming that the diffusion voltage V_d equals the barrier height Φ_b or neglecting $V_T=kT/q$ term in eqn. (9). Assuming a noncontinuous junction model, i.e. that at $V_b\geqslant V_d$ the junction is replaced by the zero capacitance and resistance, does not solve the problem either. Another approach is the numerical modelling of the junction, for example [17]. This approach is difficult to implement in a general computer program used to analyze microwave circuits with diodes. It also loses ease of physical interpretation and intuitive control over diode and circuit modelling. In these respects the analytical

model has a great advantage and has been always appreciated by engineers and researchers. Some solution to the problem might be achieved by reexamining the approximatinos made in the derivations of eqns. (9) and (25), as was suggested in [18].

Let us clearly state what is meant by a "*flat-band*" condition. The flat-band condition is defined as the case where the forward voltage applied to the junction is large enough to shrink the depletion region length to zero [9]. At this point the potential barrier from the semiconductor to the metal is reduced to zero and the conduction band is gradient free (or flat) provided that the electric field in the undepleted epitaxial layer is neglected. The flat-band voltage V_{fb} is thus defined as

$$V_{fb}=\Phi_b-V_b=V_d. \tag{41}$$

The electric field in the barrier is one of major factors which determine the properties of the junction. Unless the barrier height is so large that the surface is strongly *p*-type, the maximum value of the strength of electric field in the barrier is given by [1]

$$\mathscr{E}^2_{\max}=\frac{2qN_{De}}{\varepsilon_s}\left[(V_d-V_b-V_T)+V_T\exp\left(\frac{-(V_d-V_b)}{V_T}\right)\right]. \tag{42}$$

let us notice that if $q(V_d-V_b)>3kT$, i.e. the difference between the bias voltage and the flat-band voltage is greater than $3V_T=3kT/q\approx 80$ mV at room temperature, eqn. (42) reduces to $\mathscr{E}^2_{\max}=(2qN_{De}/\varepsilon_s)(V_d-V_b-V_T)$ which is the form of eqn. (4) used in deriving C/V and I/V relationships for small and medium bias conditions.

It is important to realize that in this more exact theory the depletion region does not have a precisely defined width, since the bottom of the conduction band approaches its position in the bulk of the semiconductor asymptotically. Making use of the relationship $|\mathscr{E}_{\max}|=qN_{De}w/\varepsilon_s=Q_{sc}/\varepsilon_s$ the effective width of the depletion region can be given by

$$w=\left(\frac{2\varepsilon_s}{qN_{De}}\right)^{\frac{1}{2}}\left\{V_d-V_b-V_T\left[1-\exp\left(\frac{-(V_d-V_b)}{V_T}\right)\right]\right\}^{\frac{1}{2}} \tag{43}$$

and eqn. (7) can be replaced by a more general form of

$$Q_{sc}=\varepsilon_s|\mathscr{E}_{\max}|=(2q\varepsilon_sN_{De})^{\frac{1}{2}}\left\{V_d-V_b-V_T\left[1-\exp\left(\frac{-(V_d-V_b)}{V_T}\right)\right]\right\}^{\frac{1}{2}}. \tag{44}$$

Then the expression for the depletion layer capacitance is derived as

$$c_b\equiv\left|\frac{\partial Q_{sc}}{\partial V_b}\right|=\left(\frac{q\varepsilon_sN_{De}}{2}\right)^{\frac{1}{2}}\frac{1-\exp\left(\frac{-(V_d-V_b)}{V_T}\right)}{\left\{V_d-V_b-V_T\left[1-\exp\left(\frac{-(V_d-V_b)}{V_T}\right)\right]\right\}^{\frac{1}{2}}} \tag{45}$$

which can be written in the form corresponding to eqn. (9)

$$c_b = c_{b0} \frac{1-\exp\left(\frac{-(V_d-V_b)}{V_T}\right)}{\left[1-\frac{V_b}{V_d-V_T}+\frac{V_T}{V_d-V_T}\exp\left(\frac{-(V_d-V_b)}{V_T}\right)\right]^{\frac{1}{2}}}, \tag{46}$$

where c_{b0} is the zero bias junction (barrier) capacitance per unit area and, as before, is given by eqn. (10). If $V_d - V_T \approx V_d$ is assumed (i.e., the term $V_T = kT/q$ is neglected) the above expression corresponds to expression given in [18].

More precise description of the depletion layer also calls for reexamining derivations of eqns. (17) and (25) in search for analytical expression which would extend J/V relationship for bias close to the flat-band condition. A better approximation of the solution of equation (14) was given in [19] after original derivations of Mott and Schottky. Using our notation this more general current/voltage relationship may be written as

$$J_b = J_S \frac{\exp\left(\frac{V_b}{\eta V_T}\right)-1}{1-\exp\left(\frac{-2(V_d-V_b)}{V_T}\right)}, \tag{47}$$

where $J_S(V_b) \propto w(V_b)$ slowly decreases when V_b increases. Assuming $J_S = \text{const}(V_b)$ eqn. (47) for $J_b \gg J_S$ and small values η may be written in a more convenient form similar to that in [18]

$$J_b = J_{fb} \frac{1}{2\sinh\left(\frac{(V_d-V_b)}{\eta V_T}\right)}, \tag{48}$$

where $J_{fb} = J_S \exp[V_d/(\eta V_T)]$. An analytic approximation of the diode characteristic which is in even better agreement with exact numerical solution resulting from the Dowson integral is given in [20]. In that approximation the denominator in eqn. (47) is replaced by its square root. The barrier incremental resistance is now

$$R_b = \frac{1}{(\partial I_b/\partial V_b)} = \frac{\eta V_T}{J_b S} \tanh\left(\frac{(V_d-V_b)}{\eta V_T}\right). \tag{49}$$

Comparison between the "*classical*" model of the metal-semiconductor junction and the model which extends to voltages close to the flat-band condition is given at Figures 6 and 7. Examining these figures and equations (46), (47) and (49), it is seen that for bias voltages such that the difference between the bias and the flat-band voltage is greater than $3kT/q$, the extended model reduces to the classical one. When the forward bias voltage would approach the flat-band voltage $V_d = \Phi_b - V_n$, the depletion layer would disappear, the space-charge would reduce to zero and the incremental capacitance of the junction would be limited. At the same voltage, the junction resistance would decrease to zero and the current flow would be determined

by the drift and diffusion of electrons through the semiconductor. The limited value of the current (i.e. $J_b < \infty$) causes that the limiting case $V_b = V_d$ can never be reached.

The presented analytical model is easy to implement in a computer program used for analysis of microwave circuits with diodes. It is often sufficient to replace old expressions such as that of (9) and (25) with the new ones. This allows more practical and realistic cases of diode operation, preserving clear physical interpretation of junction behaviour.

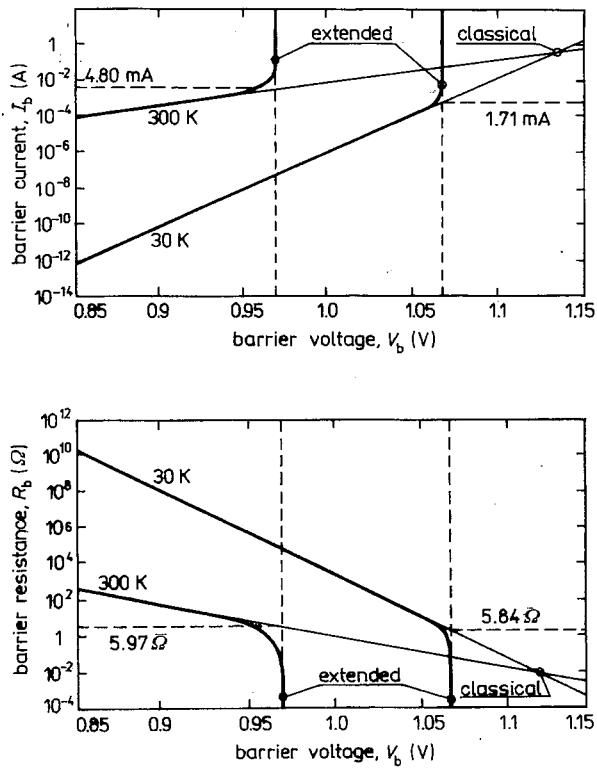


Fig. 6. Comparison between extended and classical models of a metal-semiconductor barrier at high forward bias — barrier current and incremental resistance. Parameters for 300 K (30 K) are: $I_s = 1.0 \cdot 10^{-17}$ A ($5.6 \cdot 10^{-50}$ A), $\eta = 1.11$ (3.86), $C_{b0} = 1.5$ fF (1.5 fF) and $\Phi_b = 1.04$ eV (1.07 eV)

6. REVERSE CHARACTERISTICS

According to the thermionic-emission/diffusion theory, the reverse bias current density of a Schottky diode should saturate, eqn. (25), at the value $J_b = A^{**}T^2 \exp[(-q\Phi_b)/(kT)]$. There are several causes of departure from this behaviour.

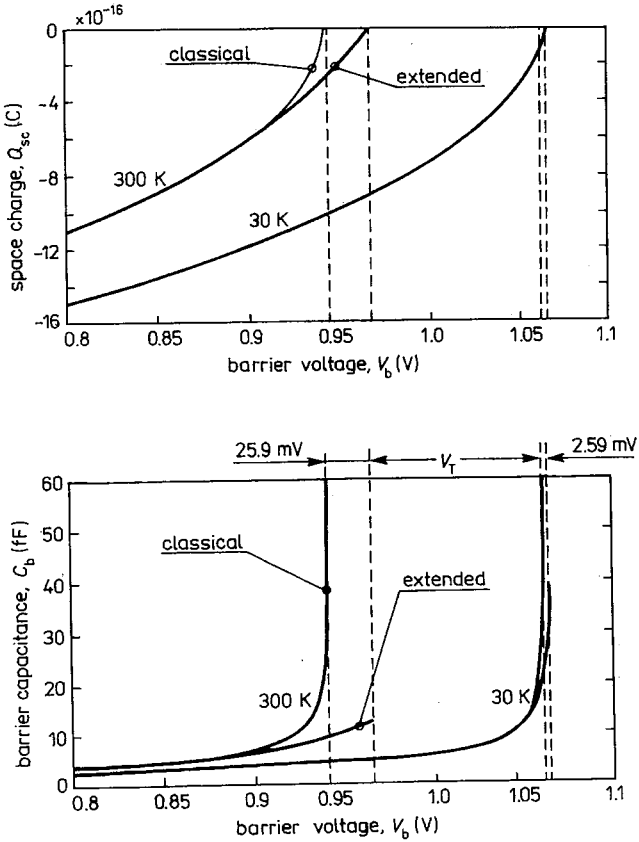


Fig. 7. Comparison between extended and classical models of a metal-semiconductor barrier at high forward bias — space-charge and incremental barrier capacitance. Parameters for 300 K (30 K) are: $I_s = 1.0 \cdot 10^{-17}$ A ($5.6 \cdot 10^{-50}$ A), $\eta = 1.11$ (3.86), $C_{b0} = 1.5$ fF (1.5 fF) and $\Phi_b = 1.04$ eV (1.07 eV)

One of them is the dependence of barrier height Φ_b on the electric field strength $\mathscr{E}$ in the barrier. Because $|\mathscr{E}_{max}|$ increases with reverse bias voltage, it follows that Φ_b decreases with increasing $|V_b|$ (V_b is negative at reverse bias). Thus the current does not saturate, but increases proportionally to $\exp[(q\Delta\Phi_b)/(kT)]$, where $\Delta\Phi_b$ is the lowering of the barrier due to the field. $\Delta\Phi_b$ resulting from the image-force effect (12) is too small to explain the barrier height lowering usually occurring in practical diodes. It is supposed that such additional barrier lowering is due to the presence of an interfacial layer, or to interdiffusion of the metal and semiconductor which can produce an effect equivalent to an interfacial layer.

Another cause for lack of reverse current saturation is the tunnelling through the barrier. Tunnelling becomes significant at lower doping, and at higher temperatures in the reverse direction than in the forward direction. This is because even at moderately high reverse bias the potential barrier becomes thin enough for electrons in the metal to tunnel into the semiconductor at energies below the top of the barrier. Tunnelling is one of the most common causes of "*soft*" reverse characteristics. It is particularly

important near to the edge of the junction, because the distortion of the field can cause a big increase in field strength. Edge effects are minimized by proper surface preparation, shape of the anode or by using diffused p-type guard rings [5, 7, 21–23].

The next cause for unsaturated reverse current is the generation of electron-hole pairs in the depletion region. Generation current is most important in high barriers and in low-lifetime materials like gallium arsenide. It is more pronounced at low temperatures than at high, because of lower activation energy than the thermionic emission. It is a common cause of lack of saturation of the reverse current in GaAs Schottky diodes.

Most practical junctions are made from epitaxial semiconductor material with a lightly doped layer of thickness t_e. Such a layered structure allows the use of a highly doped substrate to minimize resistive loses, while forming the junction on the layer doped to lower concentration optimizing junction performance. The width of the depletion layer increases with increasing the reverse bias voltage $|V_b|$. The voltage for which the depletion layer thickness w, eqn. (43), becomes equal to the epitaxial layer thickness t_e is called the "*punchthrough voltage*" V_p [2]. The junction capacitance remains almost constant above this voltage with a relatively small decrease with increasing voltage because the electric field penetrates a small distance into the heavily doped substrate. At sufficiently high reverse bias the electric field in the junction becomes large enough to generate avalanche multiplication, because of secondary and higher order ionization of holes and electrons from injected or thermally generated carriers. The ionization rates depend not only on the electric field but also on the junction temperature and the orientation of the crystal. The "*breakdown voltage*" V_{br} under the condition of punchthrough is nearly independent of the donor concentration because the field intensity is mainly determined by the thickness of the epitaxial layer and the applied voltage. The breakdown voltage is thus determined in this case by the thickness of the epitaxial layer; e.g. for GaAs $t_e \approx 0.1$ μm is sufficient to obtain a breakdown voltage of approximately 6 V while $t_e \approx 1$ μm and low doping yield $|V_{br}| \approx 30$ V [2].

The junction behaviour at reverse voltages close to the breakdown voltage, V_{br}, is well modeled by a power-law dependence of the breakdown current on voltage. If we adopt a definition, commonly accepted in experimental determination, namely breakdown voltage as a voltage yielding a reverse current of 10 μA, then breakdown current I_{br} in μA expressed as

$$I_{br}(V_b) = 10 \cdot (1 + V_{br} - V_b)^E \qquad \text{for} \quad V_b < (1 + V_{br}), \tag{50}$$

where exponent E is usually 10, provides ver good fit to experimental data in a wide variety of cases.

Comparing silicon and gallium arsenide Schottky diodes, it should be noted that owing to lower carrier mobility, silicon diodes require an extremely thin epitaxial layer for low resistive losses (series resistance). Therefore they might have lower breakdown voltages, and might be more easily damaged by electrostatic discharge or excessive RF power, and might generate noise due to avalanching on negative peaks of the RF signal cycle [2, 4].

7. DIODE SERIES IMPEDANCE

Practical diodes are fabricated by growing a thin lightly doped epitaxial layer on a low resistivity heavily doped substrate (buffer layer). The epilayer is used for the junction as is shown in Figure 1. A high quality ohmic contact (cathode) is made by metalizing surfaces of the substrate.

The series impedance consists of all impedances between the edge of the depletion region and the ohmic contact metalization. It is typically composed of three components: the impedance of the undepleted epitaxial layer, the impedance of the substrate and the ohmic contact impedance. Typically the ohmic contact impedance is quite small, both at dc and at operational frequencies [25], and can therefore be neglected.

Epitaxial layer impedance The greatest part of the series impedance is the impedance of the undepleted epitaxial layer under the junction. In *varactor* diodes the epilayer is normally made thick enough (see Figure 1.b) to contain the depletion region at high reverse barrier voltages. The optimum situation is when the entire epitaxial layer is just depleted at the onset of reverse breakdown [26–28]. A thicker epitaxial layer would increase the series resistance without any increase in breakdown voltage (which increases with reducing doping of the epilayer). A relatively thick epitaxial layer results in thick undepleted high-resistance epitaxial material under the barrier at bias voltages close to zero, but a thinner epitaxial layer would degrade the capacitance modulation (C_b is almost constant when the epilayer is punched through). In varactor applications barrier capacitance variation is crucial and may not be sacrificed to obtain lower series resistance.

In diodes designated to be used as *varistors* epitaxial layer may be thinner (see Figure 1.a) because it is nonlinear resistance that is used in frequency conversion rather than nonlinear capacitance. In this case the capacitance may be kept constant at reverse voltages. Too thick epitaxial layer in a varistor would add series resistance without any benefit; a very thick layer can also lead to excess noise because of intervalley scattering effects. Typically, an epitaxial layer thickness equal to the zero-bias depletion depth in the epitaxial material is used in varistor diodes [29–31] but even thinner layers are sometimes preferred [6, 21]. This ensures that the depletion region extends into the heavily doped semiconductor under reverse or even small forward bias.

Using a planar junction model, i.e. assuming that the diode diameter is much larger than the epilayer thickness, the epilayer series impedance at low frequency can be approximated as [6, 32]

$$R_{su}(V_b)=\frac{t_u(V_b)}{S\sigma_{e0}}=\frac{t_u(V_b)}{Sq\mu_{e0}N_{De}}, \tag{51}$$

where $t_u=w_{br}-w$ is the distance between the junction's edge w_{br} and the edge of the depleted region w, S is the area of the junction, $\sigma_{e0}=q\mu_{e0}N_{De}$ is the epilayer low-field

conductivity, μ_{e0} is the mobility in the layer and N_{De} is donor concentration in the epilayer. Because of voltage dependence of t_u, resistance R_{su} is also *voltage-dependent*. t_u can be determined if the depletion region width w is known. For voltages lower than $V_{jb}-3V_T$, considering eqn. (3) and derivations in [33], w can be expressed as

$$w=\begin{cases}\left[\dfrac{2\varepsilon_s}{qN_{De}}(V_d-V_b-V_T)\right]^{\frac{1}{2}} & \text{for } V_b\geqslant V_p\\ \left[\dfrac{2\varepsilon_s}{qN_{Db}}(V_d-V_b-V_T)+\left(1-\dfrac{N_{De}}{N_{Db}}\right)t_e^2\right]^{\frac{1}{2}} & \text{for } V_b<V_p\end{cases} \tag{52}$$

where

$$V_p=(V_d-V_T)-\frac{qN_{De}t_e^2}{2\varepsilon_s} \tag{53}$$

is the punchthrough voltage at which the depletion region thickness w becomes equal to the epitaxial layer thickness t_e and $V_T=kT/q$.
Making use of eqn. (52), the junction series resistance for low junction current density and $w_{br}>t_e$ can be written as

$$R_{su0}=\frac{1}{S}\begin{cases}\rho_{b0}(w_{br}-t_e)+\rho_{e0}(t_e-w) & \text{for } w\leqslant t_e\\ \rho_{b0}(w_{br}-w) & \text{for } w>t_e\end{cases} \tag{54}$$

where w_{br} denotes the edge of the junction (i.e., the depletion region width at the breakdown voltage V_{br} [34]), and $\rho_{e0}=1/\sigma_{e0}$ and $\rho_{b0}=1/\sigma_{b0}$ are resistives of epitaxial and buffer layers for low electrical field strength.

Carrier velocity saturation at high fields (i.e., at high current densities resulting from high bias and/or RF power applied to the diode) causes increase in the resistance of the undepleted epitaxial material [12, 35]. This increase was modeled in [35] by a dependence of the form

$$R_{su}(I_j)=R_{su0}(1+aI_j^9) \tag{55}$$

with the parameter "a" chosen empirically. Nonlinear series resistance $R_{su}(I_j)$ explains increase of the series resistance, and thus decrease of the multiplier efficiency, with increase of the RF power.

Because of non-zero effective mass, carriers exhibit inertia effects which may be modeled by an inductance expressed as

$$L_{su}(V_b)=\frac{R_{su}(V_b)}{\omega_{s,eff}}. \tag{56}$$

In an epitaxial diode, electrons enter the epilayer from the substrate with randomized net momentum. They are then accelerated by the electric field toward the anode where they are either emitted through the potential barrier of reflected back toward the substrate. On average, each electron emitted is replaced by a thermalized electron from the substrate. This is equivalent to a "*quasi-scattering*" event that tends

to randomize the net momentum of the electron distribution, thus modifying classical scattering mechanism in bulk semiconductors. The mean time between quasi-scattering events is approximated by the transit-time of the electrons through the epilayer. Resulting *effective scattering frequency* can be approximated by [36]

$$\omega_{s,eff} \approx \frac{q}{m^* \mu_{e0}} + \frac{v_d}{t_e}, \tag{57}$$

where m^* is the effective carrier mass and v_d is the mean drift velocity of electrons. The epilayer's scattering frequency is actually a function of transit time, which can be decreased by ballistic transport effects [37–39].

Another effect which should be considered when analyzing current flow in undepleted epilayer is the displacement current, which can be accounted for by introducing capacitance shunting the series connected R_{su} and L_{su}

$$C_{su}(V_b) = \frac{\varepsilon_s S}{t_u(V_b)}. \tag{58}$$

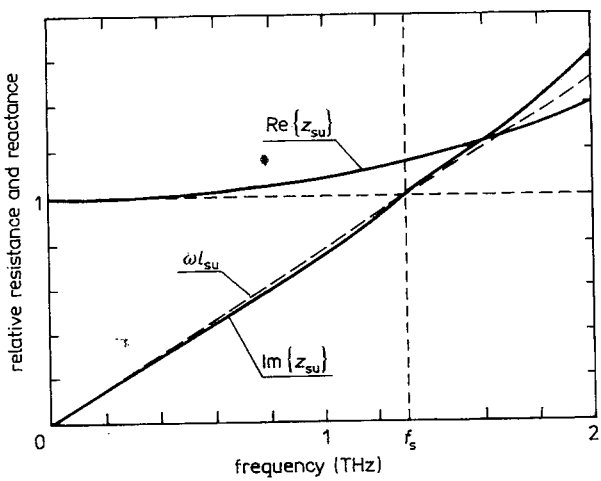


Fig. 8. Comparison of the undepleted epilayer relative impedance $Z_{su}/R_{su} = Re\{z_{su}\} + j\mathrm{Im}\{z_{su}\}$ with the model consisting of series connected voltage-dependent resistance $R_{su}(V_b)$ and voltage-dependent inductance $L_{su}(V_b)$

$N_{De} = 2 \cdot 10^{17}$ cm^{-3}
$v_d = 2 \cdot 10^5$ m/s
$T = 300$ K

$f_{s,eff} = 1.31$ THz
$f_p = 5.00$ THz
$f_d = 19.0$ THz

At relatively "*low*" frequencies (below ~1 THz) and small drive powers, reactance of $C_{su}(V_b)$ is much larger then the resistance $R_{su}(V_b)$. Most of the RF current will flow through the resistance and the effect of displacement current will be negligible. Therefore, at lower frequencies, series junction impedance may be modeled by *voltage-dependent* resistance and *voltage-dependent* inductance, as it is shown in Figure

8. However, if frequency and/or drive level increase the situation is different. According to eqn. (55) the resistance R_{su} increases with the RF drive, and C_{su} shunts larger fraction of the current. Increasing frequency also makes the current divider worse because of decreasing capacitive reactance. In addition, a plasma resonanse between the carrier inertia (inductive) and dielectric relaxation (capacitive) occurs at very high frequency seriously degrading performance of the junction. Eventually the situation would be reached when junction behaves like series connection of barrier and undepleted epilayer capacitances shunting barrier and epilayer resistances. The net capacitance would go down and become constant. As a result, the device would become useless above the plasma frequency [36, 40]. Fortunately plasma resonanse in the epilayer occurs at approximately 3 to 5 THz (see Figure 9.a), and can be even increased by increasing doping density of the epilayer [29, 31, 41].

Substrate impedance The mathematical analysis of the spreading impedance of a Schottky diode which recognized the importance of the skin effect was presented in [42]. Electron inertia and dielectric effects were first considered in analysis given in [43, 44]. Extended analysis of [33] resulted in a series approximation of the substrate spreading impedance that considers these effects and is highly accurate at frequencies up to the terahertz range. The above analyses assume a simplified diode structure with uniform conductivity in the semiconductor substrate material and ideal ohmic contact surrounding the diode chip. Accurate analytic solution can not be readily obtained for more realistic and complicated configurations and computer simulations based on finite element analysis are required to obtain results for particular diode configurations [6, 45–47]. The qualitative description of the effects occurring in the substrate may be based on analytical expressions and be supported by conclusions drawn from computer simulations.

The dc resistance of the substrate (buffer layer) is given by [42]

$$R_{sb}=\frac{1}{2d\sigma_{b0}}=\frac{1}{2dq\mu_{b0}N_{Db}} \tag{59}$$

where d is the anode diameter, μ_{b0} is the electron mobility in the substrtate doped to N_{Db} and σ_{b0} is the conductivity of the buffer layer (substrate). R_{sb} is often called the spreading resistance, because the majority of the resistance occurs near the surface of the substrate adjacent to the epilayer. Current is confined here to a small area and spreads across a larger area as it moves further into the substrate.

The dc series resistance does not represent an accurate picture of the substrate because several important effects cause the substrate impedance to increase with frequency. Most important is the skin effect which constrains the high-frequency current to flow along the surface of the substrate, greatly reducing the cross-sectional area. This not only increases the real pat of the series impedance, but also causes a substantial inductive component. For large substrates Z_{sb} is given by

$$Z_{sb}=Z_{spr}+Z_{skin}=\frac{1}{\pi\sigma_{b0}d}\tan^{-1}\left(\frac{2b}{d}\right)+\frac{1+j}{2\pi\sigma_{b0}\delta_s}\ln\left(\frac{2b}{d}\right), \tag{60}$$

where δ_s is the skin depth in the substrate given by

$$\delta_s = \left(\frac{2}{\omega \mu_0 \sigma_{b0}}\right)^{\frac{1}{2}}, \tag{61}$$

b is the distance from the anode to the ohmic contact along the chip surface and μ_o is the permeability of the semiconductor. The first term in (60) is due to the substrate material near the epilayer and for $b \gg d$ reduces to (59). The second term vanishes at dc and is due to the skin effect which forces the current to flow along the substrate surface.

While the skin effect causes a steady increase in impedance with frequency, the plasma resonance causes a sharp rise in substrate impedance near the plasma frequency and then reduction of the impedance at higher frequencies. In bulk material charge carrier inertia causes a delay in the response of the electron velocity distribution to a change in the direction of the electric field. Carrier inertia becomes important when the frequency of the field approaches the mean *scattering frequency* (the reciprocal of the mean *scattering time*). At lower frequencies the carrier inertia will be dissipated quickly by random scattering. However, as the frequency increases toward the scattering frequency the scattering rate will be too low to dissipate the inertia and inductive effects will be seen in the device's characteristics. The mean scattering frequency is approximated by

$$\omega_s \approx \frac{q}{m^* \mu_{b0}}. \tag{62}$$

The current flowing through the substrate (which is not perfect conductor) is composed of conductive and displacement components. As frequency is increased the displacement component becomes more pronounced and above *dielectric relaxation frequency* dominates over conductive component. The dielectric relaxation frequency is given by

$$\omega_d \approx \frac{\sigma_{b0}}{\varepsilon_s}. \tag{63}$$

A resonance between the carrier inertia and dielectric relaxation occurs at the plasma resonance frequency

$$\omega_p = (\omega_s \omega_d)^{\frac{1}{2}} = \left(\frac{q\sigma_{b0}}{m^* \mu_{b0} \varepsilon_s}\right)^{\frac{1}{2}} = \left(\frac{q^2 N_{Db}}{m^* \varepsilon_s}\right)^{\frac{1}{2}}. \tag{64}$$

Considering the plasma resonance yields the substrate impedance $Z_{sb}(\omega) = Z_{spr}(\omega) + Z_{skin}(\omega)$ approximated by the expressions

$$Z_{spr}(\omega) = \frac{1}{\pi \sigma_{b0} d} \tan^{-1}\left(\frac{2b}{d}\right)\left[\frac{1}{1 + j(\omega/\omega_s)} + j\frac{\omega}{\omega_d}\right]^{-1} \tag{65}$$

and

$$Z_{skin}(\omega)=\frac{\ln(2b/d)}{2\pi}\left(\frac{j\omega\mu_o}{\sigma_{b0}}\right)^{\frac{1}{2}}\left[\frac{1}{1+j(\omega/\omega_s)}+j\frac{\omega}{\omega_d}\right]^{-\frac{1}{2}}. \tag{66}$$

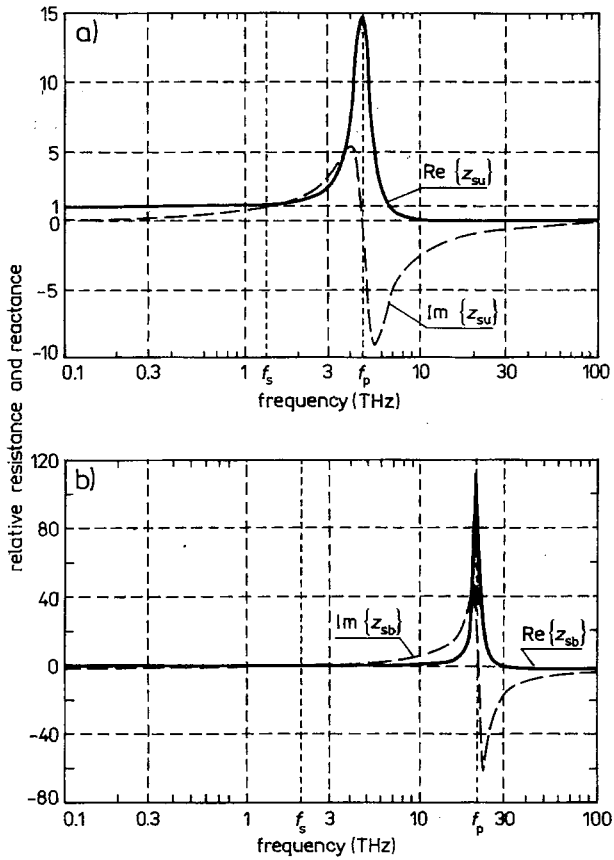


Fig. 9. Plasma resonance effect in a) epitaxial layer and b) substrate of a Schottky diode

$N_{De}=2\cdot10^{17}$ cm^{-3}
$v_d=2\cdot10^5$ m/s
$T=300$ K
$f_{s,eff}=1.31$ THz
$f_p=5.00$ THz
$f_d=19.0$ THz

$N_{Db}=5\cdot10^{18}$ cm^{-3}
$T=300$ K
$f_s=2.05$ THz
$f_p=21.7$ THz
$f_d=231$ THz

Plasma resonance is dependent on doping density, eqn. (64), and because of much higher substrate it occurs in the substrate at much higher frequency than in the epitaxial layer. State-of-the-art diodes have substrate doping density equal to $5\cdot10^{18}$ cm^{-3} or greater and plasma resonanse frequency is higher than 18 THz. Therefore the effect of plasma resonance in the bulk of substrate semiconductor may be neglected in diodes operating at frequencies below few THz. However, the skin effect may not be neglected for operating frequencies above $\approx$50 GHz (see Figure 10). At short

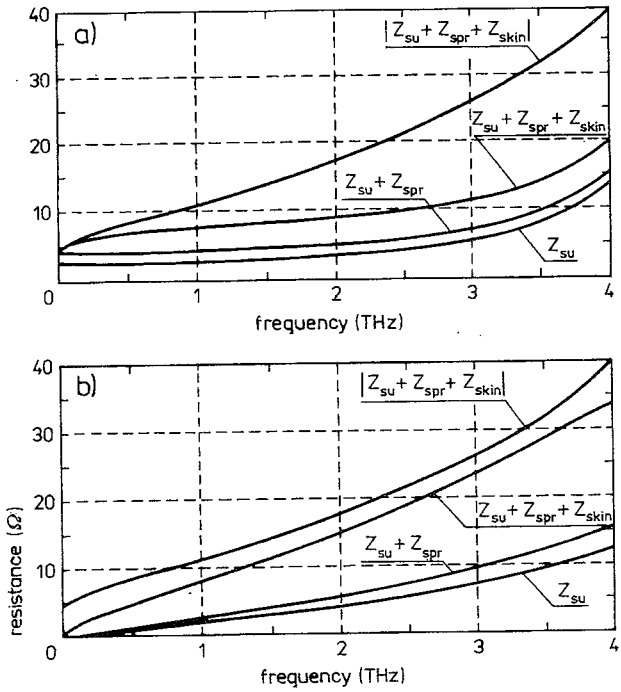


Fig. 10. Frequency dependence of the series impedance $Z_s=Z_{su}+Z_{spr}+Z_{skin}$ of a Schottky diode
$N_{De}=2\cdot10^{17}$ cm^{-3} $a=2$ μm
$N_{Db}=5.10^{18}$ cm^{-3} $b=250$μm
$t_e=0.1$ μm $T=300$ K

millimeter waves the skin effect could increase the series resistance by as much as a factor of two.

It is difficult to theoretically estimate spreading resistance and skin effect impedance because they are strongly dependent on diode structure. In some cases computer simulations are necessary. But general suggestions can be drawn on how to reduce the parasitic series impedance of the diode. Series resistance is inversely proportional to the junction area and in many cases limits the useful diode size to 1.0 to 2.0 μm (except at THz frequencies where anode diameter ≈0.5 μm is required). The substrate must be as heavily doped as possible to reduce its resistivity and avoid plasma resonance. A typical dot-matrix diode chip structure has a long path for current flowing due to skin effect along the upper surface and sides of the substrate. This can be minimized in several ways. In notch-front diodes [2, 48] sides of diode chips are metalized. Because of metalized sides the chip can be mounted on its side in microstrip or suspended-substrate stripline circuits creating a very short path between the epilayer and the mounting surface. A step further is the metalization deposited also on the upper surface of the n^+ substrate to within 10 to 20 μm of the junction periphery [49]. A similar technique is used in planar Schottky diodes optimized for

operation in short millimeter-wave range [25, 27, 28, 30, 50]. The use of extra thin substrate in a membrane form [46] also greatly reduces current path thus reducing substrate series resistance. Another way to reduce edge effects which increase series resistance is to use a diode anode having a large periphery relative to its area (i.e., having shape other than a circle) [2, 6, 23, 45]. Such diodes are more difficult to fabricate and whisker while their RF performance is not significantly better than that of more conventional diodes.

In conclusion it should be stated that the impedance of the substrate is usually smaller and less significant than the impedance of the undepleted epitaxial material. Substrate impedance is *not* dependent on barrier voltage and therefore it is a reasonable and justified common practice to include this impedance in the linear part of a circuit when analyzing frequency converting circuits.

8. DIODE NOISE

All currents and voltages in the dc biased junction of a diode are in fact random stochastic signals of the form

$$x(t)=x_0+x_n(t), \tag{67}$$

where x_0 is the average value (constant component) of the signal and $x_n(t)$ is its noise component having zero average value. Although noise is a time-domain process it can be represented also in frequency domain. Frequency description of the noise is very convenient and highly desirable in circuit analysis.

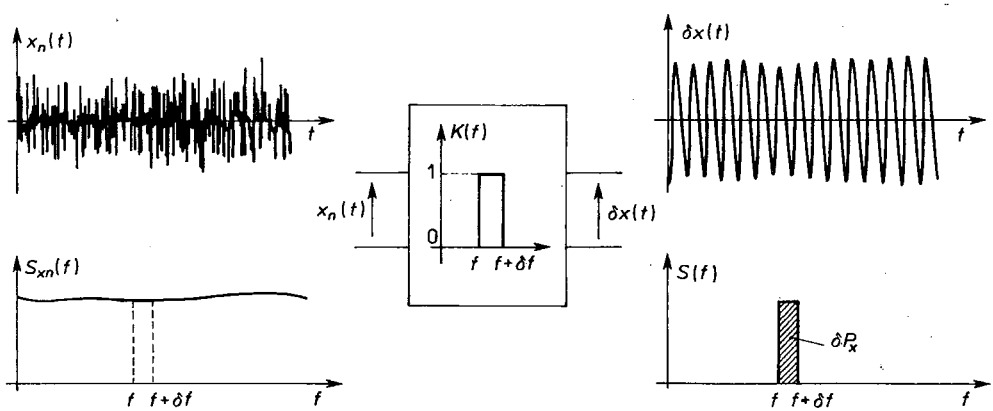


Fig. 11. Pseudosinusoidal narrow-band noise component $\delta x(t)$ and its power δP_x resulting from filtering nois process $x(t)$ characterized by self-power spectral density $S_{xn}(f)$. (plots are not to scale)

Filtering the noise signal $x_n(t)$ through an ideal band-pass filter of the band $(f, f+\delta f)$ yields pseudosinusoidal narrow-band noise

$$\delta x(t)=\sqrt{2}\,\mathrm{Re}\{\delta \mathrm{X} e^{j2\pi f t}\}, \tag{68}$$

where $\delta \mathrm{X}= \mid \delta \mathrm{X} \mid e^{j\varphi}$ is the complex root-mean-square (r.m.s) value with random magnitude and random phase. Power of this signal (mean-square value) is equal to the average value of the square of magnitude of the complex r.m.s. value

$$\delta P_x=\langle \mid \delta \mathrm{X} \mid ^2\rangle=\langle \delta \mathrm{X} \delta \mathrm{X}^*\rangle. \tag{69}$$

Function

$$S_x(f)=\lim_{\delta f\to 0}\frac{\delta P_x(f,\delta f)}{\delta f}=\lim_{\delta f\to 0}\frac{\langle \delta X \delta X^*\rangle}{\delta f} \tag{70}$$

is called *self-power spectral density* of the signal. Signal power spectral density (power per unit bandwidth) $S_x(f)$ is the basic and fundamental way for describing noise signals in the frequency domain. Random complex r.m.s. values are very convenient in the analysis of linear noisy circuits because they are governed by the same rules as sinusoidal signals in the steady-state analysis of linear circuits.

Available power spectral density of a voltage noise source is equal to

$$S_{av}(f)=\frac{S_{vg}(f)}{4\mathrm{Re}\{Z_g(f)\}}, \tag{71}$$

where $Z_g(f)$ is the source internal impedance and $v_g(t)$ is its open-circuit voltage described by self-power spectral density $S_{vg}(f)$. Available power of this source in frequency band $(f, f+\delta f)$ is thus

$$\delta P_{av}=S_{av}\delta f=\frac{S_{vg}(f)\delta f}{4\mathrm{Re}\{Z_g(f)\}}=\frac{\langle \delta V_g \delta V_g^*\rangle}{4\mathrm{Re}\{Z_g(f)\}} \tag{72}$$

And for current noise source ($Y_g(f)$, $i_g(t)$ similar form is

$$\delta P_{av}=\frac{\langle \delta I_g \delta I_g^*\rangle}{4\mathrm{Re}\{Y_g(f)\}}. \tag{73}$$

At low current levels, noise engendered in Schottky diodes is caused by fluctuations of the number of electrons crossing the barrier (the shot noise) and their velocity fluctuations (the diffusion or thermal noise).

Thermal noise is present in any power-dissipative medium having a temperature above absolute zero. It arises from continual random motion of thermally agitated electrons which impact on molecules of the lattice. Hance, in thermal equilibrium, there is a cloud of chaotically moving („*velocity*" fluctuations) electric charges. Since charage in motion constitutes an electric current, there must appear some electrical signal across the terminals of a device. Available power spectral density (i. e., power per unit bandwidth) of this thermal noise signal is dependent both on temperature and frequency [52]:

$$S_{th}(f)=\mathrm{h}f\cdot\left[\frac{1}{2}+\frac{1}{\exp((\mathrm{h}f)/(\mathrm{k}T))-1}\right] \tag{74}$$

where h is the Planck's constant and k is the Boltzmann's constant. If we assume that $(\mathrm{h}f)/(\mathrm{k}T)\ll 1$ then the above expression reduces to

$$S_{th}(f)=\mathrm{k}T=\mathrm{const}(f), \tag{75}$$

which means that spectral power density of the thermal noise does not depend on frequency ("*white*" noise) for frequencies below a few THz and temperatures above cryogenic conditions. The above assumption not always is acceptable. If we substitute k and h constants then the white noise condition becomes $f/T\ll 2.08\cdot 10^{10}\ \mathrm{s^{-1}K^{-1}}$ which for room temperature ($T=290$ K) limits considerations to $f\ll 6$ THz or, equivalently, at frequency of 1 THz to temperatures $T\gg 48$ K. Therefore, even though the quantum correction might be small, its omission must be justified carefully.

Available thermal noise power is given by

$$\delta P_{th}=S_{th}(f)\delta f\approx \mathrm{k}T\delta f \tag{76}$$

which leads to the equivalent noise source of the thermal noise generated in the diode's serises resistance $\mathrm{Re}\{Z_s\}$ in incremental bandwidth δf expressed as

$$\langle\delta E_{th}\delta E_{th}^{*}\rangle=4\mathrm{k}T\,\mathrm{Re}\{Z_s\}\delta f. \tag{77}$$

The fact that the temperature at which a noise process occurs defines its available power spectral density leads to the noise temperature. The *noise temperature* of a single-port device is defined as the temperature at which available thermal-noise power is equal to the amount of available noise power from the output of the device

$$T_n=\frac{S_n}{\mathrm{k}}. \tag{78}$$

If S_n is $4\cdot 10^{-21}\ \mathrm{WHz^{-1}}$, then $T_o=290$ K, which is the temperature adopted as a standard reference condition. Note that the noise temperature is in general a function of frequency.

Recalling eqn. (7.75), from the definiton (7.78) the noise temperature of a thermal noise is simply $T_{th}=T$.

The shot noise arises from the fluctuations of the number of electrons crossing the potential barrier of the junction. The current consists of a series of random pulses ("*shots*") that occur as each electron crosses the barrier. The averge number of such pulses in every second is constant and proportional to the dc current. But the instantaneous current vary with time and the resulting fluctuations are a noise process. Its mean-squared magnitude current observed in δf bandwidth is proportional to the dc current

$$\langle\delta I_{sh}\delta I_{sh}^{*}\rangle=2\mathrm{q}I_{eq}\delta f, \tag{79}$$

where I_{eq} is the sum of the magniutudes of the forward I_b+I_S and reverse I_S currents in the barrier. This expression assumes that the transit-time effects are negligible in the frequency range of interest.

Current pulse caused by a single electron is very short which results in a very broad spectrum. Therefore spectral density of the shot-noise available power is not dependent on frequency. Considering barrier incremental resistance R_b given by (49) and $V_b > 3\eta V_T$, available power spectral density of the shot noise can be expressed as

$$S_{sh}=\frac{2qI_{eq}}{4R_b^{-1}}=\frac{q\eta V_T}{2}\tanh\left[\frac{(V_d-V_b)}{\eta V_T}\right] \tag{80}$$

and because $V_T=kT/q$ then from the definition (78) the noise temperature of the shot noise generated in the barrier is

$$T_{sh}=\frac{\eta T}{2}\tanh\left[\frac{(V_d-V_b)}{\eta V_T}\right], \tag{81}$$

which for bias voltage V_b lower than $V_d-3\eta V_T$ reduces to a classical model, e.g. [53], and the shot-noise temperature can be written as

$$T_{sh}=\frac{1}{2}\eta T=\frac{1}{2}\theta, \tag{82}$$

where junction ideality factor η and barrier effective temperature θ are given by (34) and (40), respectively. Let us notice that both η and θ are dependent on the mechanism of carrier transport through the barrier. For pure thermionic emission $\eta \approx 1$ and T_{sh} is approximately half of the barrier physical temperature. For field emission (i.e., tunnelling through the barrier) $\eta \approx qE_{00}/kT$ and $T_{sh} \approx \theta_F/2 = qE_{00}/2k$ is not dependent on temperature and, on cooling, sets higher noise limit than that which could have been achieved if tunnelling of electrons was not present. For example, in GaAs diode with doping $N_{De} \approx 1 \cdot 10^{17}$ cm^{-3} tunnelling gives $T_{sh}=36.2$ K instead of 10 K which would have been due to thermionic emission at barrier temperature of 20 K. Tunnelling is a relatively noisy mechanism and is tried to be avoided by, for example, proper epilayer doping for particular application – as it was discussed in preceding sections.

At high current densities other mechanisms are the source of a so-called "*excess noise*". They are voltage or current dependent and raise the diode noise temperature above that resulting from the classical model, i.e., model assuming electron temperature equal to the lattice temperature and considering the shot and thermal noise only. These additional noise mechanisms are related to excess fluctuations of the number and velocities of electrons traversing the junction [16]. The excess electron velocity fluctuations are mainly due to the excess electron temperature ("*hot*" electrons). The excess fluctuations of the number of electrons (electron density fluctuations) are attributed to intervalley scattering and trapping of electrons in the undepleted epilayer and in the vicinity of the metal-semiconductor interface.

There are also additional mechanisms which affect the noise performance of the diode. Among them are: the formation of microclusters at the metal-semiconductor interface [10, 54], the graininess of the junction [16] (which can occur due to the relatively small number of dopants in the thin epitaxial layer), stresses at the $GaAs-SiO_2$ interface [7] and defects due to surface preparation [5]. These effects can be avoided or reduced to a very small (negligible) level by proper device technology and/or by eliminating diodes with "*strange*" and not "*well-behaved*" noise characteristics. Therefore we shall limit our considerations to those mechanisms which are inevitable in diode operation or which can not be eliminated. Besides shot and thermal noise presented above, the most important noise processes considered here are generation of hot electrons, intervalley scattering, trapping of electrons and the flicker (or 1/f) noise.

Hot electron noise is generated in the undepleted part of the epilayer if the electric field is high enough [12, 15, 55–57]. In thermodynamic equilibrium the free electrons existing in a semiconductor material have the same average temperature as the crystal lattice and the average kinetic energy of electrons is kT. Electrons are randomly scattered which results in zero net velocity of the electron concentration. In non-thermodynamic equilibrium, when an electric field is applied and/or net current flows, a net electron velocity is not zero. The electron gains an amount of energy from the electric field during a mean free-time. At high fields increase of energy with respect to kT can be significant and the electron energy distribution can no longer be described by a Maxwell–Boltzmann distribution function.

Applying the above to a Schottky diode it turns out that electrons which are accelerated through the undepleted epitaxial layer have significantly higher average temperature than the lattice, if the magnitude of electric field in the undepleted epitaxial region is high, i.e. the current density is large. Heating of the electron energy distribution causes that the noise exceeds that predicted by (75) for the pure thermal noise.

First order models (for fields $< \approx 3$ kV/cm) of the non-equilibrium electron distribution [13, 24, 58] make use of the fact that the average electron in the epilayer must gain energy from the electric field at the same rate it loses energy to the lattice. Equating the two rates yields the electron distribution temperature as

$$T_e = T + K_{he} I_j^2 \tag{83}$$

and the noise factor K_{he} given by

$$K_{he} = \frac{2\tau_e}{3kq\mu_e N_{De}^2 S^2}, \tag{84}$$

where I_j is the current flowing through the undepleted region of the epilayer, τ_e is the average energy relaxation time of the electrons, μ_e is the mobility of electrons, N_{De} is the epilayer doping concentration and S is the diode area.

The excess noise temperature due to the hot electron noise, i.e., noise temperature increase above thermal noise temperature $T_{th} = T$, is the difference between T_e and T_{th}

$$T_{he} = K_{he} I_j^2 \tag{85}$$

Intervalley scattering occurs in GaAs if the electric field $\mathscr{E}$ in the epilayer is high enough to accelerate electrons to energies higher than 0.31 eV (the energy separation between the central high-mobility Γ valley and the satellite low-mobility L valley). At such high electron energies, the probability is high that electrons are transferred from the Γ valley to the L valley. For bulk GaAs at room temperature, this occurs at an electric field near and above the intervalley scattering threshold field 3.2 kV/cm. The transfer of electrons eliminates the hottest electrons from the conduction band and in result the rate of increase of the electron temperature T_e is slowed near and above threshold field, as compared to the case of the hot electrons generation [58, 59]. Eventually, scattering of the hottest electrons to the low-mobility valley limits the electron temperature to approximately 1200 K.

Transfer of electrons to the upper valley is a random process and as such results in an intervalley scattering noise. Because of the energy threshold in electron intervalley transfer, the noise temperature T_{iv} describing the intervalley scattering noise increases abruptly near the threshold field (the threshold current):

$$T_{iv} = \frac{q\mu_e \tau_2}{k[1+(\omega\tau_0)^2]} \cdot \frac{p(1-p)^2 \mathscr{E}^2}{(1-p) - \mathscr{E}\dfrac{dp}{d\mathscr{E}}}, \tag{86}$$

which, equivalently, can be expressed as

$$T_{iv} = K_{iv} \frac{1}{1+(\omega\tau_0)^2} \cdot \frac{p(1-p)^2 I_j^2}{(1-p) - \mathscr{E}\dfrac{dp}{d\mathscr{E}}} \tag{87}$$

with the noise factor K_{iv} given by

$$K_{iv} = \frac{\tau_2}{kq\mu_e N_{De}^2 S^2}, \tag{88}$$

where μ_e is the central valley mobility, τ_2 is the upper valley mean lifetime (1.8 ps for GaAs), and p is the population probability in the satellite valley:

$$p(\mathscr{E}) = \frac{\tau_2}{\tau_1 + \tau_2} \tag{89}$$

where τ_1 is the electron mean lifetime in the central valley. τ_0 is the characteristic time constant

$$\tau_0(\mathscr{E}) = \frac{1}{1/\tau_1 + 1/\tau_2}, \tag{90}$$

which below the threshold field (when $\tau_2 \gg \tau_2$) reduces to $\tau_0 \approx \tau_2$. τ_0 then corresponds to a characteristic frequency of the order 90 GHz. Above this frequency the intervalley scattering noise disappears.

The above given expressions describe the intervalley scattering in bulk GaAs. In thin epilayer of a Schottky diode electrons might not gain enough energy from electric field in their mean free path and intervalley scattering might be reduced. Both theoretical calculations and measurements confirm this expectation [58] — equivalent noise temperature $T_{he}+T_{iv}$ corresponds well to eqn. (85) which means that the intervalley scattering noise in thin epilayer seems to be reduced and is effectively negligible as compared to the bulk case. In silicon diodes intervalley scattering does not occur at practical electric field strengths because the energy separation between valleys is greater than 1 eV.

Trapping of electrons is a mechanism which leads to a fluctuation of the number of electrons flowing through the diode. This results in an additional modulation-type noise whose spectrum can extend even to microwave frequencies. Density of traps in the vicinity of the junction interface is usually much higher than their density in the epilayer. Traps located at the interface and in the space charge region will affect the height of the potential barrier and the barrier resistance and noise caused by them will add to the shot noise generated in the barrier. Noise due to traps in the undepleted epilayer will add to the thermal, hot-electron and intervalley scattering noise generated in that region of the diode. At frequencies comparable to the reciprocal value of the trap lifetime, trapping noise will decrease, levelling off at lower and higher frequencies [10, 55].

Scattering by the traps causes fluctuations of the number of electrons in the conduction band whose spectral density S_{nm} is approximately equal to the sum of contributions of different traps. If we assume that trapping in a specific region of the diode can be approximated by one time constant τ_{tm}, then the spectral density of fluctuations of the number of carriers caused by these traps $n_m(t)$ is

$$S_{nm}=\frac{\langle \delta n_m \delta n_m^* \rangle}{\delta f}=\alpha_m n_{Tm}\frac{\tau_{tm}}{1+(\omega\tau_{tm})^2}, \tag{91}$$

where n_{Tm} is the number of traps in the considered region and α_m is a constant adjusted to fit the experimental data; different α_m but the same trap time constant τ_{te} are expected in the depleted and undepleted epilayer.

The self-power spectral density of the short-circuit noise current associated with $n_m(t)$ is equal to

$$S_{im}=\frac{\langle \mid q\bar{v}\delta n_m \mid ^2 \rangle}{\delta f}=(q\bar{v})^2 S_{nm}=\frac{I^2}{n^2V^2}S_{nm}, \tag{92}$$

where $\bar{v}$ is the average velocity of electrons and nV is their average number, equal to $N_{De}V$ when the diode operates in low-injection regime, with N_{De} being the donor density in the epilayer and V being the volume of the considered region of the diode.

The excess noise temperature caused by traps located in the vicinity of the junction *interface* ($m=i$) is

$$T_{ti}=\frac{S_{ii}}{4k}R_b=\frac{\eta T}{2}\tanh\left[\frac{(V_d-V_b)}{\eta V_T}\right]\cdot K_{ti}\frac{1}{1+(\omega\tau_{ti})^2}\cdot I_b \tag{93}$$

with the noise factor

$$K_{ti}=\frac{\alpha_i N_{Ti}\tau_{ti}}{2qN_{De}^2V}, \tag{94}$$

where N_{Ti} is the surface density of traps at the interface.

Similarly, the excess noise temperature caused by traps in the *depleted* region of the epilayer ($m=b$) is

$$T_{tb}=\frac{S_{ib}}{4k}R_b=\frac{\eta T}{2}\tanh\left[\frac{(V_d-V_b)}{\eta V_T}\right]\cdot K_{tb}\frac{1}{1+(\omega\tau_{te})^2}\cdot I_b \tag{95}$$

with the noise factor

$$K_{tb}=\frac{\alpha_b N_{Te}\tau_{te}w(V_b)}{2qN_{De}^2V} \tag{96}$$

where N_{Te} is the density of traps in the epilayer and the depletion region width $w(V_b)$ is given by (43). Because usually $n_{Ti}=N_{Ti}S\gg n_{Tw}=N_{Te}Sw(V_b)$ and the depletion region width $w(V_b)$ is small for high forward bias, then $T_{ti}+T_{tb}\approx T_{ti}$, i.e., trapping noise in the depleted region may be neglected.

The noise temperature of the excess noise due to traps in the *undepleted* epilayer ($m=u$) is

$$T_{tu}=\frac{S_{iu}}{4k}R_{su}=\frac{\alpha_u N_{Te}St_u}{4kN_{De}^2V^2}\cdot\frac{\tau_{te}}{1+(\omega\tau_{te})^2}\cdot R_{su}\cdot I_j^2. \tag{97}$$

Substituting the resistance of undepleted epilayer R_{su} given by eqn. (51) yields the noise temperature of the trapping noise in the undepleted epilayer expressed as

$$T_{tu}=K_{tu}\frac{1}{1+(\omega\tau_{te})^2}\cdot I_j^2 \tag{98}$$

with the noise factor

$$K_{tu}=\frac{\alpha_u N_{Te}\tau_{te}t_u^2(V_b)}{4kq\mu_e N_{De}^3V^2}, \tag{99}$$

where $t_u(V_b)=w_{br}-w(V_b)$ and $w(V_b)$ is given by (43).

The excess noise caused by traps loading and unloading strongly depends on the quality of the material and technology used to make diodes. It varies both in magnitude and in trap time constant (τ_{tm} from 66 ps to 450 ps were reported in [16, 24, 58]) from batch to batch of diodes. The trap noise can be reduced by using advanced growth and processing techniques for fabricating the epitaxial layer and metal-semiconductor contact. In particular, it can be neglected in the undepleted region

of the epilayer [60] where its level is usually much lower than the hot-electron noise. In good quality diodes, density of traps in the vicinity of the junction interface is also low and the noise due to these traps is negligible comparing to the shot noise generated in the barrier.

Trapping is believed to be also the major mechanism of another low-frequency excess noise process usually called *flicker* (or $1/f$ *noise*). This noise also depends strongly on the quality of epitaxially grown semiconductor materials and quality of the surfaces making metal-semiconductor contacts. But, in contrast to the above discussed trapping noise, it is inherent to diode operation and can not be eliminated or reduced to negligible level at low frequencies. A comprehensive overview of $1/f$ noise sources in semiconductors, semiconductor devices and collision-free devices is presented in [61].

One of the popular models of $1/f$ noise in Schottky diodes considers that filling and emptying of traps at or near the metal-semiconductor interface results in modulation of the barrier height and so leads to current fluctuations. Associated with each trap is a characteristic time constant. Because the distribution of these time constants is wide, then the resulting power spectrum has a $1/f$ component at low frequencies. Furthermore, it may be shown that the magnitude of the $1/f$ noise is proportional to the density of traps and surface states at the interface and inversely proportional to the carrier effective mass in semiconductor (therefore it is higher in GaAs than in Si diodes).

The flicker noise is usually characterized by the self-power spectral density of the short-circuit noise current or, equivalently, by the excess noise temperature of the barrier incremental resistance approximated by

$$T_{fn}=\frac{S_{fn}}{4kR_b^{-1}}=K_{fn}\frac{I_b^{\beta}}{f^{\alpha}}R_b(V_b), \tag{100}$$

where noise factor K_{fn} and constants α and β are determined experimentally — usually $\alpha\approx 1$ and $\beta\approx 2$. In measurements and for comparison purposes the flicker noise is often characterized by a "*corner frequency*" which is the frequency at which the $1/f$ noise lowers down to a device "*noise floor*" (i.e. shot and thermal noise). For example, it is ≈ 100 kHz for low barrier Si Schottky diodes and ≈ 500 kHz for high frequency GaAs Schottky diodes.

The above discussed noise processes are independent on each other and hence are uncorrelated. Therefore the noise temperature T_d of a dc biased diode can be calculated by summing up the contributions from the different parts of the diode [62], as it is shown at Figure 12.

Assuming for simplicity that the impedance of undepleted epilayer is voltage independent[1], the diode noise temperature can be expressed as

[1] It has been shown in [63] that, although fundamentally wrong, this assumption in the case of metal-semiconductor junctions gives an error not exceeding $\sim 1-2\%$. It is not necessarily the case for other nonlinear semiconductor devices.

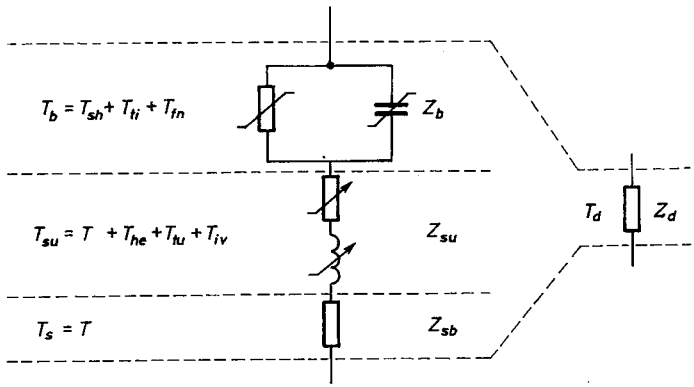


Fig. 12. Noise temperatures representing noise processes localized in different parts of the diode and resulting equivalent noise temperature of the diode

$$T_d = \frac{T_b \cdot \mathrm{Re}\{Z_b\} + T_u \cdot \mathrm{Re}\{Z_{su}\} + T_s \cdot \mathrm{Re}\{Z_{sb} + Z_{sc}\}}{\mathrm{Re}\{Z_b + Z_{su} + Z_{sb} + Z_{sc}\}}. \tag{101}$$

The first term represents noise generated in the barrier, which is modeled by the impedance $Z_b(\omega, I_j)$ and thus $\mathrm{Re}\{Z_b\} = R_b/[1 + (\omega R_b C_b)^2]$ with the barrier incremental resistance $R_b(I_b)$ given by (49) and the depletion layer capacitance $C_b(V_b)$ by (46). The noise temperature T_b results from the shot noise, noise from traps at the junction interface and the flicker noise

$$T_b(\omega, I_b) = T_{sh}(I_b) + T_{ti}(\omega, I_b) + T_{fn}(\omega, I_b). \tag{102}$$

The second term in (101) results from several different processes in the undepleted epilayer and is in general contributed to by the thermal, hot-electron, intervalley scattering and trapping noise. In thin epilayer of a good quality Schottky diode intervalley scattering and trapping noise are negligible. The impedance of the undepleted epilayer Z_{su} is modeled as the series connection of resistance $R_{su}(V_b)$ given by (54) and (51) and inductance $L_{su}(V_b)$ expressed by (56). The noise temperature T_u is the sum of noise temperatures characterizing particular noise processes occurring in the undepleted region of the epilayer, and for practical diodes is equal to

$$T_u(\omega, I_j) = T + T_{he}(I_j) \tag{103}$$

The diode noise is augmented by the thermal noise generated in the substrate (buffer layer) and diode contacts, which is characterized by the thermal noise temperature $T_s = T_{th} = T$ with the impedance $Z_{sb}(\omega) = Z_{spr}(\omega) + Z_{skin}(\omega)$ given by (65) and (66).

Substituting the above expressions into (101) gives the noise temperature of the dc biased diode as

$$T_d = \frac{(T_{sh}+T_{ti}+T_{fn})\cdot R_b/[1+(\omega R_b C_b)^2]+(T+T_{he})\cdot R_{su}+T\cdot \mathrm{Re}\{Z_{spr}+Z_{skin}+Z_{sc}\}}{R_b/[1+(\omega R_b C_b)^2]+R_{su}+\mathrm{Re}\{Z_{spr}+Z_{skin}+Z_{sc}\}}. \tag{104}$$

Noise temperatures characterizing noise generated in the diode and the impedances of different parts of the diode are bias and frequency dependent, causing T_d to be bias and frequency dependent in complicated way. The noise temperature calculated for a represenative dc biased diode at different frequencies is presented at Figures 13 and 14.

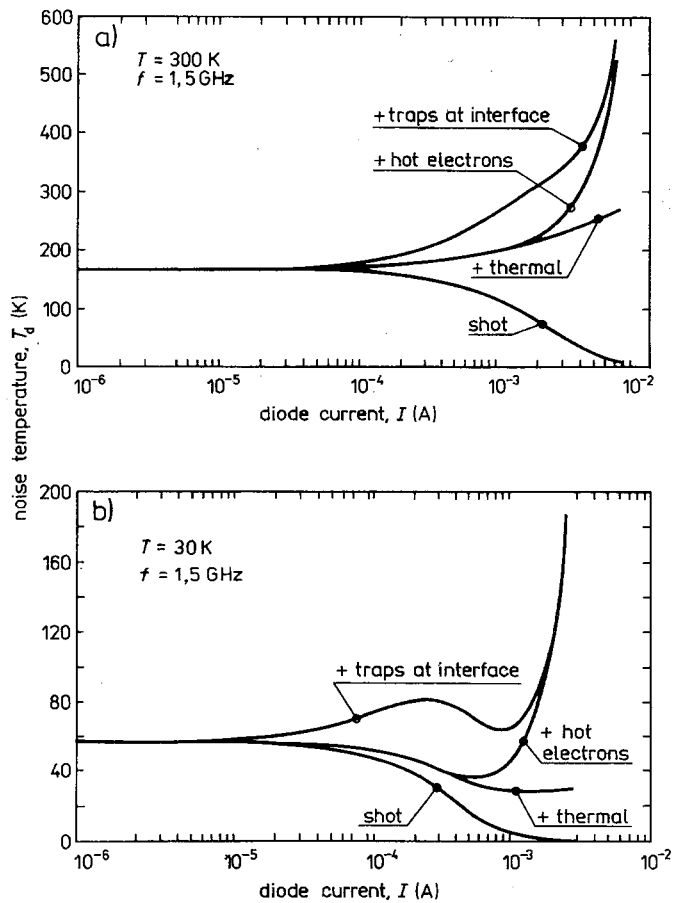


Fig. 13. Contributions of different noise processes to the equivalent noise temperature of the dc biased diode. Diode noise temperature is calculated by summing-up noise contributions with proper weights (i.e., impedance ratios), according to (104). First, only the shot noise is assumed to be present in the diode, then contributions from the thermal noise, the hot electron noise and trapping noise are successively added. In this illustrative example, parameters of representative diodes are used in calculations at frequency 1.5 GHz and temperatures of a) 300 K and b) 30 K

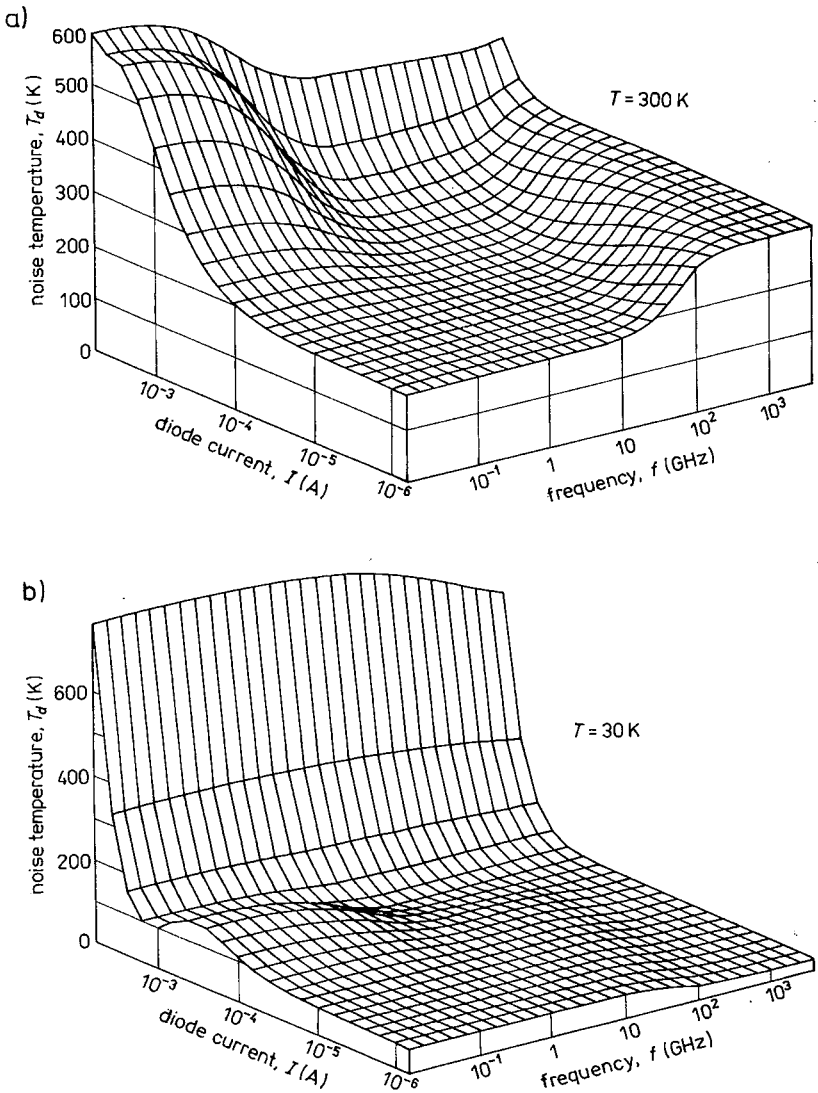


Fig. 14. Noise temperatures of representative diodes as functions of frequency and dc bias current, calculated at temperatures of a) 300 K and b) 30 K. Diode noise is assumed to be contributed to by the shot, thermal, hot-electron and trap noise

9. CIRCUIT MODEL AND TEMPERATURE CONSIDERATIONS

Considerations presented in the preceding sections lead to a diode's circuit model presented at Figure 15. Equations describing the model have been derived or introduced in proper previous sections and therefore only key expressions are given here.

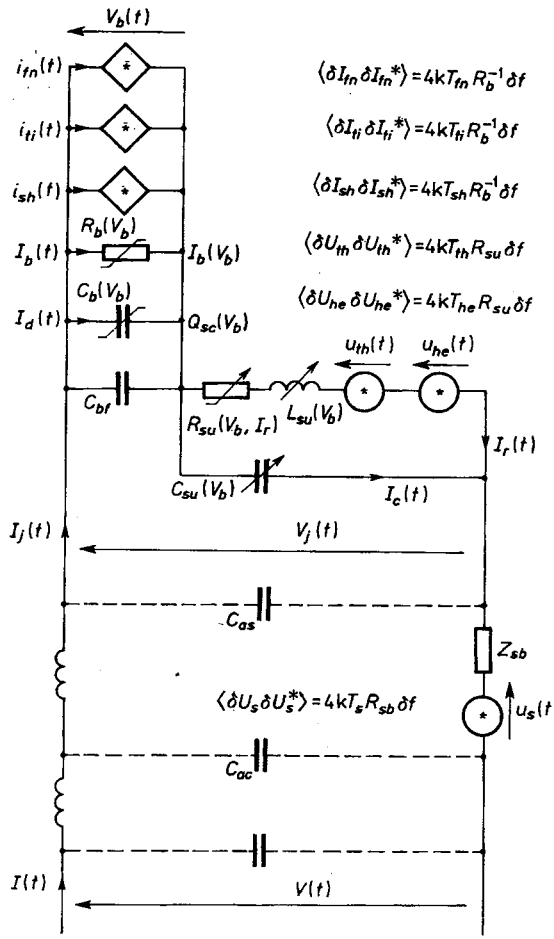


Fig. 15. Circuit model of Schottky diode

The nonlinear junction is the rectifying metal-semiconductor contact and its direct vicinity within the epitaxial layer. The potential barrier is associated with a depletion region (layer). Depletion layer width $w(t)$, space-charge contained in this layer $Q_{sc}(t)$ and the barrier current $I_b(t) = SJ_b(t)$ (S is the junction area) are all dependent on the voltage drop $V_b(t)$ across the barrier and conditions at the barrier edge. These conditions are functions of the junction current $I_j(t)$ which is the sum of the barrier current $I_b(t)$ and the displacement current $I_d(t) = \mathrm{d}Q_{sc}/\mathrm{dt}$.

The depletion layer capacitance (barrier capacitance) is expressed using (46) as

$$C_b(V_b) = C_{b0} \frac{1 - \exp\left(\dfrac{-(V_d - V_b)}{V_T}\right)}{\left[1 - \dfrac{V_b}{V_d - V_T} + \dfrac{V_T}{V_d - V_T} \exp\left(\dfrac{-(V_d - V_b)}{V_T}\right)\right]^{\gamma_0(1+\kappa V_b)}}, \tag{105}$$

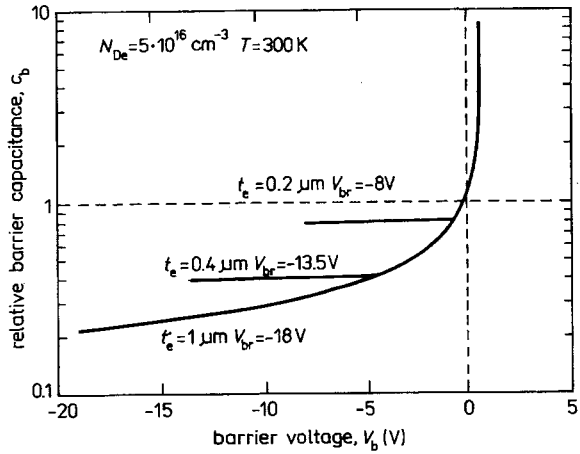


Fig. 16. Relative barrier capacitance versus voltage characteristics of Schottky diodes having different thickness, t_e, of the epilayer. Punch through effect is visible for $t_e=0.2$ μm and $t_e=0.4$ μm

where V_d is the diffusion (or build-in) voltage, $V_T=kT/q$, k is the Boltzmann's constant, q is the charge of an electron and T is the lattice temperature; C_{b0} is the zero bias capacitance, e.g., [64] and $\gamma=\gamma_0(1+\kappa V_b)$. γ_0 is introduced to allow different than abrupt (when $\gamma_0=0.5$, $\kappa=0$) doping profile. Usually the varactor capacitance is less nonlinear with high reverse bias. This is taken into consideration by choosing a positive value for parameter κ. Typical values are $\gamma_0=0.45-0.5$ and $\kappa=0-0.01$ [40]. In the case of small high-frequency diodes the edge effects must be included in the diode model [2, 50, 64 – 67] by adding capacitance C_{bf} in parallel with C_b and parasitic capacitances between the anode and the substrate C_{as} and between the anode and the ohmic contact (the cathode of the diode) C_{ac}. For simplicity, C_{bf}, C_{as} and C_{ac} are assumed to be independent on the diode bias (in fact C_{bf} depends on V_b). If the junction has a circumference U, then $C_{bf}=3\varepsilon_s U/4$ [2], which for cicrular junction with a diameter d and area S gives $C_{bf}=3\varepsilon_s S/d$ (ε_s is the permittivity of the semiconductor). Determinatoin of C_{bf} for other shapes of anodes is more compliated, e.g., [50, 65 – 67], and as C_{as} and C_{ac}, usually employs some experimental evaluation.

Barrier capacitance is dependent on temperature through $V_d=\Phi_b-V_n$ and V_T voltages; e.g., for $N_{De}\approx 3\cdot10^{16}$ cm^{-3} V_n decreases from 75 mV at 300 K to 2 mV at 20 K while Φ_b increases by about 80 mV [10]. Thus V_d would increase by about 150 mV on cooling the diode from 300 K to 20 K, if there was no lattice heating above ambient temperature. In fact lattice temperature change is smaller and V_d increases by about 100 mV. In the $C_b(V_b)$ model given by (105) the degree of capacitance nonlinearity is not temperature dependent. In all, the effect of cooling on the junction capacitance is so small (see Figure 7), that it has an almost negligible effect on the efficiency of a frequency multiplier [68, 69].

Current $I_b(t)$ flowing through the barrier depends on the voltage drop V_b across the depletion layer

$$I_b(V_b)=J_S S\frac{\exp\left(\frac{V_b}{\eta V_T}\right)-1}{1-\exp\left(\frac{-2(V_d-V_b)}{V_T}\right)}-I_{br}(V_b), \tag{106}$$

where $I_{br}(V_b)$ is the breakdown current and J_S is the saturation current density given by

$$J_S=A^{**}T^2\exp\left[\frac{-\Phi_b}{\eta V_T}\right], \tag{107}$$

where A^{**} is the modified Richardson constant [8.2–8.6 $\mathrm{Acm^{-2}K^{-2}}$ for GaAs and 96 $\mathrm{Acm^{-2}K^{-2}}$ for silicon), Φ_b is the height of the potential barrier (0.86–0.94 V for n-type GaAs and 0.90 V for n-type Si-platinum combination) and ideality factor η is given by (34). η depends on the carrier transport mechanism and varies from $\eta\approx 1$ for pure thermionic emission to $\eta\approx 3-7$ for field emission.
For $I_b\gg I_S$ and small values of η the barrier incremental resistance is derived as

$$R_b=\frac{1}{(\partial I_b/\partial V_b)}=\frac{\eta V_T}{I_b}\tanh\left(\frac{(V_d-V_b)}{\eta V_T}\right). \tag{108}$$

It should be noted that for bias voltages such that the difference between the bias V_b and the flat-band voltage $V_{fb}=V_d$ is greater than $3\eta V_T$, the flat-band correction factor $\exp[-(V_d-V_b)/(\eta V_T)]\approx 0$ and the barrier model reduces to the classical one given by (9) and (25).

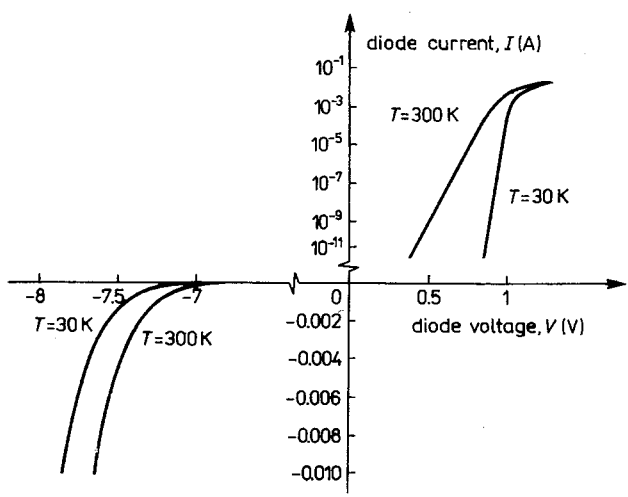


Fig. 17. Current versus bias voltage characteristics of a Schottky diode at temperatures 300 K and 30 K

The temperature has a strong effect on the I/V characteristic of a diode as it was discussed in previous sections and is summarized at Figure 17. Examining this figure it should be pointed out that when the diode is cooled the turn-up point is shifted toward

higher voltages and the steepness of the I/V curve is increased. Thus the voltage range where the frequency conversion is mainly reactive is increased and also the resistive conversion is slightly more effective.

Junction behaviour at reverse voltages close to the breakdown voltage V_{br} ($\approx$2 V for Si and 6–8 V for GaAs mm-wave diodes) is well modeled by a power-law dependences of the breakdown current on voltage

$$I_{br}(V_b)=\begin{cases}10^{-5}\cdot(1+V_{br}-V_b)^E & \text{for} \quad V_b<(1+V_{br}) \\ 0 & \text{for} \quad V_b\geqslant(1+V_{br})\end{cases} \tag{109}$$

where exponent E is usually 10. The magnitude of the greakdown voltage V_{br} slightly increases when the diode is cooled because of the increased energy gap in GaAs [3].

The undepleted layer impedance Z_{su} (i.e., series junction impedance) is modeled by the series connection of the voltage-dependent nonlinear resistance $R_{su}(V_b,I_j)$, and voltage-dependent linear inductance $L_{su}(V_b)$ both shunted by the voltage-dependent capacitance $C_{su}(V_b)$. Nonlinear series resistance $R_{su}(V_b,I_j)$ resulting from carrier velocity saturation at high fields, explains increase of the series resistance, and thus decrease of the conversion efficiency, with increase of the RF power. When the diode is cooled the maximum drift velocity v_{dmax} increases and because of that the maximum electron current ($i_{max}=qN_{De}Sv_{dmax}$) also increases and the effect of the current saturation is less significant [35, 68] (e.g., i_{max} can increase by 50%). This makes possible to pump the junction capacitance more effectively increasing multiplier efficiency, especially at high power levels and at high frequencies.

Both undepleted layer impedance Z_{su} and the substrate impedance $Z_{sb}=Z_{spr}+Z_{skin}$ (given by (65) and (66)) are temperature dependent through the electron mobility in GaAs. When the diode is cooled the mobility increases in the low doped epilayer ($N_{De}\leqslant 2\cdot 10^{17}$ cm^{-3}) but decreases slightly in the heavily doped substrate ($N_{Db}\geqslant 2\cdot 10^{18}$ cm^{-3}). When N_{De} is low ($\approx 1\cdot 10^{16}$ cm^{-3}) the mobility greatly increases and reaches maximum value at $\approx$50 K. At high doping concentration ($N_{De}\approx 2\cdot 10^{17}$ cm^{-3}) the optimum temperature is higher ($\approx$150 K) and the mobility increases only slightly [68]. Combining this with Z_{sb} increase, the net effect may be decrease in series resistance on cooling the diode down to optimum temperature (e.g., from 10.5 Ω to 6 Ω at 77 K for a lightly doped diode, or from 12 Ω to 8.5 Ω for a higher doped smaller area diode). Further cooling below optimum temperature does not reduce the series resistance. The plasma resonance frequency does not change on cooling, because it is independent of the electron mobility. Because the scatteirng frequency and the dielectric relaxation frequency are temperature dependent, the Q-factor of the plasma resonance is also temperature dependent and increases on cooling of the diode.

The noise generated in the diode is modeled by noise sources characterized by the equivalent noise temperatures, and located in proper parts of the diode. The noise generated in the barrier is composed of the shot noise, noise from traps at the junction interface and the flicker noise. The respective noise temperatures are given by (see eqns. (81), (93) and (100))

$$T_{sh}=\frac{\eta T}{2}\tanh\left[\frac{(V_d-V_b)}{\eta V_T}\right] \tag{110}$$

$$T_{ti}=\frac{S_{ii}}{4k}R_b=\frac{\eta T}{2}\tanh\left[\frac{(V_d-V_b)}{\eta V_T}\right]\cdot K_{ti}\frac{1}{1+(\omega\tau_{ti})^2}\cdot I_b \tag{111}$$

$$T_{fn}=\frac{S_{fn}}{4kR_b^{-1}}=K_{fn}\frac{I_b^{\beta}}{f^{\alpha}}R_b(V_b)\,. \tag{112}$$

The noise in the undepleted epilayer results from several different processes and is in general contributed to by the thermal, hot-electron, intervalley scattering and trapping noise. In thin epilayer of a good quality Schottky diode intervalley scattering and trapping noise are negligible. The remaining noise processes are described by the equivalent noise temperatures

$$T_{th}=T \tag{113}$$

$$T_{he}=K_{he}I_j^2, \tag{114}$$

where the noise factor K_{he} is given by (84).

The diode noise is increased by the thermal noise generated in the substrate and diode contacts, which is characterized by the thermal noise temperature $T_s=T_{th}=T$.

The noise process occurring in the diode are independent on each other and hence uncorrelated. Therefore, the noise temperature T_d of a dc biased diode is calculated by summing up the contributions from the different parts of the diode — see eqn. (104). The noise of a diode is a very compliated function of temperature, bias and frequency, as was discussed in Section 7. Complexity and thoroughness of the noise model were reflected at Figures 13 and 14. Here, only two curves of $T_d(I)$, i.e., at room and low temperatures, extracted from that model (trap noise is neglected) are presented for comparison at Figure 18.

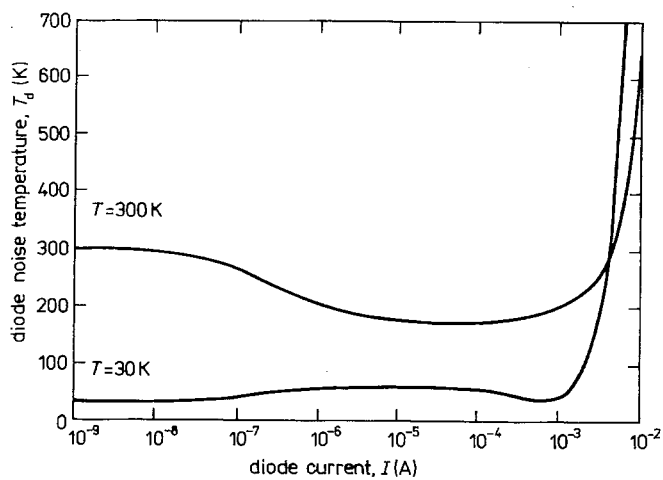


Fig. 18. Diode noise temperature of a dc biased Schottky diode at temperatures 300 K and 30 K

The lattice temperature T depends not only on the ambient temperature T_a but also on the power P_T dissipated within the diode and the rate of heat transfer from the diode to the ambient (which is described by the thermal resistance R_ϑ)

$$T = T_a + P_T \cdot R_\vartheta \tag{115}$$

The heat originates from the ohmic losses in the series impedance and the energy released at the anode by the electrons coming from the semiconductor. The thermal resistance is dominated by the thermal "*spreading*" resistance of the diode chip

$$R_\vartheta = \frac{\rho_\vartheta}{2d}, \tag{116}$$

where $\rho_\vartheta(T)$ is the temperature-dependent thermal resistivity ($22 \cdot 10^{-3}$ KmW^{-1} for GaAs at $T_a = 300$ K). For submicron size diode at room temperature $T_a = 295$ K calculated in [58] junction temperature increase at $I = 1$ mA is of the order of 7 K and may be as high as 100 K at 10 mA. At cryogenic temperatures thermal resistivity is much lower and, for example, at $T_a = 20$ K the lattice heating at 10 mA is less than 10 K and is of minor importance even for such the small area diodes. Power dissipation and hence the diode temperature is much more important in power generating devices, such as IMPATT or Gunn diodes — detailed discussion can be found in [78].

The diode circuit model is augmented by parasitic capacitances and inductances of a possible package (e.g. [70]) in which the diode is encapsulated. Packages can seriously degrade the performance of a diode depending on the ratio of the package parasitics to the device parameters. Therefore, in high-frequency applications bare chips are used rather than encapsulated devices. All the parasitics are linear and, as other linear elements of the diode model, are to be included into a linear part of a circuit which is to be analyzed.

Summing up in short, it should be stated that the above presented model includes both the voltage and frequency dependence of the series impedance and incorporates voltage modulated thickness of the depletion layer, driving up to flat-band conditions, skin effect (mainly in the substrate), dielectric relaxation (mainly in the epilayer), carrier scattering (both in the epilayer and in the substrate) and carrier velocity saturation. The model is adequate at frequencies through millimeter waves and up to about 1 THz. Elements of the model are related to technological parameters and physical dimensions of the varactor.

To check the validity of our model we have used data available in the literature for representative state-of-the-art diodes to calculate diode's parameters from our model. Comparison of our model with real diodes is given in Table 1. Very good agreement with measurements is achieved for all parameters except the breakdown voltage. Measured values are generally lower than theoretically predicted voltages, e.g. [2], because of high electric field intensities near the periphery of the device or surface leakage. These are dependent on technology used in making the diode and are very difficult to predict theoretically. Punch-through voltage V_p, given in the table, substantially higher than the breakdown voltage V_{br} indicates too thick epilayer and suggest that the thickness t_e could have been smaller.

Table 1

Parameters of representative state-of-the-art Schottky-barrier varactor diodes and their comparison with the circuit model

Diode type	N_{De} [cm^{-3}]	d [μm]	t_e μm]	R_s [Ω]	C_{j0} [fF]	C_{b0} [fF]	$C_{bf}+C_{as}$ [fF]	V_{br} [V]	V_p [V]	f_c [THz]	Reference
2T2	$1.0 \cdot 10^{17}$	2.5	0.59	12.0	5.5			11.0		2.41	[26, 28]
	$1.0 \cdot 10^{17}$	2.47	0.50	12.7	5.52	4.64	0.88	14.3	17.1	2.27	our model
VDO12	$6.5 \cdot 10^{16}$	3.6	0.80	11.0	12.5			14.0		1.16	[40]
	$6.5 \cdot 10^{16}$	4.21	0.85	11.0	12.50	10.87	1.60	18.2	33.0	1.16	our model
6P4	$3.5 \cdot 10^{16}$	6.3	1.00	9.5	20.0			20.0		0.84	[28]
	$3.5 \cdot 10^{16}$	6.30	1.00	9.6	20.40	17.90	2.50	26.9	24.4	0.81	our model
3A1	$5.0 \cdot 10^{16}$	6.0	1.00	6.5	33.6			13.5		0.73	[27]
	$5.0 \cdot 10^{16}$	7.05	1.10	6.7	29.90	26.70	3.20	21.3	42.8	0.79	our model
UVA#2	$2.5 \cdot 10^{16}$	10.0	1.20	7.0	40.0			20.0		0.57	[28]
	$2.0 \cdot 10^{16}$	10.20	1.23	7.0	39.80	35.40	4.43	39.5	20.9	0.57	our model

CONCLUDING REMARKS

A reliable device for frequency conversion above 100 GHz has been, and still is, the Schottky diode. Its operating principles and properties has been presented and its circuit model adequate at frequencies up to about 1 THz has been derived. There is a continuous development of millimeter wave Schottky barrier diodes and several laboratories have developed whiskerless diodes intended to be surface mounted in planar strucures, e.g. [27, 28, 50]. In a low-parasitic-capacitance version of the diode an etched surface channel and associated air bride are used or the substrate GaAs is removed and replaced by quartz [30, 71].

The Schottky diodes have proven to be very useful in millimeter-wave application. However, they are still hampered by a substantial parasitic series impedance and inherent response time arising from such mechanisms as velocity saturation in the semiconductor. The need of idler circuits in high-order multipliers and degrading efficiency with increasing power are also their drawbacks. Therefore, there is a great interest in developing technologies and new devices which would have potential to generate greater amounts of power at higher frequencies. In response to this demand improved and new devices have been proposed, many of them conceptually different to well known microwave diodes. A systematic classification and review of these rapidly emerging devices is given in [72].

REFERENCES

1. E. Rhoderick and R. Williams: *Metal-Semiconductor Contacts (2nd ed.)*. Monographs in Electrical and Electronics Engineering, No. 19, Oxford Science Publ., 1988
2. M.V. Schneider: *Metal-Semiconductor Junctions as Frequency Converters*, Infrared and Millimeter Waves, Vol. 6, pp. 209–275, K.J. Button, Ed., Academic Press, New York 1982
3. S.M. Sze: *Physics of Semiconductor Devices (2nd ed.)*, John Wiley and Sons, New York, 1981
4. S.A. Maas: *Microwave Mixers (2nd ed.)*. Artech House, Norwood 1993
5. K.M. Kattmann, T.W. Crowe, R.J. Mattauch: *Noise Reduction in GaAs Schottky Barrier Mixer Diodes*, IEEE Trans. Microwave Theory Techn., Vol. MTT-35, No. 2, pp. 212–214, Feb. 1987
6. W.M. Kelly, G.T. Wrixon: *Optimization of Schottky-Barrier Diodes for Low-Noise, Low-conversion Loss Operation at Near-Millimeter Wavelengths*, Infrared and Millimeter Waves, Vol. 3, pp. 77–110, K.J. Button, Ed., Academic Press, New York, 1980
7. G.K. Sherrill, R.J. Mattauch, T.W. Crowe: *Interfacial Stress and Excess Noise in Schottky-Barrier Mixer Diodes*. IEEE Trans. Microwave Theory Tech., vol. MTT-34, No. 3, pp. 342–345, March 1986
8. T.W. Crowe, R.J. Mattauch: *An Analysis of the I-V Characteristics of Cryogenically Cooled Schottky Barrier Mixer Diodes*. IEEE Southeastcon Proc., 1984, pp. 184–188
9. T.W. Crowe, R.J. Mattauch: *Conversion Loss in GaAs Schottky-Barrier Mixer Diodes*. IEEE Trans. Microwave Theory Tech., Vol. MTT-34, No. 7, pp. 753–760, July 1986
10. E.L. Kollberg, H.H.G. Zirath, A. Jeleński: *Temperature-Variable Characteristics and Noise in Metal-Semiconductor Junctions*, IEEE Trans. Microwave Theory Tech., Vol. MTT-34, No. 9, pp. 913–922, Sept. 1986
11. H.H.G. Zirath, S.M. Nilsen, H. Hjelmgren, J.P. Ramberg, E.L. Kollberg: *Temperature Variable Noise and Electrical Characteristics of Au-GaAs Schottky Barrier Millimeter-Wave Mixer Diodes*. IEEE Trans. Microwave Theory Tech., Vol. MTT-36, No. 11, pp. 1469–1475, Nov. 1988

12. H.H.G. Zirath, S.M. Nilsen, E.L. Kollberg, T. Andersson, W. Kelly: *Noise in Microwave and Millimeter Wave Pt-GaAs Schottky Diodes.* Proc. 14th Eurpean Microwave Conf., Liege 1984, pp. 477–482
13. T.W. Crowe, R.J. Mattauch: *Analysis and Optimization of Millimeter- and Submillimeter-Wavelength Mixer Diodes.* IEEE Trans. Microwave Theory Tech., Vol. MTT-35, No. 2, pp. 159–168, Feb. 1987
14. G.T. Wrixon: *Schottky Diode Realization for Low-Noise Mixing at Millimeter Wavelengths.* IEEE Trans. Microwave Theory Tech., Vol. MTT-24, No. 11, pp. 702–706, Nov. 1976
15. R.J. Mattauch, T.W. Crowe, W.L. Bishop: *Frequency and Noise Limits of Schottky Barrier Mixer Diodes.* Microwave Journal, vol. 28, No. 3, pp. 101–116, March 1985
16. A. Jeleński, M.V. Schneider, A.Y. Cho, E.L. Kollberg, H. Zirath: *Noise Measurements and Noise Mechanisms in Microwave Mixer Diodes.* 1984 IEEE MTT-S Int'l. Microwave Symp. Digest, San Francisco 1984, pp. 552–554
17. I. Mehdi, P.H. Siegel, J. East: *Improved Millimeter-Wave Mixer Performance Analysis Using a Drift Diffusion Capacitance Model.* 1991 IEEE MTT-S Int'l. Microwave Symp. Digest, Boston 1991, pp. 887–890
18. A. Jeleński, A. Grüb, V. Krozer, H.L. Hartnagel: *New Approach to the Design and the Fabrication of THz Schottky Barrier Diodes.* IEEE Trans. Microwave Theory Tech., Vol. MTT-41, No. 4, pp. 549–557, April 1993
19. H.K. Henisch: *Semiconductor Contacts.* Clarendon Press, Oxford 1984
20. M.E. Adamski, M.T. Faber: *Modelling of Schottky-Diode Characteristics for Terahertz Applications.* Proc. X Int'l. Microwave Conf. MIKON 94, Książ, Poland, May 1994, pp. 266–270
21. N.J. Keen: *Low-Noise Millimeter-Wave Mixer Diodes: Results and Evaluation of a Test Programme.* IEE Proc. Vol. 127, Pt. 1, No. 4, pp. 188–198, Aug. 1980
22. W.C.B. Peatman, T.W. Crowe: *Design and Fabrication of 0.5 Micron GaAs Schottky Barrier Diodes for Low-Noise Terahertz Receiver Applications,* Int'l. J. Infrared Millimeter Waves, Vol. 11, No. 3, pp. 355–365, March 1990
23. G.T. Wrixon: *Low-Noise Diodes and Mixers for the 1-2 mm Wavelength Region.* IEEE Trans. Microwave Theory Tech., Vol. MTT-22, No. 12, pp. 1159–1165, Dec. 1974
24. A. Jeleński, E.L. Kollberg, H.H.G. Zirath: *Broad-Band Noise Mechanisms and Noise Measurements of Metal-Semiconductor Junctions.* IEEE Trans. Microwave Theory Tech., Vol. MTT-34, No. 11, pp. 1193–1201, Nov. 1986
25. K. Bhaumik, B. Gelmont, R.J. Mattauch: *Series Impedance of GaAs Planar Schottky Diodes Operated to 500 GHz,* IEEE Trans. Microwave Theory Tech., Vol. MTT-40, No. 5, pp. 880–885, May 1992
26. T.W. Crowe, W.C.B. Peatman, E. Winkler: *GaAs Schottky Barrier Varactor Diodes for Submillimeter Wavelength Power Generation,* Microwave and Optical Technology Letters, Vol. 4, No. 1, pp. 49–53, Jan. 1991
27. J.W. Archer, R.A. Batchelor, C.J. Smith: *Low-Parasitic, Planar Schottky Diodes for Millimeter-Wave Integrated Circuits.* IEEE Trans. Microwave Theory Tech., Vol. MTT-38, No. 1, pp. 15–22,Jan. 1990
28. B.J. Rizzi, J.L. Hester, H. Dossal, T.W. Crowe: *Varactor Diodes for Millimeter and Submillimeter Wavelengths.* Proc. Third Int'l. Symp. on Space Terahertz Technology, Ann Arbor 1992, pp. 73–92
29. T.W. Crowe, W.C.B. Peatman, P.A.D. Wood, X. Liu: *GaAs Schottky Barrier Diodes for THz Applications.* 1992 IEEE MTT-S Int'l. Microwave Symp. Digest, Albuquerque 1992, pp. 1141–1144
30. W.L. Bishop, T.W. Crowe, R.J. Mattauch, H. Dossal: *Planar GaAs Diodes for THz Frequency Mixing Applications.* Proc. Third Int'l. Symp. on Space Terahertz Technology, Ann Arbor 1992, pp. 600–615
31. T.W. Crowe, R.J. Mattauch, H.P. Röser, W.L. Bishop, W.C.B. Peatman, X. Liu: *GaAs Schottky Diodes for THz Mixing Applications.* Proc. IEEE, Vol. 80, No. 11, pp. 1827–1841, Nov. 1992

32. W.M. Kelly, G.T. Wrixon: *Conversion Losses in Schottky-Barrier Diode Mixers in the Submillimeter Region.* IEEE Trans. Microwave Theory Tech., Vol. MTT-27, No. 7, pp. 665–672, July 1979
33. O. von Roos, Ke-Li Wang: *Conversion Losses in GaAs Schottky-Barrier Diodes.* IEEE Trans. Microwave Theory Tech., Vol. MTT-34, No. 1, pp. 183–187, Jan. 1986
34. E. Bava, G.P. Bava, A. Godone, G. Rietto: *Analysis of Schottky-Barrier Millimetric Varactor Doublers.* IEEE Trans. Microwave Theory Tech., vol. MTT-29, No. 11, pp. 1145–1149, Nov. 1981
35. E.L. Kollberg, T.J. Tolmunen, M.A. Frerking, J.R. East: *Current Saturatoin in Submillimeter Wave Varactors.* IEEE Trans. Microwave Theory Tech., Vol. MTT-40, No. 5, pp. 831–838, May 1992
36. T.W. Crowe: *GaAs Schottky Barrier Mixer Diodes for the Frequency Range from 1 – 10 THz.* Int'l. J. Infrared Millimeter Waves, Vol. 10, No. 7, pp. 765–777, July 1989
37. A. van der Ziel: *Infrared Detection and Mixing in Heavily Doped Schottky–Barrier Diodes.* J. Appl. Phys., Vol. 47, No. 5, pp. 2059–2068, May 1976
38. R.O. Grondin, P.A. Blakey, J.R. East: *Effects of Transient Carrier Transport in Millimeter-Wave GaAs Diodes.* IEEE Trans. Electron Devices, Vol. ED-31, No. 1, pp. 21–28, Jan. 1984
39. M. Trippe, G. Bosman, A. van der Ziel: *Transit-Time Effects in the Noise of Schottky-Barrier Diodes.* IEEE Trans. Microwave Theory Tech., Vol. MTT-34, No. 11, pp. 1183–1192, Nov. 1986
40. A. Räisänen, M. Sironen: *Capability of Schottky Diode Multipliers as Local Oscillators at 1 THz.* Microwave and Optical Technology Letters, Vol. 4, No. 1, pp. 29–33, Jan. 1991
41. A. Räisänen: *Frequency Multipliers for Millimeter and Submillimeter Wavelengths,* Proc. IEEE, Vol. 80, No. 11, pp. 1842–1852, Nov. 1992
42. L.E. Dickens: *Spreading Resistance as a Function of Frequency.* IEEE Trans. Microwave Theory Tech., Vol. MTT-15, No. 2, pp. 101–109, Feb. 1967
43. K.S. Champlin, D.B. Armstrong, P.D. Gunderson: *Charge Carrier Inertia in Semiconductors.* Proc. IEEE, Vol. 52, No. 6, pp. 677–685, June 1964
44. K.S. Champlin, G. Eisenstein: *Cutoff Frequency of Submillimeter Schottky-Barrier Diodes.* IEEE Trans. Microwave Theory Tech., Vol. MTT-26, No. 1, pp. 31–34, Jan 1978
45. J.S. Campbell, G.T. Wrixon: *Finite Element Analysis of Skin Effect Resistance in Submillimeter Wave Schottky Barrier Diodes.* IEEE Trans. Microwave Theory Tech., Vol. MTT-30, No. 5, pp. 744–750, May 1982
46. L.K. Seidel, T.W. Crowe: *Fabrication and Analysis of GaAs Schottky Barrier diodes Fabricated in Thin Membranes for Terahertz Applications.* Int'l. J. Infrared Millimeter Waves, Vol. 10, No. 7, July 1989
47. U.V. Bhapkar, T.W. Crowe: *Analysis of the High Frequency Series Impedance of GaAs Schottky Diodes by a Finite Difference Technique.* IEEE Trans. Microwave Theory Tech., Vol. MTT-40, No. 5, pp. 886–894, May 1992
48. E.R. Carlson, M.V. Schneider, T.F. McMaster: *Subharmonically Pumped Millimeter-Wave Mixers.* IEEE Trans. Microwave Theory Tech., Vol. MTT-26, No. 10, pp. 706–715, Oct. 1978
49. J.A. Calviello: *Advanced Devices and Components for the Millimeter and Submillimeter Systems.* IEEE Trans. Electron Devices, Vol. ED-26, No. 9, pp. 1273–1281, Sept. 1979
50. W.L. Bishop, K. McKinney, R.J. Mattauch, T.W. Crowe, G. Green: *A Novel Whiskerless Schottky Diode for Millimeter and Submillimeter Wave Application,* 1987 IEEE MTT-S Int'l. Microwave Symp. Digest, Las Vegas 1987, pp. 607–610
51. A. Kreisler, N. Boucenna, M. Pyée: *Theoretical Model for Submillimeter-Wave Schottky Diode Mixers,* Proc. 15th European Microwave Cofn., Paris 1985, pp. 273–278
52. K.M. van Vliet, A. van der Ziel: *The Quantum correction of the Einstein Relation for High Frequencies.* Solid-State Electronics, Vol. 20, pp. 931–933, 1977
53. T.J. Viola, R.J. Mattauch: *Unified Theory of High-Frequency Noise in Schottky Barriers,* J. Appl. Phys., Vol. 44, No. 6, pp 2805–2808, June 1973

54. E. Kollberg, H. Zirath, M.V. Schneider, A.Y. Cho, A. Jeleński: *Characteristics of Millimeter-Wave Schottky Diodes with Micro-cluster Interface.* Proc. 13th European Microwave Conf., Nüremberg 1983, pp. 561–566
55. N.J. Keen, H.H.G. Zirath: *Hot-Electron Noise Generation in Gallium-Arsenide Schottky-Barrier Diodes*, Electronic Letters, Vol. 19, No. 20, pp. 853–854, Sept. 1983
56. T.W. Crowe, R.J. Mattauch: *GaAs Schottky Barrier Diodes for High Sensitivity Millimeter and Submillimeter Wavelength Receivers.* 1987 IEEE MTT-S Int'l. Microwave Symp. Digest, Las Vegas 1987, pp. 753–756
57. G.M. Hegazi, A. Jeleński, K.S. Yngvesson: *Limitations of Microwave and Millimeter-Wave Mixers Due to Excess Noise.* IEEE Trans. Microwave Theory Tech., Vol. MTT-33, No. 12, pp. 1404–1409, Dec. 1985
58. H.H.G. Zirath: *High-Frequency Noise and Current-Voltage Characteristics of mm-Wave Platinum n-n$^+$-GaAs Schottky Barrier Diodes.* J. Appl. Phys., Vol. 60, No. 4, pp. 1399–1407, August 1986
59. T.J. Maloney, J. Frey: *Transient and Steady-State Electron Transport Properties of GaAs and InP.* J. Appl. Phys., Vol. 48, No. 2, pp. 781–787, Feb. 1977
60. S. Palczewski, A. Jeleński, A. Grüb, H.L. Hartnagel: *Noise Characterization of Schottky Barrier Diodes for High-Frequency Mixing Applications.* IEEE Microwave and Guided Wave Letters, Vol. 2, No. 11, pp. 442–444, Nov. 1992
61. A. van der Ziel: *Unified Presentation of 1/f Noise in Electronic Devices: Fundamental 1/f Noise Sources.* Proc. IEEE, Vol. 76, No. 3, pp. 233–258, March 1988
62. M.T. Faber, M.E. Adamski: *Unified Noise Model of a Submillimeter-Wave Schottky-Barrier Diode.* Proc. X Int'l. Microwave Conf. MIKON 94, Książ, Poland, May 1994, pp. 271–276
63. M.E. Adamski, M.T. Faber: *Transmission of Noise in Nonlinear Devices as Applied to Metal-Semiconductor Junctions.* Proc. X Int'l. Microwave Conf. MIKON 94, Książ, Poland, May 1994, pp. 338–342
64. B. Gelmont, M. Shur, R.J. Mattauch: *Capacitance-Voltage Characteristics of Microwave Schottky Diodes.* IEEE Trans. Microwave Theory Tech., Vol. MTT-39, No. 5, pp. 857–863, May 1991
65. P. Philippe, W. El-Kamali, V. Pauker: *Physical Equivalent Circuit Model for Planar Schottky Varactor Diode.* IEEE Trans. Microwave Theory Tech., Vol. MTT-36, No. 2, pp. 250–255, Feb. 1988
66. J.M. Dieudonné, B. Adelseck, K.E. Schmegner, R. Rittmeyer, A. Colquhoun: *Technology Related Design of Monolithic Millimeter-Wave Schottky Diode Mixers.* IEEE Trans. Microwave Theory Tech., Vol. MTT-40, No. 7, pp. 1466–1474, July 1992
67. J.A. Wells, N.J. Cronin: *Determination and Reduction of the Capacitance Associated with the Bonding Pads of Planar Millimeter-Wave Mixer Diodes.* IEEE Microwave and Guided Wave Letters, Vol. 2, No. 7, pp. 297–299, July 1992
68. J.T. Louhi, A.V. Räisänen, N.R. Erickson: *Cooled Schottky Varactor Frequency Multipliers at Submillimeter Wavelengths.* IEEE Trans. Microwave Theory Tech., Vol. MTT-41, No. 4, pp. 565–571, April 1993
69. J.T. Louhi, A.V. Räisänen: *Cooled Cascaded Frequency Multipliers at 1 THz.* Proc. 22nd European Microwave Conf., Helsinki 1992, pp. 597–602
70. K. Kazi, B.B. Szendrenyi, I. Mojzes: *Lossy Model of Diode Packages: an Alternative Metod for Exact Evaluation of Active Chip Parameters.* 1989 IEEE MTT-S Int'l. Microwave Symp. Digest, Long Beach 1989, pp. 1267–1270
71. W.L. Bishop, T.W. Crowe, R.J. Mattauch, P.H. Ostdiek: *Planar Schottky Barrier Mixer Diodes for Space Applications at Submillimeter Wavelengths,* Microwave and Optical Technology Letters, Vol. 4, No. 1, pp. 44–49, Jan. 1991
72. M.T. Faber, J. Chramiec, M.E. Adamski: *Microwave and Millimeter Wave Diode Frequency Multipliers.* In press, Artech House, Boston, London, 1995

M.T. FABER, M.E. ADAMSKI

PÓŁPRZEWODNIKOWE DIODY *m-n-n*$^+$ DO PRZEMIANY CZĘSTOTLIWOŚCI W ZAKRESACH FAL MILIMETROWYCH I SUBMILIMETROWYCH

Streszczenie

W artykule przedstawiono obszerne i wszechstronne omówienie zasad działania i właściwości współczesnych diod półprzewodnikowych *m-n-n*$^+$ stosowanych do przemiany częstotliwości w zakresach fal milmetrowych i submilimetrowych. W oparciu o zjawiska fizyczne występujące w diodach Schottky'ego (tak zwykle nazywane są struktury *m-n-n*$^+$) sformułowano model obwodowy takich diod. Model ten uwzględnia zależność impedancji szeregowej diody zarówno od napiecia jak i od częstotliwości. Obejmuje on modulację grubości warstwy zubożonej złącza napięciem złącza i wysterowanie aż do napięcia zrównania pasm. Uwzględnia on także takie zjawiska jak nasycanie prędkości nośników, efekt naskórkowy (głównie w podłożu), relaksację dielektryczną (głównie w warstwie epitaksjalnej) oraz rozpraszanie nośników w podłożu i warstwie epitaksjalnej. Model ten jest wystarczająco dokładny aż do częstotliwości sięgajacych 1 THz.

Słowa kluczowe: dioda *m-n-n*$^+$, bateria Schottkyego, waraktory, szumy, model obwodowy. GaAs, mikrofale, przemiana częstotliwości, temperatura szumowa.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 2, ss. 257–285

Practical Millimeter- and Submillimeter-Wave Diode Frequency Multipliers

MAREK T. FABER

Instytut Podstaw Elektroniki, Politechnika Warszawska

Received 1995.02.15

Authorized 1995.03.20

Development of a practical millimeter-wave frequency multiplier has been, because of its complexity, raditionally more art than science and usually required extensive, empirical trial-and-error adjustment and tailoring of the circuit structure and the diode. Translating theoretical (numerical) results into practical microwave realization is of critical importance in the development process, and is generally the point at which ultimate success or failure in achieving a particular aim occurs. To alleviate these difficulties and to lower the risk of failure, a comprehensive systmatic review of state-of-the-art practical diode frequency multipliers is provided in the paper. Several specific design examples of practical multipliers are presented not only to indicate existing limitations and necessary trade-offs, but also to illustrate potentials of contemporary millimetr- and submillimeter-wave diode multipliers.

Key words: frequency multiplier, p- varactors, Schottky varactors, waveguide circuits, quasi-optical circuits, planar circuits.

1. INTRODUCTION

From the beginning of radio communication and broadcasting signal generation has been one of major technical problems. Frequency multiplication has been early recognized as one of the possible signal generation techniques, permitting one to obtain and utilize harmonics of a fundamental frequency signal. Germanium and silicon point-contact diodes were experimentally evaluated as microwave frequency multipliers during the World War II, but the most significant progress in this area should be attributed to the decades 1950–1970 when semconductor technology had advanced enough to fabricate good quality *p-n* junction diodes. The diffused *p-n* junction diode biased in reverse direction could respond at microwave speeds, since changing its capacitance only required movement of charge in and out of the diode's

depletion region, which could ideally occur at the majority carrier saturation velocities. The *p-n* varactor opened new possibilities in exploring the microwave region of electromagnetic waves. It stimulated enormous research efforts in the theory, design, development and application of varactor devices. Very simple (but justified at relatively low frequencies) varactor models were used and the research aimed on getting most from a device resulted in fundamental ("*low-frequency*") limits and set firm foundations for efficient generation of microwaves and millimeter waves using semiconductor diodes.

Nowadays diode frequency multipliers have most important applications as signal sources at higher millimeter- and submillimeter-waves because at those frequencies signals can not be generated directly as it is feasible at lower frequencies where either transistor oscillators or Gunn diode or IMPATT diode oscillators are available. *p-n* junction devices are subject to minority carrier recombination and suffer from the diffusion charge-storage effects which limit their application to varactor mode, typically at microwave frequencies. In these respects Schottky-barrier diodes are superior because they are majority carrier devices. Furthermore, metal-semiconductor junctions can be fabricated more precisely than *p-n* junctions and excellent, repeatable characteristics can be achieved. High quality junctions, sometimes with diameters below one micron, have been successfully developed making frequency multipliers feasible even at frequencies above 1000 GHz. Therefore, the dominant devices used in frequency multipliers are the metal-semiconductor junction (i.e., Schottky) diodes. They are, however, still hampered by a substantial parasitic series impedance and inherent response time arising from such mechanisms as carrier velocity saturation in the semiconductor. In response to the demand of generating greater amount of power at higher frequencies, improved and new devices have been proposed recently, many of them conceptually different from well known microwave diodes.

Successful development of a frequency multiplier hinges on proper design and accurate modelling of both the diode and the embedding circuit of the multiplier. The circuits are electrically complex and difficult to model theoretically, and the diodes are highly nonlinear, thus even more difficult to model. In result, the development of a practical frequency multiplier has been traditionally more art than science and usually required extensive, empirical trial-and-error adjustment and tailoring of the circuit structure and the diode.

Diode frequency multipliers may be generally classified as being of varistor or varactor type. In the first case frequency multiplication is performed by a nonlinear resistance or conductance with a consequent poor conversion efficiency but a very large potential bandwidth. In the second case a nonlinear reactive element (usually a nonlinear capacitance) is used. Varactor frequency multipliers have high potential conversion efficiency (the theoretical low-frequency limits is 100 percent, i.e., no conversion loss at all) but they exhibit narrow fix-tuned bandwidth and higher sensitivity to operating conditions, and, sometimes, stability problems. Practical diodes for frequency multipliers usually exhibit both the nonlinear conductance and capacitance so the classification of such a device depends on which one of the nonlinear elements plays a dominant role in the frequency multiplication process.

According to the Manley–Rowe relations, which are valid for any nonlinear capacitor, all the input power applied to the diode's reactive junction (i.e., not including the series resistance) must be converted to output power at the harmonics of the input signal frequency. The output power can not be dissipated in the nonlinear capacitance but it must be dissipated in real parts of impedances seen by the junction at all the harmonics (embedding impedances). In real multipliers these impedances are composed of the junction's series impedance, impedances of the multiplier mount (external circuit) and external loading impedances. In practice much of the output power may be dissipated in the junction's series resistance and in circuit losses and care must be taken in multiplier design to maximize conversion efficiency and/or power delivered to the load at the desired output harmonic.

In a varactor multiplier the highest possible value of real power generated in the nonlinar capacitance at the desired (output) frequency f_n occurs when real powers only at the input f_1 and output f_n frequencies are not zero. In this case the power P_n at the output frequency is equal to the power P_1 at the input frequency and, if P_1 is equal to the available power of the source (i.e., input matching), then neglecting power loss in the junctions's series resistance, the efficiency of power conversion approaches 100%. This can only be achieved if the diode's junction capacitance is terminated in a pure reactance at all harmonics other than that desired. In practice the junction's series resistance, which is in series with the terminating impedances, prevents to present a pure reactance to the nonlinear capacitance. Therefore, it is desirable to design the embedding circuit to open-circuit the diode at all unused harmonics, which would inhibit the harmonic currents in the series resistance and thus no power would be dissipated at unused harmonics. If open-circuiting is not feasible then short-circuting would limit power dissipation to the series resistance and thus prevent harmonic power dissipation in the outer embedding network.

It has been shown, e.g. [1–4], that in diodes having C/V characteristics close to that of the ideal Schottky or *p-n* junction (i.e., abrupt junctions), it is impossible to generate harmonics higher than the second if varactor currents are allowed only at the input and output frequencies. In order to generate higher harmonics, it is necessary to allow intermediate harmonic currents to flow in the varactor. Such intermediate harmonics are known as idlers. The abrupt-junction diode can be thought of as providing two basic functoins: frequency doubling and frequency mixing. Thus only frequency doubler is possible without idlers. A tripler can be obtained only with a second harmonic idler — the doubling function produces the second harmonic idler, which is then mixed with the fundamental to produce the third harmonic output. This is termed a 1-2-3 tripler. A quadrupler may have one idler, i.e., 1-2-4 scheme, or two idlers if 1-2-3-4 scheme is used. A quintupler requires at least two idlers, preferably 1-2-3-5 scheme [1, 3].

The necessity of idlers applies only to ideal abrupt-junction varactor. C/V characteristics of real varactors deviate somewhat from this ideal model which allows nonzero higher harmonics without idlers. Furthermore, charge storage in a p^+-n diode causes that overdriving such varactor increases C/V nonlinearity significantly

which permits higher harmonics even without idlers. However, both theory and experiments prove that the use of idlers greatly increases the output power and efficienty of reactive frequency multipliers [5, 6].

Idler circuits are usually realized as short-circuit resonators which allows large currents at idler frequencies. To minimize power dissipation and thus to obtain high efficiency, it is essential to use high unloaded Q (i.e., low-loss) idler resonators. High-order multipliers are most efficient when idler resonators are provided at all idler frequencies. This implies that an efficient higher-order varactor multiplier would have narrow instantaneous bandwidth and usually some trade-off between efficiency and bandwidth is necessary. In addition, at millimeter-wave frequencies it is very difficult to employ many idler circuits and hence a multiplier with two idler resonators is probably a practical limit.

From the above considerations it is clear that achieving optimum efficiency in a frequency multiplier is not an easy task and requires careful design of the multiplier. In general, a varactor must be used that has low series resistance and is appropriate for the frequency band (i.e., with adequate capacitance or high enough cut-off frequency) and power level (i.e., having proper breakdown voltage and junction area) at which it is to be operated. Input power must be efficiently coupled to the diode which necessitates impedance matching at the input. Significant real power should exist in the diode only at the input and output frequencies which means that low loss (i.e., high Q) resonators must be used as idler short-circuits and unused (unwanted) harmonics should be open-circuited. Power generated in the varactor at the output frequency should be efficiently transferred to the external output load. This means that the embedding impedance which maximizes the power in the varactor at the output frequency must be transformed to the output load with low loss and some tuning is necessary in the output circuit to resonate reactances.

2. SINGLE-DIODE FREQUENCY MULTIPLIERS

Development of a microwave frequency multipliers starts with setting of some preliminary assumptions such as: input and output frequency (hence the multiplication factor), the required output power and multiplier instantaneous bandwidth (hence the choice between varistor and varactor operationg mode), need for device external biasing or for the DC output port (e.g., for monitoring purposes), available devices and their parameters, input and output transmission media with corresponding connectors or waveguide flanges. The particular feature of very high frequency multipliers is that they often employ different input and output waveguides as a consequence of different input and output frequency bands or circuit techniques used. In addition, the existence of high order harmonics in the multiplier circuit will often lead to modelling uncertainties, resulting from the possibility of higher modes excitation in the waveguides.

2.1. PLANAR MULTIPLIERS

At present single-diode frequency multipliers find applications mostly as sources of power at higher millimeter- and submillimeter-wave frequencies at which multi-diode circuits are difficult to realize. Therefore, only one example of a planar hybrid-integrated single-diode frequency multiplier is given here to illustrate applications at longer millimeter waves.

A 24/48 GHz planar microstrip doubler [7] is presented at Figure 1. In hybrid planar microstrip integrated circuits, series varactor diode configuration avoids the need for holes to be drilled in the substrate and utilizes only a single stitch bond. This results in better repeatibility, ruggedness and reduced assembly costs.

The multiplier was realized on 127 μm (0.005″) thick alumina mounted on a gold plated kovar carrier. A commercially available varactor diode had cutt-off frequency >900 GHz and exhibited a zero bias capacitance of 0.25 pF, a series resistance of 2 Ω and a breakdown voltage $\approx$25 V. The circuit dimensions are 16.5 mm × 6.1 mm (0.65″ × 0.24″) and its layout is shown in Figure 1. Isolation between the input and output is provided by the input lowpass filter and output stub filter. Proper impedances at the input (match) and output of the diode are provided by quarter-wave impedance transformers and the rejection filters. The input lowpass filter passes the input (24 GHz) signal with less than 1 dB insertion loss and provides 40 dB rejection at 48 GHz. The output filter is a simple narrowband stub filter which passes the 48 GHz signal and provides greater than 30 dB rejection at 24 GHz. The filters are phased so as to provide the proper reactive termination at the diode for the 24 GHz and 48 GHz signals. The effective bond wire inductance fine tunes the multiplier circuit.

The input signal (24 GHz) enters the multiplier on the microstrip and the output signal (48 GHz) is extracted through waveguide which is cut-off to the input signal frequency. The microstrip to waveguide transition is a simple narrowband transition incorporating a 50 Ω half-way into the wavequide E-plane. The nominal conversion loss of the multiplier reported in [7] was 8 dB when driven by 200 mW. Measured performance over −60°C to +100°C showed ±1 dB variation in the output power. Several units were throughly tested under vibration with 100% survivability.

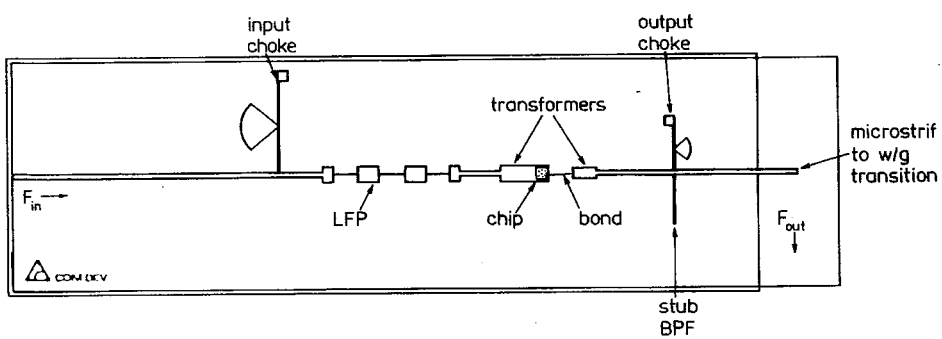


Fig. 1. Circuit layout of the 24 GHz to 48 GHz single-diode frequency doubler. After: E. Boch, *Design of a Rugged Millimeter-wave Doubler Using a Series Varactor Configuration,* 1988 IEEE MTT-S Int. Microwave Symp. Digest, New York 1988, pp. 785–787.

Single-diode planar frequency multipliers find applications also in monolithic microwave integrated circuits (MMICs) in which simplicity of circuit topology (and thus small chip area) is of primary importance. A representative example of a state-of-the-art 94 GHz MMIC frequency doubler is given in Figure 2 [8–10]. The doubler was fabricated on a vapor phase epitaxy substrate that had a buried n^+ layer ($N_{Db}=8\cdot10^{18}$ cm^{-3}) to minimize the diode series resistance. Schottky diode active layer was doped to $N_{De}=7\cdot10^{16}$ cm^{-3}. Ti-Pt-Au metallization was used for Schottky barier contact. Having 16 μm diameter the diode was characterized by $R_s=0.87\ \Omega$, $C_{jo}=0.17$ pF and a breakdown voltage $V_{br}=18$ V. Following completion of the circuits through the front side, via-holes were etched in the thinned wafer. Gold was then plated onto the back side and via-holes to a thickness of 10 μm.

The circuit schematic is shown in Figure 2.a while Figure 2.b shows a microphotograph (top side) of the MMIC doubler chip. A $\lambda/4$ (at 47 GHz) short-circuited (through a via-hole) stub at the input side of the diode, which is equivalent to $\lambda/2$ at 94

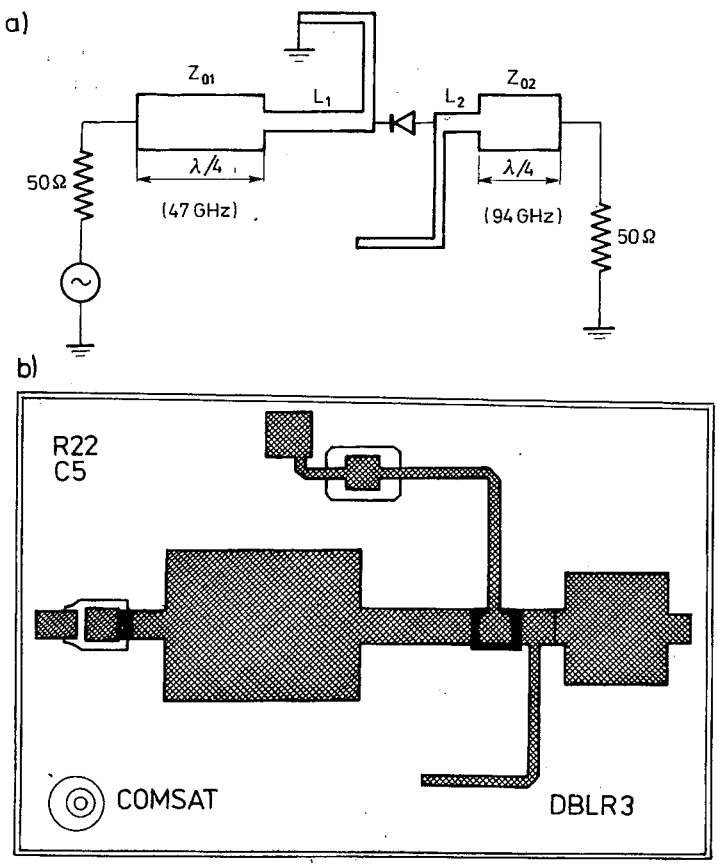


Fig. 2. a) Circuit schematic and b) microphotograph of the 94 GHz single-diode monolithic microwave integated circuit (MMIC) frequency doubler. After: S.-W. Chen, T.C. Ho, K. Pande, P.D. Rice: *Rigorous Analysis and Design of a High-Peformance 94 GHz MMIC Doubler*. IEEE Trans. Microwave Theory Tech., Vol. MTT-41, No. 12, pp. 2317–2322, Dec. 1993

GHz, is used to create an RF short circuit at 94 GHz to prevent the output power generated in the diode from traveling backward. Similarly, a $\lambda/4$ (at 47 GHz) open-ended stub at the output side of the diode creates an RF short at 47 GHz and causes the input signal penetrating through the diode to be reflected back to the diode. A section of transmission line is used as an inductor to resonate the diode junction capacitance. $\lambda/4$ impedance transformers at the input and output are used to transform 50 Ω source and load impedances to optimum diode terminations.

In order to test the doubler the chip was mounted into a test fixture consisting of a U-band input and W-band output microstrip to ridged waveguide transitions. Results of measurements at 94 GHz output frequency are shown in Figure 3. The reverse bias applied to the varactor diode was 7 V. The doubler exhibited maximum efficiency of 25% (6 B conversion loss) and output power of 55 mW at an input power of 220 mW. At an input power level of 330 mW the output power reached 65 mW.

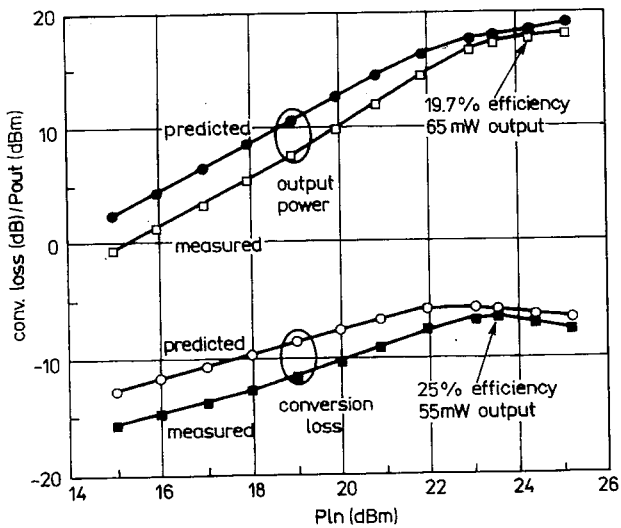


Fig. 3. Measured and predicted performance of the 94 GHz MMIC frequency doubler. After: S.-W. Chen, T.C. Ho, K. Pande, P.D. Rice: *Rigorous Analysis and Design of a High-Peformance 94 GHz MMIC Doubler.* IEEE Trans. Microwave Theory Tech., Vol. MTT-41, No. 12, pp. 2317–2322, Dec. 1993

2.2. WAVEGUIDE MULTIPLIERS

A single-diode frequency multiplier structure which is commonly used at millimter and submillimeter waves and which may be considered as a classical one, is the arrangement of crossed rectangular waveguides of widths specific for the input and output frequency bands. An example of such a design is shown in Figures 4 and 5.

The main advantages of the crossed waveguide structure are as follows:

— the input signal does not excite the output waveguide which should be at cut-off at the input signal frequency;

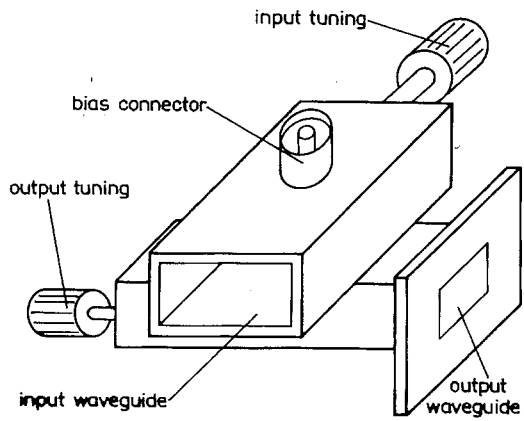


Fig. 4. Configuration of a crossed-waveguide frequency multiplier

— rectangular waveguides exhibit low losses which makes them very attractive in the realization of diode frequency multipliers;

— the height of waveguide in the diode mounting plane may be chosen to facilitate the electrical matching conditions or/and to match the diode physical dimensions;

— movable short-circuits (with the possibility of adding other matching elements) enable to experimentally optimize the diode operating conditions at the input and output frequency; the circuit is therefore particularly attractive in narrowband with the possibility of tuning for optimum performance,

— application of external bias to the diode mounting structure is relatively easy;

— diodes stacked in a single encapsulation may be conveniently used.

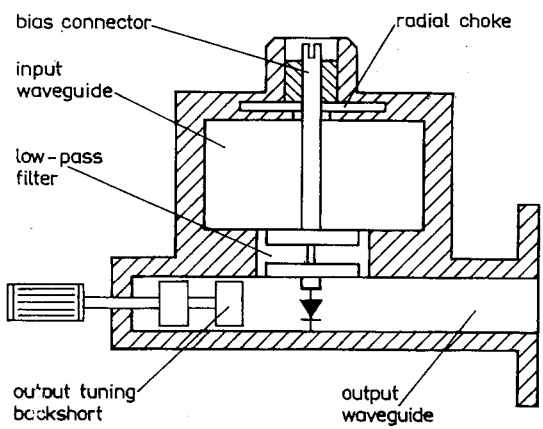


Fig. 5. Cross-section of a crossed-waveguide frequency multiplier

Harmonics of the input signal engendered in the diode may excite the input waveguide, so a low-pass filter or at least a radial choke is required between the input and output guide to prevent transmission of harmonics into the input (lower frequency) waveguide.

Depending on the type of the structure coupling/decoupling input and output waveguides, crossed-waveguide millimeter-wave mounts can be divided into two groups. One group uses a coaxial structure, e.g. [11 – 17] and originates from the mount designed by Erickson [11]. The other group employs a stripline structure, first used by Takada *et al.* [18], and refined by Archer in his design of the frequency multiplier mount [19] and used successfully in various frequency multipliers, e.g. [20 – 27]. An excellent review of state-of-the-art multiplier performance was recently given by Räisänen [17].

An example of a frequency doubler which employs crossed-waveguide mount of the latter design is given in Figure 6. The split-block mount was designed to provide output power at frequencies around 100 GHz and was optimized to achieve highest doubling efficiency at low levels of input power [23, 24]. The same doubler was then used as a first stage in a x2 x3 frequency multiplying chain [25] to drive a quasi-optical frequency tripler [22].

Pump power, incident in the full height WR-15 input waveguide, is fed, via a waveguide-to-stripline transition, to a suspended substrate stripline low-pass filter, fabricated on crystalline quartz 0.076 mm thick. The seven-section filter (0.2 dB ripple Chebyshev design) passes the fundamental frequency with low loss, but is cut off for higher order harmonics. The low-pass filter also transforms the impedance of the pumped varactor at the input frequency to a convenient value at the plane of the waveguide-to-stripline transition. Pump circuit impedance matching is achieved using a sliding contacting backshort.

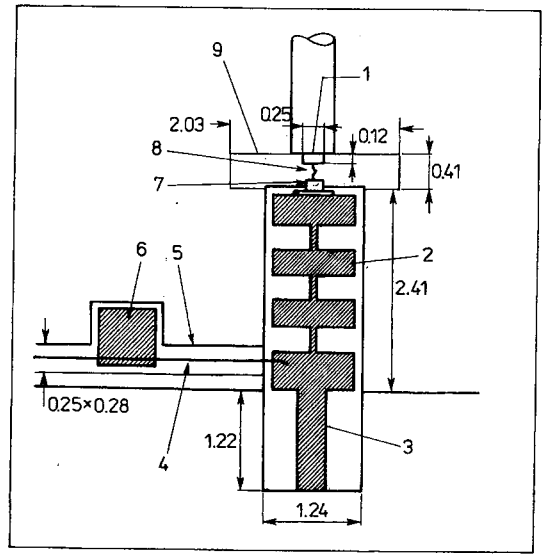
The varactor diode chip is mounted in the reduced-height (by the factor of 0.4) WR-8 output waveguide which provides $Z_o = 226\ \Omega$ at 100 GHz. The diode is contacted by a 12.7 μm diameter, gold-plated, BeCu alloy pin which is an interference fit in the doubler body. The mount is matched to the impedance of the diode at the second harmonic of the pump with the aid of an adjustable contacting loop type backshort [21] in the reduced height guide. A quarter-wave, three-section, step impedance transformer is used to couple the reduced-height guide to the full-height output waveguide.

DC bias is brought to the diode via a transmission line bias filter. The outer shield of the bias line is a rectangular cross section channel milled into the surface of one of the blocks forming the mount. The center conductor is a length of 25 μm diameter gold wire bonded at one end to a low impedance section of the low-pass filter. The line is then connected to a 100 pF metallized quartz dielectric bypass capacitor. The line terminates on the center pin of the SMA bias connector.

The mount is estimated to have the following losses: 0.1 dB in the input waveguide; 0.1 dB in the input sliding backshort; 0.3 dB in the waveguide-to-stripline transition and the low-pass filter; 0.15 dB in the reduced height output waveguide and the sliding backshort; and 0.25 dB in the output waveguide transformer. This gives total losses of 0.5 dB and 0.4 dB in the mount input and output circuits, respectively.

Considering these losses the maximum efficiency of the doubler was expected in [23, 24] to be $\approx 40\%$ if the practical doubler would approach near-optimal operating conditions, namely:

1 — whisker post
2 — low pass filter (quartz substrate)
3 — input waveguide coupling probe
4 — bias wire
5 — $\lambda/4$ bias line
6 — quartz bypass capacitor
7 — varactor diode chip
8 — whisker 0.16 mm long
9 — reduced height outut waveguides



1 — bias filter structure (see inset)
2 — whisker pin
3 — output waveguide (contacting loop type backshort in removed part)
4 — input waveguide backshort (contacting type)
5 — suspended substrate stripline low pass filter and waveguide coupling probe (see inset)
6 — split block mount
7 — WR-15 pump input waveguide flange
8 — WR-8 output waveguide flange (on hidden side)
9 — D.C. bias input

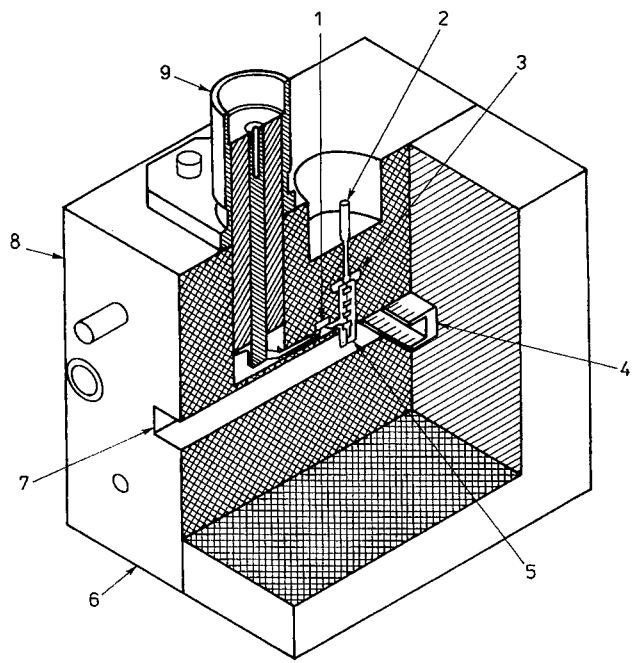


Fig. 6. An isometric drawing and sketch showing the main features of the 100 GHz doubler design and details of the varactor mounting and biasing structure. After: M.T. Faber, J.W. Archer, R.J. Mattauch: *A High Efficiency Frequency Doubler for 100 GHz*. 1985 IEEE MTT-S Int. Microwave Symp. Digest, St. Louis 1985, pp. 363–366

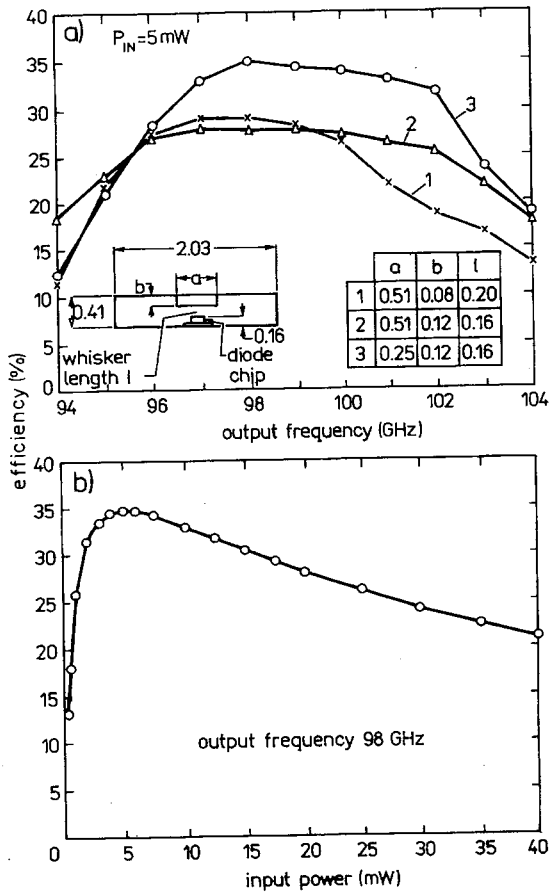


Fig. 7. Conversion efficiency of the 100 GHz doubler a) versus output frequency and b) versus pump power for the final diode mounting structure (case 3 in Fig. 7.a). After: M.T. Faber, J.W. Archer, R.J. Mattauch: *A High Efficiency Frequency Doubler for 100 GHz*. 1985 IEEE MTT-S int. Microwave Symp. Digest, St. Louis 1985, pp. 363–366. ©1985 IEEE

1) The pump circuit is conjugately matched to the diode impedance at the pump frequency (50 GHz).

2) The output waveguide is perfectly decoupled from the pump circuit by the low-pass filter and the filter presents a short circuit to second and higher order pump harmonics at the guide wall.

3) The output waveguide tuning short is adjusted for maximum second harmonic conversion efficiency. This is achieved when varactor reactance is resonated out by the mount at the output frequency (100 GHz).

4) The pump drive is small enough ($P_{in}=10$ mW) to avoid forward current flow in the varactor and not to exceed the reverse breakdown voltage.

and the varactor is modeled by $R_s=10\ \Omega$, $C_{jo}=20$ fF, $\Phi_b=0.92$ V and $\gamma=0.5$.

The performance of the initially assembled doubler was measured and compared with the predicted limit. Then fine changes to the diode embedding circuit were made and the doubler performance reevaluated (Figure 7.a). The whisker post diameter was finally reduced to 0.25 mm, the length of the post in the guide increased to 0.12 mm and the whisker length shortened to 0.16 mm (see Insert of Figure 6). Bias and tuning were adjusted for best performance at each measurement frequency and each pump level. For a 5 mW input power the efficiency of the doubler was greater than 32% at any frequency between 97 and 102 GHz and reaches 35% at the frequency of 98 GHz. Figure 7.b shows the efficiency of the optimized doubler as a function of input power level at 98 GHz. This compares very well to the efficiency reported in [28] for quasi-optical mount but is lower than 45% peak efficiency reported in [15] for a W-band doubler employing a better diode and noncontacting sliding backshorts.

Numerical optimization of frequency multiplier operating conditions lead in both varistor and varactor operating modes to similar qualitative requirements on the diode junctoin embedding impedances, namely source and load optimum impedances providing the best transfer of power together with optimum load reactances at the idle frequencies and open circuited terminations at the higher harmonics. In the case of varistor frequency doublers the conversion loss difference between the best and the worst circuit providing reactive loads at the idle frequencies does not exceed 2 dB but it is more significant for varactor doublers. In the case of frequency triplers and quadruplers the advantages resulting from the use of optimum operating conditions are so significant that the search for suitable circuit configurations is of great practical importance.

In a frequency tripler reported in [21] and [26] the multiplier mount of the design similar to that of Figure 6 was utilized and the output circuit and the low-pass stripline filter were optimized to achieve the best compromise termination of the various pump harmonics for good broad-band performance in 200–290 GHz frequency range. A quarter-wave, two-section impedance transformer was used to couple a 1.14 mm × 0.23 mm reduced-height waveguide in which the diode chip was mounted, to a 0.76 mm × 0.38 mm output waveguide. The second harmonic power could flow in the wider guide, whereas the output guide was cut off at this frequency. The transformer was spaced approximately $\lambda_g/2$ (at the second harmonic wavelength) from the plane of the diode and thus implemented a waveguide resonator and resulted in reactive second-harmonic idler termination. More than 2 mW output power between 200 and 290 GHz with 80 mW input power was provided by this tripler. The peak output power obtained was 4.6 mW at 225 GHz while highest efficiency was 7.5% at 220 GHz output frequency with 30 mW input power.

Crossed-waveguide mounts employing coaxial-line low-pass filter are useful even at frequencies as high as 750 GHz [16] and allow the achievement of excellent performance of a frequency multiplier. An example of such mount is given in Figure 8.a [15]. The input power is fed through the larger (input) waveguide to a low-pass filter and via it further to the diode chip, which is soldered at the end of the filter [17]. One of the diodes on the surface of the chip is contacted by a thin wire (whisker), which acts like an antenna coupling the harmonic power to the smaller (output)

waveguide. The low-pass filter prevents the harmonic power from propagating back to the input waveguide. Sliding back-shorts in input and output waveguides allow tuning at input and output frequencies. The same pin forms the center conductor of both the bias filter and the RF filter between the waveguides. This pin is supported and locked by two Macor ($\varepsilon_r \simeq 5.7$) rings. The bias filter was designed [15] so that it was seen as a short circuit at the input waveguide wall. In the design of the RF filter the coaxial-wavegude junction was taken into consideration as a part of the filter. This filter was optimized for slow impedance variations versus the input backshort positions.

The construction consists of several wafers to make machining possible. The top layer holds the Macor rings and a milled channel for the input waveguide and taper. The next wafer is just a flat plate containing holes drilled to form outer conductors of

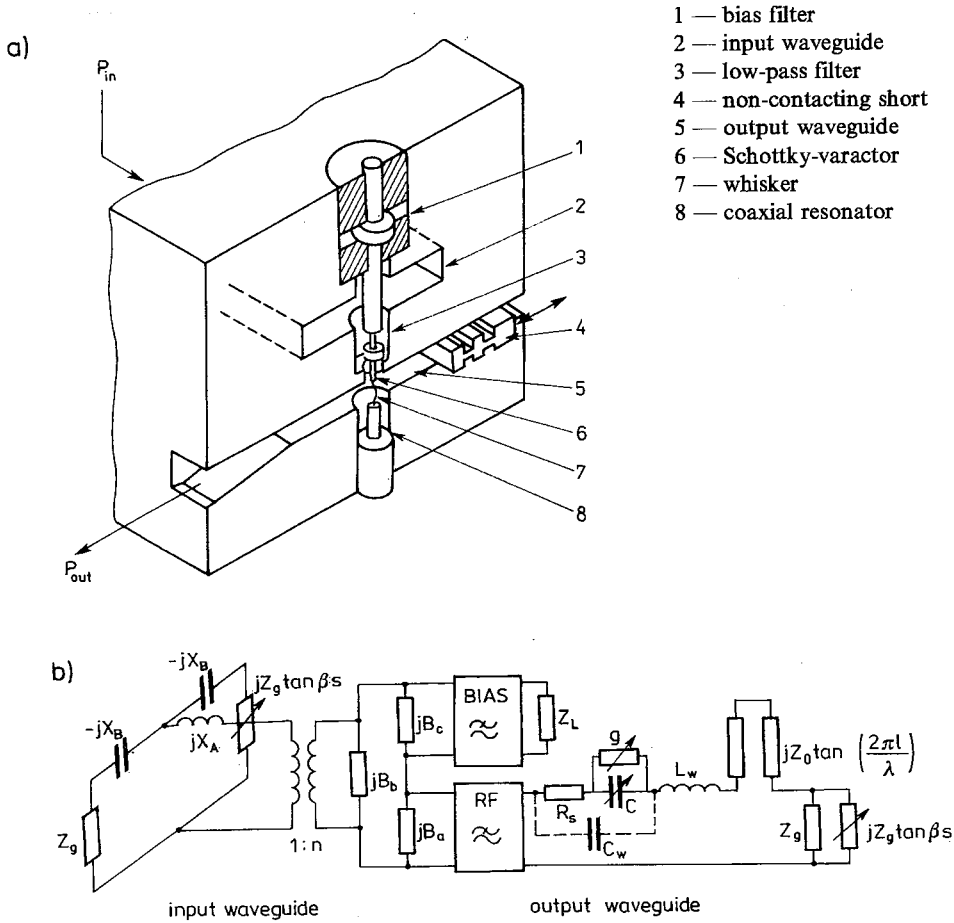


Fig. 8. a) Drawing showing main features of the W-band tripler design and b) the equivalent circuit of the mount. After: T.J. Tolmunen, A.V. Räisänen: *An Efficient Schottky-Varactor Frequency Multiplier at Milimeter Waves*. Part II: *Tripler*, Int. J. Infrared Millimeter Waves, Vol. 8, No. 10, pp. 1337–1353, Oct. 1987

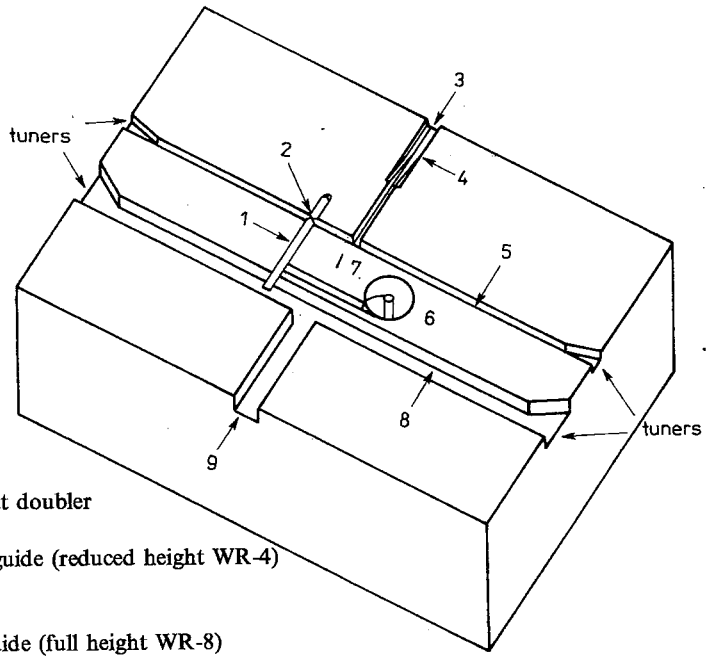
the RF filter, and forms the fourth side of the input and output waveguides. The third wafer has a milled channel for the output waveguide, and the hole for the whisker pin. The remaining two wafers are the bottom covering plate and the top plate supporting a bias connector (SMA).

In the frequency tripler presented in Figure 8.a, a fixed lossless idler termination is provided by means of a short circuited coaxial-line realized in a way proposed by Erickson [11]. In this mount a coaxial resonator is made by modifying the whisker pin [15, 17]. The modified whisker pin forms a short-circuited $\lambda/6$ long transmission line and appears as an inductance at the input frequency. At the output frequency (third harmonic) the line is $\lambda/2$ long and has no effect. At the second harmonic this coaxial resonator ($\lambda/3$) reflects a capacitive reactance to the varactor terminals. This is utilized in second harmonic idler circuit together with the whisker and output waveguide (cut off at the second harmonic), which cause a high inductive reactance at the second harmonic. At the fourth harmonic the resonator tends to increase the inductive reactance of the embedding circuit, which is beneficial.

Experimental measurements performed with a scaled model of the output waveguide and the resonator of a tripler for 100 – 120 GHz enable to formulate equivalent circuit of the mount and calculate values of its elements [15]. The main components of the equivalent circuits presented in Figure 8.b are the waveguides, bias and RF filters, the coaxial resonator and the diode which is contacted by a whisker. The input waveguide-coaxial line junction is represented by X_A, X_B, B_a, B_b, B_c and transformer ratio n. In the output waveguide port, the susceptance caused by the RF filter/waveguide junction is small for typical mount dimensions. On the contrary the susceptance caused by the whisker-varactor gap is significant. Owing to these the equivalent circuit of the output waveguide part is simplified to a capacitance C_w, mainly due to the effect of the gap (in a reduced height waveguide C_{gap} is typically 1.5 – 3.0 fF), and an inductance of the whisker L_w. The coaxial resonator, whose input reactance is in series with inductance L_w, is represented at Figure 8.b by a short-circuited line of Z_0 characteristic impedance and the length l. The distance between the backshort and the diode is denoted by s and is used to calculate waveguide reactances seen toward the backshort. It should be noted that the output waveguide is below cut-off frequency both for the input (pump) frequency and its second harmonic (idler frequency).

Optimized mounts of the above design resulted in very efficient triplers. A peak efficiency of 22% at 100 GHz and 18% at 230 GHz were reported in [11, 13]. A tripler for 100 – 115 GHz output produced a peak efficiency of 28% at 107 GHz. Tripler designed to operate as a second stage of a x2 x3 multiplier chain provided 0.7 mW at 474 GHz [14], while similar design resulted in a tripler producing 0.12 mW at frequency of 803 GHz with 0.8% efficiency [16].

An interesting novel split-waveguide mount for millimeter- and submillimeter-wave applications has been reported recently [29]. The mount consists of only two pieces, block halves, which are mirror images of each other. One part of the mount is shown in Figure 9. The mount provides parallel and series impedance tuning with two sliding backshorts at both the input and the output frequencies while utilizing E-plane



1 — RF filter
2 — position of diode
3 — signal in (mixer), signal out doubler
4 — transformer
5 — 2nd harmonic tuner waveguide (reduced height WR-4)
6 — coaxial connector (below)
7 — bias IF filter
8 — fundamental tuner waveguide (full height WR-8)
9 — pump /LO

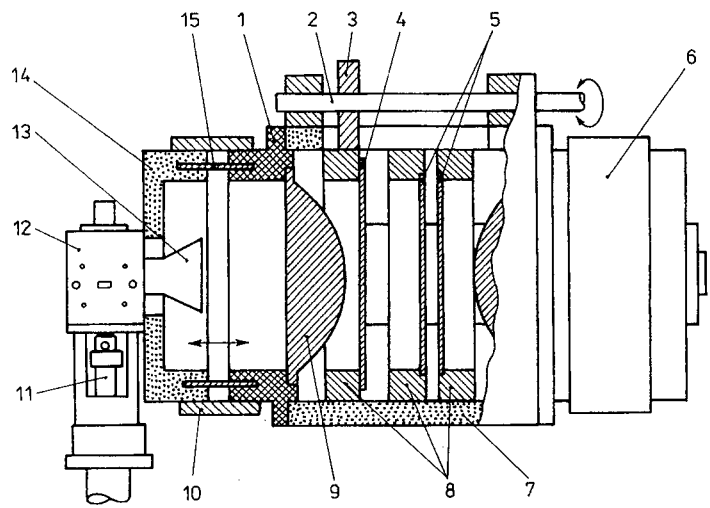
Fig. 9. Schematic drawing showing one half of the 220-GHz doubler/harmonic mixer split-block. After: A.V. Räisänen et al.: *A Novel Split-Waveguide Mount Design for Millimeter- and Submillimeter-Wave Frequency Multipliers and Harmonic Mixers.* IEEE Microwave and Guided Wave Letters, Vol. 3, No. 10, pp. 369 – 371, Oct. 1993

arms to provide an in-line waveguide input and output. Its fabrication is much easier (it can be milled and no electroforming is required) than that of a traditional multifrequency waveguide mount. Waveguide losses are minimized by a very compact design with very short input and output waveguides. Planar diodes are easily mounted on either microstrip or suspended stripline substrates in a channel milled between the tuner waveguides.

With increasing order of frequency multiplication the efficiency inevitably decreases and therefore optimum idler circuits at the second and the third harmonics are essential in a frequency quadrupler to achieve high efficiency. The multiplier mount becomes more complex and the useful operational bandwidth becomes narrower (mixed mode, i.e., varactor/varistor operation, usually results in a wider bandwidth). In a crossed-waveguide structure it is difficult to provide more than one optimum idler without further complication of the mount, which increases manufacturing difficulties. Therefore, in a practical mount the realizations of the idler reactances are based on the fixed termination. To provide near-optimum idler reactances both the coaxial resonator used in the above described tripler and the waveguide resonator introduced in [21] were employed in a quadrupler reported in [15]. The quadrupler had crossed waveguides and two backshorts, one in each

waveguide, as tuning elements at the input and output frequencies. The coaxial resonator was made by reducing the diameter of the last section of the whisker pin to form short-circuited coaxial line. At the diode plane the output waveguide was cut-off at the fundamental and second harmonic frequencies. Further away from the diode the waveguide size was reduced by a step so that it was cut-off also at the third harmonic, thus implementing a waveguide resonator. The highest efficiency obtained with this quadrupler was 11.3% at 148 GHz with input power level of 10 mW and an output power of 1.5—2.4 mW was available over the range from 142.5 GHz to 152.5 GHz.

In higher order frequency multipliers designed to operate at output frequencies above 200 GHz, the waveguide and stripline structures required to optimally tune the various harmonic frequencies for maximum conversion efficiency become extremely difficult to fabricate, especially if it is required that the multiplier be tunable over



1 — lens holder with left hand external thread (40 T.P.I.) 2.50″ external diameter
2 — threaded shaft 0.190 diam., 80 T.P.I (1 of 3)
3 — carrier drive pinion reduction ratio 2.5:1 (1 of 3)
4 — dichroic plate
5 — fused quartz plates for output tuner
6 — second lens / feed horn assembly used for performance measurements (removed when used in conjuction with L.O. diplexer)
7 — quasi-optics body, internal thread — 32 T.P.I. bore: 2.653″ diam.
external dimensions: 3″ × 3″ × 2.5″
8 — carriers for quasi-optical tuning elements made from 2.563″ diam., 32 pitch, stainless steel gear
9 — teflon collimating lens
10 — adjusting ring-internal mating left and right hand threads
11 — drive for stub tuner short
12 — varactor diode mount
13 — scalar feed horn
14 — feed horn carrier with right hand external thread (40 T.P.I.) 2.50″ diam.
15 — lock pin to prevent rotation of carrier

Fig. 10. A schematic diagram of the assembled quasi-optical frequency tripler. After: J.W. ARcher: *A Novel Quasi-Optical Frequency Multiplier Design for Millimeter and Submillimeter Wavelengths.* IEEE Trans. Microwave Theory Tech., Vol. MTT-32, No. 4, pp. 421—427, Apr. 1984. ©1984 IEEE

a wide frequency range. This problems can be overcome by using quasi-optical elements for separate idler and output frequency tuning and filtering. In the multiplier reported in [22] the varactor diode is coupled to the pump source via waveguide and stripline impedance matching and filtering structures. Output power at the various harmonics of the pump frequency engendered in the diode is fed to quasi-optical filtering and tuning elements which enable near-optimum impedances to be seen by the varactor.

The varactor diode, mounted in a broad-band single-ridge waveguide structure, is coupled to the quasi-optics via a tapered waveguide transformer and a broad-band scalar feed horn and circularly symmetric collimating lens made from teflon. The idler termination is implemented using a dichroic mirror as a high-pass filter. The dichroic plate comprises an aluminium sheet of accurately determined thickness, perforated with an equispaced array of holes of precisely machined diameter. Power at the output frequency, after passing through the dichroic reflector, is incident on a pair of parallel plates made from fused quartz (a quasi-optical dual dielectric plate tuner). The tuning plates may be positioned as necessary relative to the varactor mount and to each other. This type of tuner behaves in a similar way as a double stub tuner in a coaxial line or waveguide. Practical quasi-optical frequency triplers reported in [22] achieved 8% efficiency between 250 GHz and 280 GHz and 5% at 340 GHz. A frequency quadrupler to the 310–345 GHz was implemented in this design by installing an additional (i.e., second) dichroic plate to reflectively terminate the third harmonic. The best quadrupler efficiency was 0.75% at 332 GHz.

3. FREQUENCY MULTIPLIERS WITH DISTRIBUTION OF POWER

For a given diode, the output power of a single-diode frequency multiplier can be maximized by careful mount opitimization. To achieve a further significant increase in multiplier output power either a device with significantly greater breakdown voltage may be employed in the single diode mount or the input (pump) signal be split between many identical diodes and then the desired harmonic output component from each varactor recombined with the correct phase relationship. The first possibility has found its maturity in the development of a stacked device which is grown by epitaxial techniques on a single crystal wafer containing two, three or more active varactor diodes in series [30–32]. The latter can be implemented either in a classical type mount, e.g. [25] or by quasi-optical spatial splitting and recombining of signals [33–35].

3.1. SIMPLE POWER COMBINING TECHNIQUES

The epitaxially grown multi-junction (ISIS) varactor consists of a series arrangement of varactor *p-n* junctions, grown by multiple-layer epitaxy on a single wafer of gallium arsenide. Then the diode is etched to the desired capacitance. In this

construction the active region is adjacent to the heat sink which results in a low thermal resistance [30, 31]. If N junctions are in series, the breakdown voltage will increase by N times. If each junction has N times the capacitance of a comparison single diode, then the total capacitance of a series string will be the same as for the single diode. Increasing the capacitance of each individual junction also decreases the series resistance by the same factor N. Therefore the cut-off frequency is unchanged, as should be the efficiency. Distribution of the input power between N junctions of the ISIS varactor allows to use higher level of the pumping signal and in the ideal case should result in N^2 times higher output power. Typical ISIS varactors reported in [30, 31] have the breakdown voltage of 50 V and 100 V for two-stack and three-stack devices, respectively.

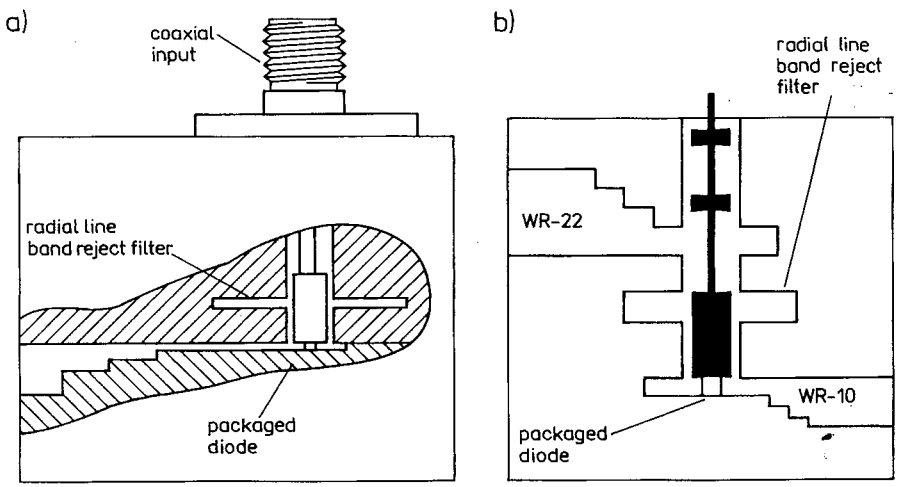


Fig. 11. Frequency multiplier mounts employing epitaxially stacked varactor diodes a) used at Ka-band and b) at W-band. After: a) P.W. Staecker, M.E. Hines, F. Occhiuti, J.F. Cushman: *Multi Watt Power Generation at Millimeter-Wave Frequencies Using Epitaxially-Stacked Varactor Diodes.* 1987 IEEE MTT-S Int. Microwave Sym. Digest, Las Vegas 1987, pp. 917–920 and b) J.F. Cushman, F. Occhiuti, E.M. Mc Donagh, M.E. Hines, P.W. Staecker: *High Power Epitaxially-Stacked Varactor Diode Multipliers: Peformance and Applications at W-band.* 1990 IEEE MTT-S Microwave Symp. Digest, Dallas 1990, pp. 923–926

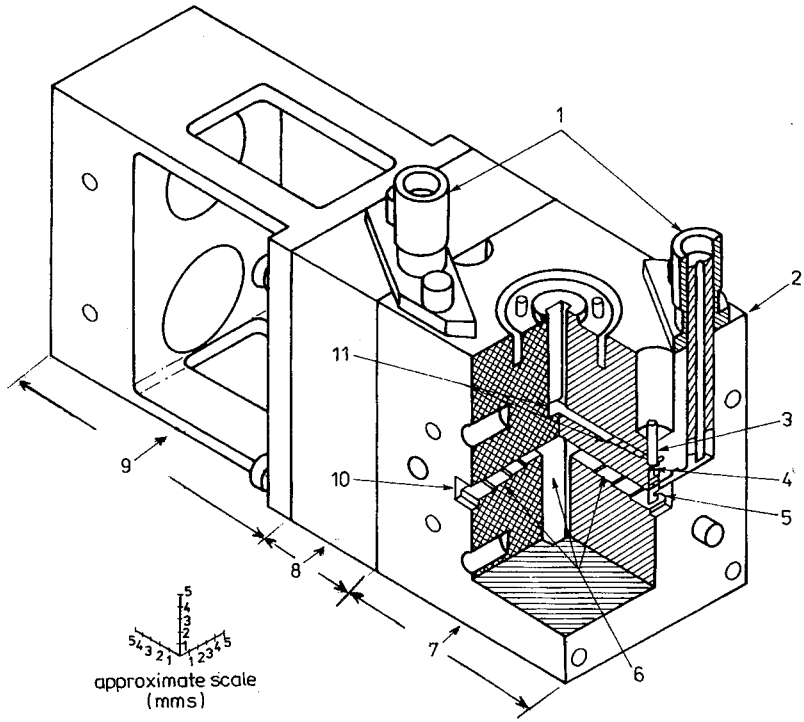
The diode mount, used for Ka-band multipliers is shown in Figure 11.a [30]. The diode is mounted in a reduced height waveguide which is transformed to the full height by a two-section step transformer. The input signal (pump) is fed to the ISIS diode through a coaxial line. A radial choke in the input coax section provides output to input isolation. The W-band doubler, mount shown in Figure 11.b [31], also combines coaxial and waveguide techniques, but has both the input and the output in the waveguide form. The input waveguide contains a two-step waveguide transformer with a transition to coax. Input-output frequency separation is accomplished using a radial line band-reject filter. The output circuit transformation from the packaged

diode to full-height waveguide is achieved through a two-section Chebyshev-design transformer and appropriate backshort position. A coaxial low-pass filter was designed as a means of biasing the varactor and to provide a short-circuited stub to tune the coax transition. Circuit tuning was achieved by adjusting input and output backshort positions and selection of proper bias. For CW operation, a resistor attached to the low-pass filter section of the coaxial center conductor provides the proper self-bias. Alternatively, a fixed DC bias voltage allows fast transition times under pulsed conditions.

A three-junction ISIS varactor for use at Ka-band had a breakdown voltage of 100 V and cut-off frequency of 1 THz when used as a doubler in the mount presented in Figure 11.a, delivered 5.5 W CW output power at 35 GHz with 60% efficiency. A two-diode combination of these varactors provided 5 W CW output power at 44 GHz with 50% efficiency. A 94 GHz single-device used as a doubler in the mount shown in Figure 11.b provided CW output power of 260 mW with 16% efficiency and 850 mW pulsed power output.

In a dual-diode frequency doubler reported in [25] and shown at Figure 12 power splitting and recombining is accomplished by waveguide hybrid devices. The doubler incorporates to identical varactor mounts coupled together, for the purpose of power dividing or combining, by waveguide 4-port T junctions. The matched waveguide hybrid "*magic T's*" used in the mount at the input and output frequencies are, except for the waveguide transformers, geometrically scaled versions of one another, and have theoretical center frequencies of 57 GHz and 105 GHz. Coupling of the magic T's to the waveguides of the varactor mounts is accomplished by means of multisection, quarter-wave step transformers. Each hybrid T junctions were separately electroformed on gold-plated aluminum mandrels in an acid copper bath. The electroforming was carried out in two stages, as shown in Figure 13, in order to maintain precise perpendicularity of the arms of the T's and to ensure excellent symmetry of the junction. The matching elements were fixed into the copper waveguide walls by inserting slightly overlenght, gold-plated posts and shims into holes cut in the mandrels prior to electroforming. The resultant accurately fabricated, internally gold-plated waveguide junctions were pressed into pre-milled slots in a pair of brass half blocks, which were then permanently fixed together to form the final doubler assembly. The waveguide flanges, sripline channels, whisker pin holes, and bias line structures were machined into the surfaces of this block to complete the device. The backshort waveguides were separately electroformed and then pressed into an additional pair of brass blocks which mate with the main assembly, as shown in Figure 12.

The varactor mounts are similar in many respect to the mount presented in Figure 3 but differ in that both the input and output waveguides are normal to the plane of the stripline low-pass filter substrate. The varactor diode chips are mounted on the low-pass filter substrate adjacent to the output waveguide. Each diode chip is contacted by a 0.38 mm-long × 12.7 μm-diam. gold plated, phosphor-bronze electrochemically-pointed whisker. The length of the contact whisker is chosen so that its inductance approximately series-resonates the average capacitance of the pumped

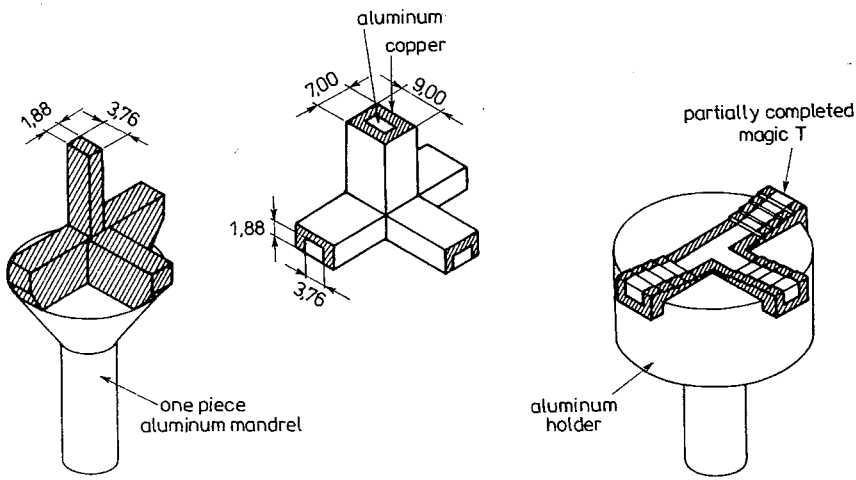


1 — d.c. bias inputs
2 — WR-8 output waveguide flange on hidden side
3 — whisker mounting post
4 — varactor diode chip
5 — low pass filter substrate
6 — input hybrid junction and impedance transformers
7 — hybrid T's and varactor mounts
8 — back short waveguide block
9 — backshort drive mounting bracket (micrometer drives and backshort holders removed)

Fig. 12. An isometric drawing showing the main features of the balanced doubler design. After: J.W. Archer, M.T. Faber: *High-Output, Single- and Dual-Diode, Millimeter Wave Frequency Doublers.* IEEE Trans. Microwave Theory Tech., Vol. MTT-33, No. 6, pp. 533–538, June 1985

varactor diode at the input frequency. Second harmonic power from each varactor mount is coupled to the output magic T via another pair of quarter-wave transformers, and the combined signal is then coupled to the output port.

The varactors used in the dual-diode doubler were selected in order to obtain two diodes with well-matched dc characteristics and were characterized by C_{jo} — 20.5 fF and 21.1 fF, R_s — 8.5 Ω and 8.0 Ω, $V_{br}(1\mu A)$ — 14.5 V and 13.8 V. In measurements performed to evaluate performance of the doubler, backshort tuning and dc bias were optimized for each of the two varactor mounts at each measurement frequency. At a maximum safe input power level of 190 mW, the reverse-bias voltage for optimum performance was about 4.0 V with a forward current between 5 and 8 mA. More than 18 mW output power was obtained at any frequency between 85 and 116 GHz. Over most of this range, more than 20 mW was available, with a peak output power of 26.6



Step 1
Copper electroform onto hatched area.

Step 2
Machine away aluminium mandrel and machine copper as shown

Step 3
Mount piece in aluminium hold. Machine transformer steps. Re-electroform over copper and aluminum of magic T.

Fig. 13. The method of fabricating and holding the "magic T" mandrels during the electroforming process. After: J.W. Archer, M.T. Faber: *High-Output, Single- and Dual-Diode, Millimeter Wave Frequency Doublers*. IEEE Trans. Microwave Theory Tech., Vol. MTT-33, No. 6, pp. 533–538, June 1985

mW at 88 GHz, corresponding to an efficiency of 14%. Higher conversion efficiencies can be obtained with lower pump power. A maximum efficiency of 16.5% was obtained at 104 GHz with a pump power of 80 mW. Obtained results suggest that the performance of the doubler was hampered by high losses caused by the very complicated waveguide structure.

3.2. QUASI-OPTICAL SPATIAL POWER DISTRIBUTION

Quasi-optical spatial power distribution potentially enables watt level of power at millimeter-wave frequencies. The use of a diode grid for frequency multiplication was reported in [33–35]. The approach is attractive because a grid is monolithically integrated with thousands of diodes thereby resulting in potentially low-cost fabrication and small-size realization. The power illuminating the grid is distributed among the diodes and power at a desired harmonic generated in the diodes is then recombined in quasi-optical structure.

A millimeter-wave quasi-optical diode-grid doubler is shown in Figure 14 [33, 34]. Input power at the fundamental frequency is fed to the diode grid through a tuner and filter. A nonlinear capacitance of the diodes illuminated by the incident electromagnetic radiation generates harmonics. The second harmonic passes through another filter and tuning network. The filters consist of a wire polarizing grid with a half-wave plate designed for the fundamental. The half-wave plate separates the fundamental

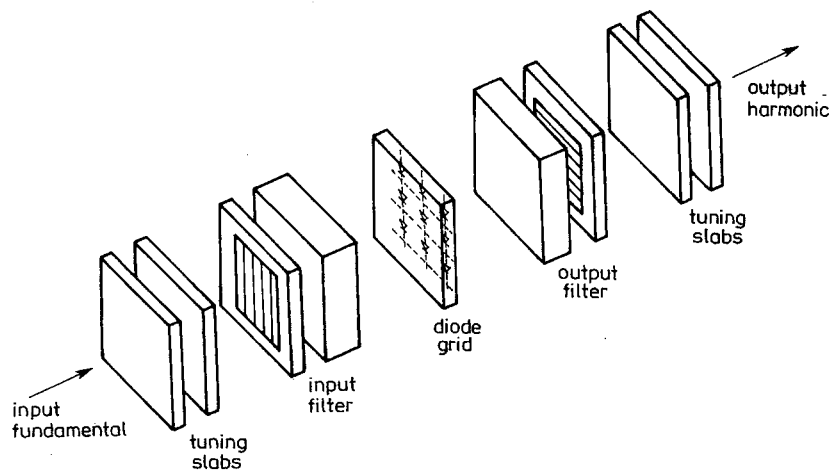


Fig. 14. Configuration of the millimeter-wave diode-grid frequency dobulers array. After: Ch.F. Jou et al.: *Millimeter-Wave Diode-Grid Frequency Doubler*. IEEE Trans. Microwave Theory Tech., Vol. MTT-36, No. 11, pp. 1507–1514, Nov. 1988

from the second harmonic because it rotates the fundamental polarization by 90°, but does not alter the second harmonic polarization. This allows the polarizing grid to select the desired frequency. The tuner is a pair of fused quartz slabs which may be positioned relative to the diode grid and to each other. The tuner behaves in a fashion similar to a double stub tuner in a coaxial line or waveguide. The tuning slabs, filters and grid were all mounted on micrometers, so they can be easily positioned relative to each other.

In a quasi-optical multiplier no waveguides are necessary, the design and modelling are simpler and the losses due to the waveguide wall are eliminated. However, care must be taken to reduce the losses due to diffraction. The input and output filters act effectively as a mirror for the second harmonic and for the fundamental frequency, respectively. Therefore, tuning is done independently at the input and output, with dual dielectric slabs. The power handling capability increases as the size of the grid increases. The design can be easily scaled to higher frequencies.

Monolithic Schottky diode grids were fabricated on 2 cm square GaAs wafers for a proof-of-principle test of the quasi-optical varactor doubler. In the experimental tests, the doubler circuit was placed at the far field of both the transmitting horn and the receiving horn. A helium-neon laser was used to align the system. An efficiency of 9.5% and output power of 0.5 W were achieved at 66 GHz when the diode grid was pumped with a pulsed source at 33 GHz. The performance of the monolithic diode grid was limited by the diode breakdown voltage (which was only -3 V) and losses in the series resistance (27 Ω).

Development of back-to-back Barrier-Intrinsic-N^+ (bbBIN) diodes opened the possibility to use these diodes in a quasi-optical diode grid frequency tripler. Symmetric capacitance-voltage characteristics of bbBIN diodes result in cancellation of even harmonics. This favors bbBIN diodes in tripler application beacuse there is no

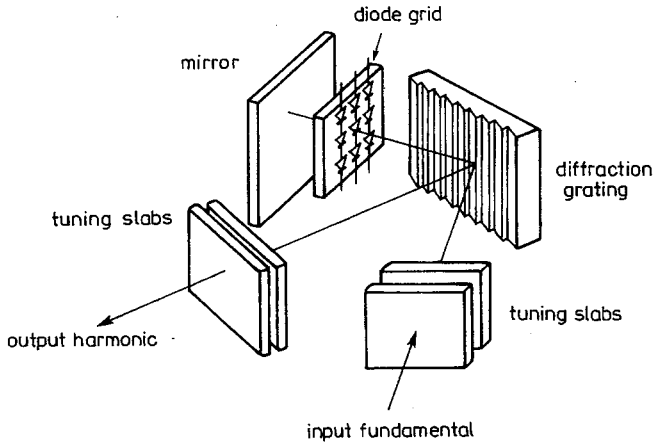


Fig. 15. Configuration of the millimeter-wave diode-grid frequency tripler array. After: R.J. Hwu et al.: *Quasi-Optical Watt-Level Millimeter-Wave Monolithic Solid-State Diode-Grid Frequency Multipliers.* 1989 IEEE MTT-S Int. Micowave Symp. Digest, Long Beach 1989, pp. 1069 – 1072

need for an even (second) harmonic idler circuit in the tripler, which greatly simplifies the circuit design. A complete quasi-optical diode-grid tripler is shown in Figure 15 [35]. Power at the fundamental frequency enters the tripler from the bottom (in the Figure) and passes through an input tuner. The blazed grating plate (which functions as a high-pass transmission filter) reflects (as a mirror) the incident pump power at the fundamental frequency to the diode grid on the left of it, and the metal mirror behind the diode grid again reflects all the harmonics back to the grating plate. Different harmonics are then diffracted in different directions. The third harmonic is designed to exit in the desired direction passing through an output tuner. It should be recalled that, due to the use of bbBIN diodes and thus the elimination of even harmonics, second harmonic idler circuit is unnecessary in the diode-grid tripler. Using the quasi-optical tripler with monolithic planar array containing thousands of GaAs back-to-back Barrier-Intrinsic-N^+ (bbBIN) diodes for a proof-of-principle test, a watt output power at an output frequency of 100 GHz with a tripling efficiency 8.5% was experimentally obtained from approximately 4 mW incident power on each diode. It should be noted that diode-grid frequency multipliers are at early stages of development and significant improvement might be expected in the future.

4. BALANCED MICROWAVE FREQUENCY MULTIPLIERS

The most important advantages of balanced configurations are their inherent filtering properties as well as increased power handling capability resulting from the use of at least two diodes. Balanced frequency multipliers using distributed circuit techniques rely on microwave equivalents of the winding transformers that appear in balanced circuits used at lower frequencies. Usually a junction of two or more different waveguides or transmission lines is used to provide the required amplitude

and phase relationships and an input/output isolation. Experimental determination of diode embedding impedances in microwave and millimeter-wave balanced multiplier circuits is usually difficult, even on scaled-down low frequency models. Development of balanced frequency mulitpliers for use at higher millimeter and submillimeter waves is very difficult and successful realizations are scarce. Because of complicated microwave structure and assembly (and thus possibly high losses) only frequency doublers have practical significance. (Higher order multipliers utilize usually a single diode configuration). An example of such doubler is given in the next Section.

4.1. BALANCED FREQUENCY DOUBLERS

A crossed waveguide balanced varactor frequency doubler proposed by Erickson [14] is shown in Figure 16. The two diodes of opposite polarity are back-to-back mounted between the rectangular waveguide walls and the end of interconnecting coaxial pin. Whisker contacting of diodes is possible. Contrary to single diode crossed waveguide doublers, the second harmonic signal is generated in the input waveguide and it is transmitted to the output waveguide by a connecting coaxial structure. A radial filter confining the input signal into the input waveguide is included in the transition between the waveguides. The diodes conduct in turn during half-periods of the input signal. The resulting current in the output coaxial line has components at DC and even order harmonics of the input signal. Reduced height input waveguide is used to cut-off TE_{11} or TM_{11} mode that otherwise might be excited by the doubled frequency signal.

Development of exact equivalent circuits is very difficult because of many discontinuities existing in the structure and because of possible higher modes excitation. Experimental optimization of multiplier is necessary after initial design.

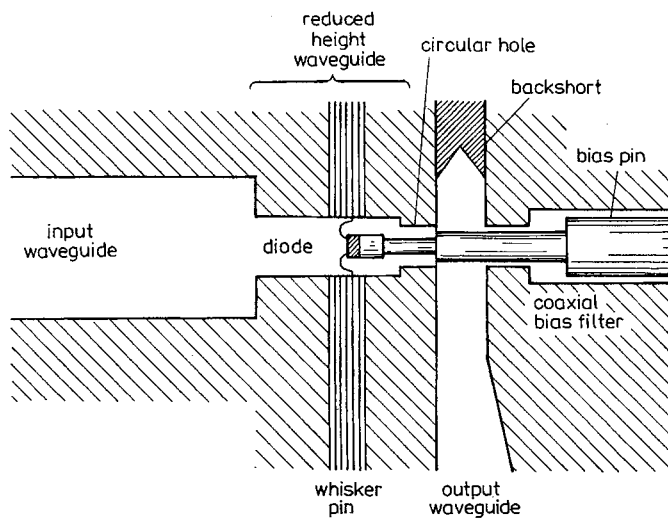


Fig. 16. Cross-section of a crossed waveguide balanced frequency doubler. After: N.R. Erickson: *High Efficiency Submillimeter Frequency Multipliers.* 1990 IEEE MTT-S Int. Micowave Symp. Digest, Dallas 1990, pp. 1301 – 1304

Very impressive results have been obtained by Erickson [14] using the above presented doubler configuration. Peak efficiency of 35 percent has been attained when doubling a 79 GHz, 35 mW signal with varactors of parameters $C_{b0}=21$ fF, $R_s=10\ \Omega$ and $V_{br}=10$ V. Despite the observed heating of the diode junctions an output power of 26 mW has been produced at 120 mW input. Such good performance has been attributed to particularly low loss structure which does not require a low-pass filter between the waveguides and does not excite the third harmonic signal in the output waveguide. A similar 166/332 GHz doubler employing varactors with parameters $C_{b0}=6.5$ fF, $R_s=12\ \Omega$ and $V_{br}=8.5$ V provided output power of 4 mW at $P_1 \cong 22$ mW. The similar circuit idea has been used next [36] to reduce carrier velocity saturation effects due to the use of a planar diode array of four monolithically integrated diodes. An 87/174 GHz doubler exhibited peak conversion efficiency of 25 percent at an input signal power of 150 mW.

5. HIGHER ORDER MULTIPLIER VERSUS MULTIPLIER CHAIN

An important question that often arises in the design of multiplier system is whether it is better to realize a high-order multiplier via a single stage or by a cascade of two or more low-order multipliers. The question is difficult to answer in general particularly at high millimeter and submillimeter waves where power sources capable to drive a multiplier are not easily available. At high frequencies complicated multiplier mounts with many idler circuits are extremely difficult to machine and usually have high losses. If the frequency is further increased then only simple, often quasi-optical, mounts are possible. It must be also considered that it is usually necessary to use isolators between multiplying stages, which introduces additional losses to the multiplier chain.

Theoretical studies [37, 38] do not give a general clear answer. A comparison of the highest theoretical output power available from multipliers of various orders at 345 GHz is given in Figure 17.a. In this case the doubler from 172.5 GHz to 345 GHz gives the best efficiency and widest operational range. However, in practice this doubler is not a good choice because there are no fundamental low noise solid state sources available at 172 GHz. Cascading two doublers overcomes this disadvantage, but resulting output power at 345 GHz is low. It would be even lower in practice because an isolator with several dBs of loss is needed between the doublers. A quadrupler from 86 GHz gives theoretically better efficiency than two cascaded doublers, if the quadrupler has idlers both at the 2nd and the 3rd harmonics. If only the 2nd harmonic idler is used the theoretical efficiency is lower than that of two cascaded doublers. But 1-2-4 quadrupler may be attractive because of no need of an isolator at 172 GHz. On the other hand the quadrupler can provide only very limited operational frequency range.

Another theoretical comparison is presented in Figure 17.b in which the highest theoretical output power available from various multipliers at 460 GHz is shown.

Again the single doubler now from 230 GHz to 460 GHz gives the best efficiency and the highest output power if the input power level is higher than 15 mW. With input power below 15 mW the doubler and the quadrupler from 115 GHz to 460 GHz have almost an equal theoretical performance.

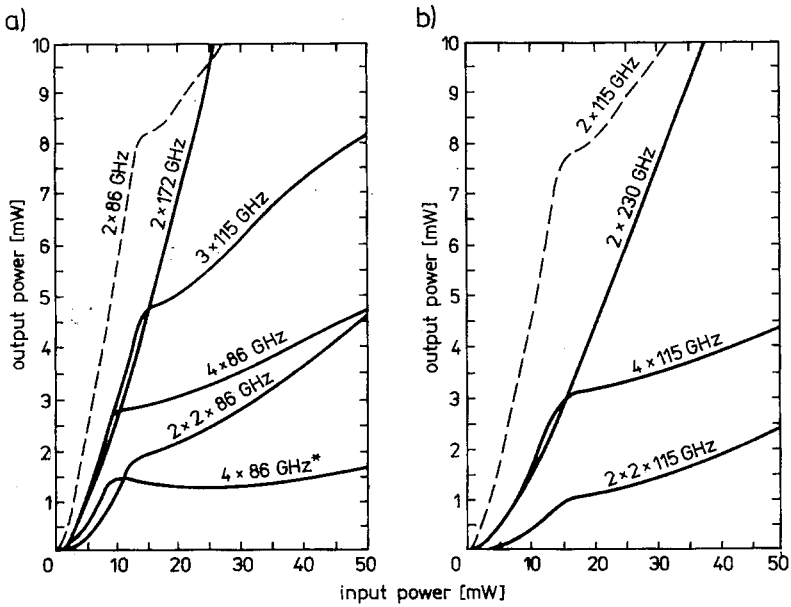


Fig. 17. A comparison of the highest theoretical output power available from frequency multipliers at: (a) — 345 GHz and (b) — 460 GHz. After: T. Tolmunen, A. Räisänen, M. Sironen: *High Order Frequency Multipliers — a Solution for mm and Submm-Wave Local Oscillator Signal Generation.* 1989 IEEE MTT-S Int. Micowave Symp. Digest, Long Beach 1989, pp. 1223–1226

At 800 GHz situation is different [39] because of lack of power sources in the 300–400 GHz range capable of pumping a single doubler or tripler. Comparison of the theoretical performance of various multipliers producing 800 GHz is given in Figure 18. At 50 mW input power a $4 \times 2 \times 100$ GHz cascaded multiplier gives by far the best result. The octupler is also clearly more efficient than the other two configurations. These performances are purely theoretical, because no metal losses are assumed and all ideal idler termination are assumed in the higher order multipliers. But when nonidealities are reasonably taken into account, the cascaded multiplier $4 \times 4 \times 100$ GHz is expected in [39] to be still by far the best and should produce about 100 μW at 800 GHz.

No experimental multiplier chain has yet reached an output power of 1 mW at 500 GHz or 100 μW at 700 GHz. Multiplier chains where a doubler is followed by a tripler are in practice more effective than chains having a tripler followed by a doubler [17]. Up to 500 GHz the direct higher-order multipliers compete quite evenly with the cascaded low-order multipliers. At frequencies above 500 GHz the power level of 100 μW has not yet been reached with direct high order multipliers. This is mainly because

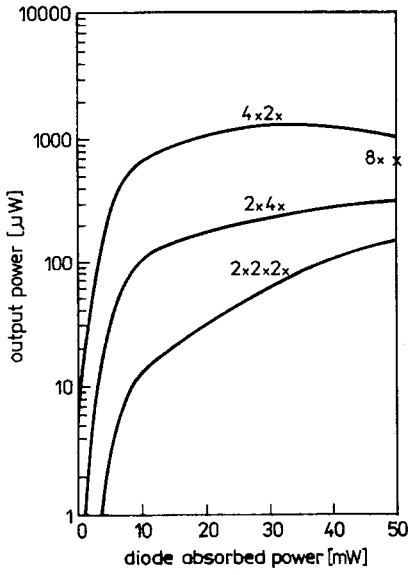


Fig. 18. Comparison of theoretical performance of varius multipliers producing 800 GHz. After: A. Räisänen, M. Sironen: *Capability of Schottky-Diode Multipliers as Local Oscillators at 1 THz.* Microwave and Optical Technology Letters, Vol. 4, No. 1, pp. 29–33, Jan. 1991

of extreme difficulty to optimize all the necessary idler terminations, and because an effective diode for this purpose can not absorb high input powers without damage.

If the additional cost of designing, manufacturing, and testing separate components and their interconnecting hardware is also considered the decision whether to use the higher-order multiplier or the multiplier chain is even more difficult to make. It must be decided individually in each particular case and universal prescriptions may not be given.

REFERENCES

1. P. Penfield, R.P. Rafuse: *Varactor Applications.* The MIT Press, Cambridge, 1962
2. C.C.H. Tang: *An Exact Analysis of Varactor Frequency Multiplier.* IEEE Trans. Microwave Theory Tech., Vol. MTT-14, No. 4, pp. 210–212, April 1966
3. M. Uenohara, J.W. Gewartowski: *Varactor Applications.* Microwave Semiconductor Devices and Their Circuit Applications, pp. 194–270, H.A. Watson, Ed., McGraw–Hill, New York 1969
4. S.A. Mass: *Nonlinear Microwave Circuits,* Artech House, Norwood 1988
5. C.B. Burckhardt: *Analysis of Varactor Frequency Multipliers for Arbitrary Capacitance Variation and Drive Level.* Bell Syst. Tech., J. Vol. 44, No. 4, pp. 675–692, April 1965
6. J.O. Scanlan: *Analysis of Varactor Harmonic Generators.* Advances in Microwaves, L. Young, Ed., Vol. 2, pp. 165–236, Academic Press, New York 1967

7. E. Boch: *Design of a Rugged Millimeter-Wave Doubler Using a Series Varactor Configuration.* 1988 IEEE MTT-S Int. Microwave Symp. Digest, New York 1988, pp. 785–787
8. S.-W. Chen et al.: *A High-Performance 94-GHz MMIC Doubler.* IEEE Microwave and Guided Wave Letters, Vol. 3, No. 6, pp. 167–169, June 1993
9. S.-W. Chen, T.C. Ho, K. Pande, P.D. Rice: *Rigorous Analysis and Design of a High-Performance 94 GHz MMIC Doubler.* IEEE Trans. Microwave Theory Tech., Vol. MTT-41, No. 12, pp. 2317–2322, Dec. 1993
10. T.C. Ho, S.-W. Chen, K. Pande, P.D. Rice: *A W-Band Integrated Power Module Using MMIC MESFET Power Amplifiers and Varactor Doublers.* IEEE Trans. Microwave Theory Tech., Vol. MTT-41 No. 12, pp. 2288–2294, Dec. 1993
11. N.R. Erickson: *A High Efficiency Frequency Tripler for 230 GHz.* Proc. 12th European Microwave Conf., Helsinki 1982, pp. 288–292
12. N.R. Erickson: *High Efficiency Millimeter and Submillimeter Frequency Multipliers.* Proc. 8th Int. Conf. Infrared and Millimeter Waves, Miami 1983, paper M3.2
13. N.R. Erickson: *Very HIgh Efficiency Triplers for 100–300 GHz.* Proc. 10th Int. Conf. Infrared and Millimeter Waves, 1985, pp. 54–55
14. N.R. Erickson: *High Efficiency Submillimeter Frequency Multipliers.* 1990 IEEE MTT-S Int. Microwave Symp. Digest, Dallas 1990, pp. 1301–1304
15. T.J. Tolmunen, A.V. Räisänen: *An Efficient Schottky-Varactor Frequency Multiplier at Millimeter-Waves, Part I: Doubler.* Int. J. Infrared Millimeter Waves, Vol. 8, No. 10, pp. 1313–1336, Oct. 1987; *Part II: Tripler.* Int. J. Infrared Millimeter Waves, Vol. 8, No. 10, pp. 1337–1353, Oct. 1987; *Part III: Quadrupler.* Int. J. Infrared Millimeter Waves, Vol. 10, No. 4, pp. 475–504, April 1989; *Part IV: Quintupler.* Int. J. Infrared Millimeter Waves, Vol. 10, No. 4, pp. 505–518, April 1989
16. A. Rydberg, B.N. Lyons, S.V. Lidholm: *On the Development of a High Efficiency 750 GHz Frequency Tripler for THz Heterodyne Systems.* IEEE Trans. Microwave Theory Tech., Vol. MTT-40, No. 5, pp. 827–830, May 1992
17. A.V. Räisänen: *Frequency Multipliers for Millimeter and Submillimeter Wavelengths.* Proc. IEEE, Vol. 80, No. 11, pp. 1842–1852, Nov. 1992
18. T. Takoda et al.: *Hybrid Inegrated Frequency Multipliers at 300 and 450 GHz.* IEEE Trans. Microwave Theory Tech., Vol. MTT-26, No. 10, pp. 733–737, Oct. 1978; and *Frequency Triplers and Quadruplers with GaAs Schottky-Barrier Diodes at 450 and 600 GHz.* IEEE Trans. Microwave Theory Tech., Vol. MTT-27, No. 5, pp. 519–523, May 1979
19. J.W. Archer: *Millmeter Wavelength Frequency Multipliers.* IEEE Trans. Microwave Theory Tech., Vol. MTT-29, No. 6, pp. 552–557, June 1981
20. J.W. Archer: *A High Performance Frequency Doubler for 80 to 120 GHz.* IEEE Trans. Microwave Theory Tech., Vol. MTT-30, No. 5, pp. 824–825, May 1982
21. J.W. Archer: *An Efficient 200–290 GHz Frequency Tripler Incorporating a Novel Stripline Structure.* IEEE Trans. Microwave Theory Tech., Vol. MTT-32, No. 4, pp. 416–420, Apr. 1984
22. J.W. Archer: *A Novel Quasi-Optical Frequency Multipliers Design for Millimeter and Submillimeter Wavelengths.* IEEE Trans. Microwave Theory Tech., Vol. MTT-32, No. 4, pp. 421–427, Apr. 1984
23. M.T. Faber, J.W. Archer, R.J. Mattauch: *A High Efficiency Frequency Doubler for 100 GHz.* 1985 IEEE MTT-S Int. Microwave Symp. Digest, St. Louis 1985, pp. 363–366
24. M.T. Faber, J.W. Archer, R.J. Mattauch: *A Frequency Doubler with 35% Efficiency at W Band.* Microwave Journal, Vol. 28, No. 7, pp. 145–152, July 1985
25. J.W. Archer, M.T. Faber: *High-Output, Single- and Dual-Diode, Millimeter Wave Frequency Doublers.* IEEE Trans. Microwave Theory Tech., Vol. MTT-33, No. 6, pp. 533–538, June 1985
26. J.W. Archer, R.A. Batchelor: *Multipliers and Parametric Devices.* Handbook of Microwave and Optical Components, vol. 2, pp. 142–191, K. Chang, Ed., John Wiley and Sons, New York 1990
27. D. Choudhury, M.A. Frerking, P.D. Batelaan: *A 200 GHz Tripler Using a Single Barrier Varactor.* IEEE Trans. Microwave Theory Tech., Vol. MTT-41, No. 4, pp. 595–599, Apr. 1993
28. J.A. Calviello: *Advanced Devices and Components for the Millimeter and Submillimeter Systems.* IEEE Trans. Electron Devices, vol. ED-26, no. 9, pp. 1273–1281, Sept. 1979

29. A.V. Räisänen et al.: *A Novel Split-Waveguide Mount Design for Millimeter- and Submillimeter-Wave Frequency Multipliers and Harmonic Mixers.* IEEE Microwave and Guided Wave Letters, Vol. 3, No. 10, pp. 369–371, Oct. 1993
30. P.W. Staecker, M.E. Hines, F. Occhiuti, J.F. Cushman: *Multi Watt Power Generation at Millimeter-Wave Frequencies Using Epitaxially-Stacked Varactor Diodes.* 1987 IEEE MTT-S Int. Microwave Symp. Digest, Las Vegas 1987, pp. 917–920
31. J.F. Cushman, F. Occhiuti, E.M. McDonagh, M.E. Hines, P.W. Staecker: *High Power Epitaxially-Stacked Varactor Diode Multipliers: Performance and Applications at W-band.* 1990 IEEE MTT-S Int. Microwave Symp. Digest, Dallas 1990, pp. 923–926
32. D.F. Peterson: *The Varactor Power Frequency Multiplir — a Device for Quietly Extending the Frequency Range of Microwave Power Sources.* Microwave Journal, Vol. 33, N. 5, pp. 135–146, May 1990
33. R.J. Hwu et al.: *Watt-Level Millimeter-Wave Monolithic Diode-Grid Frequency Multipliers,* 1988 IEEE MTT-S Int. Microwave Symp. Digest, New York 1988, pp. 533–536
34. Ch.F. Jou et al.: *Millimeter-Wave Diode-Grid Frequency Doubler.* IEEE Trans. Microwave Theory Tech., Vol. MTT-36, No. 11, pp. 1507–1514, Nov. 1988
35. R.J. Hwu et al.: *Quasi-Optical Watt-Level Millimeter-Wave Monolithic Solid-State Diode-Grid Frequency Multipliers.* 1989 IEEE MTT-S Int. Microwave Symp. Digest, Long Beach 1989, pp. 1069–1072
36. B.J. Rizzi, T.W. Crowe, N.R. Erickson: *A High-Power Millimeter-Wave Frequency Doubler Using a Planar Diode Array.* IEEE Microwave and Guided Wave Letters, Vol. 3, No. 6, pp. 188–190, June 1993
37. M. Sironen, T. Tolmunen, A. Räisänen: *Comparison of Higher-Order Multipliers to Cascaded Doublers and Triplers in Submillimeter Signal Generation.* Proc. 19th European Microwave Conf., London 1987, pp. 464–469
38. T. Tolmunen, A. Räisänen, M. Sironen: *High Order Frequency Multipliers — a Solution for mm and Submm-Wave Local Oscillator Signal Generation.* 1989 IEEE MTT-S Int. Microwave Symp. Digest, Long Beach 1989, pp. 1223–1226
39. A. Räisänen, M. Sironen: *Capability of Schottky-diode Multipliers as Local Oscillators at 1 THz.* Microwave and Optical Technology Letters, Vol. 4, No. 1, pp. 29–33, Jan. 1991

M.T. FABER

DIODOWE POWIELACZE CZĘSTOTLIWOŚCI ZAKRESÓW FAL MILIMETROWYCH I SUBMILIMETROWYCH

Streszczenie

Projektowanie i realizacja powielaczy częstotliwości zakresu fal milimetrowych wciąż jeszcze jest, z uwagi na ich złożoność, bardziej sztuką niż praktyką inżynierską i zwykle wiąże się z doświadczalną optymalizacją układów i konstrukcji powielaczy. Transformowanie wyników teoretycznych w faktyczną konstrukcję mikrofalową jest w tym procesie etapem krytycznym, często decydującym o sukcesie lub porażce całego przedsięwzięcia. Złagodzeniu tych trudności i zredukowaniu ryzyka porażki ma służyć przedstawiony w tym artykule obszerny i systematyczny przegląd rozwiązań konstrukcyjnych diodowych powielaczy częstotliwości będących szczytowymi osiągnięciami światowymi. Specyficzne realizacje mikrofalowe przedstawiane są tu nie tylko by wskazać na istniejące ograniczenia i niezbędne kompromisy układowe, ale także dla ilustracji celowości konstrukcji i potencjalnych możliwości współczesnych diodowych powielaczy częstotliwości zakresów fal milimetrowych i submilimetrowych.

Słowa kluczowe: powielacz częstotliwości, waraktory $p-n$, waraktory Schottkyego, układy falowodowe, układy quasi-optyczne, układy planetarne

INFORMACJE DLA AUTORÓW

Redakcja przyjmuje do publikowania prace oryginalne, przeglądowe i monograficzne wchodzące w zakres szeroko pojetej elektroniki. Ponieważ KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI jest czasopismem Komitetu Elektroniki i Telekomunikacji Polskiej Akademii Nauk, w związku z tym na jego łamach znajdują się prace naukowe dotyczące podstaw teoretycznych i zastosowań z zakresu elektroniki, telekomunikacji, mikroelektroniki, optoelektroniki, radiotechniki i elektroniki medycznej.

Artykuły powinno charakteryzować oryginalne ujęcie zagadnienia, własna klasyfikacja, krytyczna ocena (teorii lub metod), omówienie aktualnego stanu, lub postępu danej gałęzi techniki oraz omówienie perspektyw rozwojowych.

Artykuły publikowane w innych czasopismach nie mogą być kierowane do druku w Kwartalniku Elektroniki i Telekomunikacji w drugiej kolejności zgłoszenia.

Objętość artykułu nie powinna przekraczać 20 stron po około 1800 znaków na stronie.

Wymagania podstawowe: Artykuły należy nadsyłać w maszynopisie pisanym jednostronnie lub na wyraźnym czarno-białym wydruku komputerowym, w 2 egzemplarzach, w języku polskim lub angielskim — wybranym przez autora. Tekst artykułu musi być poprzedzony tytułem pracy, imieniem i nazwiskiem Autora wraz z podaniem miejsca jego pracy. Wszystkie strony muszą mieć numerację ciągłą.

Sposob pisania tekstu: Tekst powinien być pisany bez używania wyróżnień, a w szczególności nie dopuszcza się spacjowania, podkreślania pisania tekstu dużymi literami z wyjątkiem wyrazów, które umownie pisze się dużymi literami (np. FORTRAN). Proponowane wyróżnienia Autor może zaznaczyć w maszynopisie zwykłym ołówkiem za pomocą przyjętych znaków adiustacyjnych) np. podkreślenie linią przerywaną oznacza spacjowanie (rozstrzelenie), podkreślenie linią ciągłą — pogrubienie, podkreślenie wężykiem — kursywa. Tekst powinen być napisany z podwójnm odstępem między wierszami, tytuły i podtytuły malymi literami. Marginesy z każdej strony powinny mieć około 35 mm. Przy podziale pracy na rózdziały i podrozdziały cyfrowe ich oznaczenia nie powinny być większe niż III stopnia (np. 4.1.1).

Sposób pisania tablic: Tablice powinny być pisane na oddzielnych stronach. Tytuły rubryk pionowych i pozimych powinny być napisane małymi literami z podwójnym odstępem między wierszami. Przypisy (notki) dotyczące tablic należy pisać bezpośrednio pod tablicą. Tablice należy numerować kolejno liczbami arabskimi, u góry każdej tablicy podać tytuł. Tablice umieścić na końcu maszynopisu. Przyjmowane są tablice algorytmów i programy na wydrukach komputerowych. W tym przypadku zachowany jest ich oryginaly układ.

Sposób pisania wzorów matematycznych; Rozmieszczenie znaków, cyfr, liter i odstępów powinno być zbliżone do rozmieszczenia elementów druku. Wskaźniki i wykładniki potęg powinny być napisane wyraźnie i być prawidłowo obniżone lub podwyższone w stosunku do linii wiersza podstawowego. Znaki nad literami i cyframi, całkami i in. symbolami (strzalki, linie, kropki, daszki) powinny być umieszczone dokładnie nad tymi elementami, do których się odnoszą. Numery wzorów cyframi arabskimi powinny być kolejne i umieszczone w nawiasach okrągłych z prawej strony. Nazwy jednostek, symbole litrowe i graficzne powinny być zgodne z wytycznymi IEC (International Electronical Commission) oraz ISO (International Organization of Standarization).

Powołania: Powołania na publikacje powinny być umieszczone na ostatnich stronach tekstu pod tytułem „Bibliografia", opatrzone numeracją kolejną bez nawiasów. Numeracja ta powinna być zgodna z odnośnikami w treści artykułu. Przykłady opisu publikacji:

- — periodycznej: F.Valdoni: A new millimetre wave satelite. E.T.T. 1990, vol. 2, nr 5, p. 553
- — nieperiodycznej: K. Andersen: A resourse allocation framework. XVI International Symposium. Stockholm (Sweden), May 1991, paper A2.4
- — książki: Y.P. Tvidis: Operation and modelling of the MOS transistors. Mc Graw-Hill, New York 1987, p. 141 – 148.

Material ilustracyjny: Rysunki powinny być wykonane wyraźnie, na papierze gładkim, lub milimetrowym w formacie nie mniejszym niż 9x12 cm. Mogą być także w postaci wydruku komputerowego.

Fotografie lub diapozytywy przyjmowane są kolorowe lub czarno-białe w formacie nie przekraczającym 10×15 cm. Na marginesie każdego rysunku i na odwrocie fotografii powinno być napisane ołówkiem imię i nazwisko Autora oraz skrót tytułu artykułu, do którego są przeznaczone.

Spis podpisów pod rysunki i fotografie powinen być umieszczony na oddzielnej stronie.

Streszczenie. Do każdego artykułu musi być dołączone streszczenie z tytułem artykułu. Streszczenia oraz tytuły (Summary) muszą być w języku polskim i angielskim. Streszczenie powinno wyjaśniać główny cel pracy, wskazywać korzyści i ograniczenia, możliwe zastosowania i zalecenia dla dalszego rozwoju danej gałęzi techniki. Objętość streszczenia nie powinna przekraczać 100 słów, a jego treść nie może być identyczna ze „Wstępem", lub „Zakończeniem". Pod streszczeniami powinny być podane słowa kluczowe.

Autorowi przysługuje bezpłatnie 20 odbitek artykułu. Dodatkowe egzemplarze odbitek, lub cały zeszyt Autor może zamówić u wydawcy na własny koszt.

Autora obowiązuje korekta autorska, którą powinien wykonać w ciąu 3 dni od daty otrzymania tekstu z Redakcji i zwrócić osobiście, lub listownie pod adresem Redakci. Korekta powinna być naniesiona na przekazanych Autorowi szpaltach na marginesach ew. na osobnym arkuszu w przypadku uzupełnień tekstu większych niż dwa wiersze. W przypadku nie zwrócenia korekty w terminie, korektę przeprowadza Redakcja Techniczna Wydawcy.

Redakcja prosi Autorów o powiadomienie jej o zmianie miejsca pracy i adresu prywatnego.,

Uwaga: Od 1995 r. Redakcja przyjmuje teksty gładkie na dyskietkach z plikiem w formacie ASCI. Rysunki w formacie TIF lbu PCX. Wydruk tekstu traktowany jako podstawa powinien zawierać również wzory matematyczne.

SPIS TREŚCI

CONTENTS

INFORMATION
for „Kwartalnik Elektroniki i Telekomunikacji (KEiT) — Electronics and Telecommunications Quarterly" Authors

The KEiT is a journal of the Committee for Electronics and „Telecommunication of the Polish Academy of Sciences. Its pages contain scientific articles concerned with theory and application of electronics, telecommunications, microelectronics, optoelectronics, radioenginering and medial electronics.

Regular papers are accepted for publication in Polish or Engish. The two copies of manuscript are requested for examination and processing. Each copy should a complete set of illustrations. The complete manuscript including title, bilingual — english and polish abstract, key words, text, references, ilustrations and figure captions shold not exceed 20 typewritten pages. The manuscript should be typed on side of the paper only double spacing between the lines and a 3 cm margin. The title should indicate the subject of the manuscript as briefly possible. The abstract — from 30 to 100 words-should state the contents of the paper on a self-explanatory manner, so that it may be published in abstracts journal, separately from paper. Key words are required. In manuscript containing lot of symbol, a glossary may be included after introduction.

Text

- All text pages should be numbered.
- The first page should mention the title in full, surname of the Author's titles and functions, their affilations and relevant addresses.
- For letteral symbols, units and graphical symbols, the recommendations of IEE (International Electrotechnical Commision) and ISO (International Organization of Standarization) must be follwed.
- Formulae and footnotes should be numbered in accordance with the quotation order followed in the text.
- The serial number of each formula should be written at right side between round brackets (i.e. (1), (2), etc.).
- The serial number of footnotes should be written between round brackets and upper (i.e. $(^1)$, $(^2)$, etc.).
- References are usually gathered at the end of the text and numbered according to the order of quotation in the text, written without the brackets.

Illustrations:

The illustrations may be drawins in black ink or photograps.

- Illustrations must report besides the Author(s) name, their order numbers (All illustrations should be numbered following the order of quotation in the text, irrespective of photographs and drawings.
- Since most illustrations will be reducted in size to 8 cm width, alphanumeric characters must be written so to be legible after reduction.
- Graphs should be contained into rectangles with four sides quoted.(The use of graph paper of similar should be avoided, except for graphing recorders).
- Photographs (coloured or black and with) must not exceed 10×15 cm.

Appendixes

- Mathematical details, ancillary to the main of the paper, may be included in one more appendixes, that are printed after conclusions and before references.
- In manuscripts containing lot of symbols, a glossary may be included after introduction.

SEMINARIUM Z PODSTAW ELEKTROTECHNIKI I TEORII OBWODÓW
19-TH SEMINAR ON FUNDAMENTALS OF ELECTROTECHNICS AND CIRCUIT THEORY

Politechnika Śląska
Wydział Elektryczny
ul. Akademicka 10
44-100 GLIWICE POLAND
tel. (48) (32) 371097
fax: (48) (32) 371655
E-mail: speto @ zeus.gliwice.edu.pl

W maju 1996 roku odbędzie się tradycyjne Seminarium z Podstaw Elektrotechniki i Teorii Obwodów - **XIX SPETO'96**

Osoby zainteresowane udziałem w XIX SPETO'96 mogą uzyskać informacje pod adresem:

POLITECHNIKA ŚLĄSKA
WYDZIAŁ ELEKTRYCZNY

SPETO
Dr Magdalena UMIŃSKA-BORTLICZEK

ul. Akademicka 10
44-100 GLIWICE POLAND
Fax (48)(32) 371655, Tel. (48) (32) 37-10-97
E-mail: speto@zeus.gliwice.edu.pl

Bardzo prosimy o używanie poczty komputerowej.

KALENDARIUM

DO 29 PAŹDZIERNIKA 1995	-	ZGŁOSZENIE W FORMIE KRÓTKIEGO STRESZCZENIA
DO 20 LISTOPADA 1995	-	INFORMACJA O AKCEPTACJI PROPONOWANEGO TYTUŁU PRACY ORAZ INFORMACJE REDAKCYJNE
DO 2 LUTEGO 1996	-	WYSYŁKA PRAC W JĘZYKU POLSKIM LUB ANGIELSKIM [DRUK DWUSZPALTOWY, STRON **4**, FORMAT **A4**]. KAŻDA PRACA MUSI BYĆ POPRZEDZONA STRESZCZENIEM W JĘZYKU ANGIELSKIM (minimum **200** słów). WSZYSTKIE PRACE SĄ RECENZOWANE I PODLEGAJĄ ZATWIERDZENIU PRZEZ KOMITET PROGRAMOWY.
1-10 KWIETNIA 1996	-	WYSYŁKA INFORMACJI O PRZYJĘCIU PRACY

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

T. 41, Z. 3 (1995)

Czasopismo jest przeznaczone dla specjalistów oraz studentów zajmujących się tematyką dotyczącą podstaw teoretycznych i zastosowań z zakresu elektroniki, telekomunikacji, mikroelektroniki, optoelektroniki, radiotechniki i elektroniki medycznej. Zeszyt nr 3/95 zawiera opis nowej, efektywnej implementacji metody Braytona-Gustavsona-Hatchela (BGH) rozwiązywania równań różniczkowych zwyczajnych zalecaną dla równań sztywnych do języka symulacyjnego AMIL a także prace związane z modelowaniem i identyfikacją strukturalną wielowymiarowych systemów stacjonarnych opisanych za pomocą operatorów. Ponadto w bieżącym zeszycie zamieszczone są artykuły nt.: „Metody Symulacji komputerowej analogowych filtrów scalonych w technologii MOS", „Projektowanie scalonych wzmacniaczy operacyjnych w technologii CMOS", „Projektowanie systemów wieloprocesorowych z procesorami sygnałówymi" i obszerna praca dotycząca „metodyki charakteryzowania wad rur na podstawie sygnałów przetworników wiroprądowych".

Kwartalnik jest rozprowadzany głównie w prenumeracie, jadnakże poszczególne numery można kupić w Księgarni Wydawnictwa Naukowego PWN, ul. Miodowa 10, 00-251 Warszawa. Można je również otrzymać w redakcji: ul. Nowowiejska 15/19, 00-665 Warszawa, pok. 470.

Kwartalnik jest wydawany z funduszu Polskiej Akademii Nauk i dofinansowany z instytucji:

— Politechnika Warszawska, Wydział Elektroniki i Technik Informacyjnych
— Politechnika Gdańska, Wydział Elektryczny
— Instytut Łączności w Warszawie
— Spółka „Telekomunikacja Polska S.A."

Subscription for external subscribers:

The promotional subscription price in 1995 is $ 90 including postage for institutions. Subscriptions should be sent to the publisher, Polish Scientific Publishers PWN Ltd, Journal Division, Miodowa 10, 00-250 Warsaw, POLAND, fax (48) (22) 26 09 50, (48) (22) 26 71 63, with a copy by fast to Electronics and Telecommunications Quarterly 00-665 Warsaw, ul. Nowowiejska 15/19, p. 470. Subscription is occupated after showing a checque or transfer documents. Our bank account is as follows: Bank account: PBK VIII O/Warszawa nr 370028-1052. At subscripber's request this journal will be air mailed at additional postage to 50% dross price to European countries and 65% overseas.

Subscription orders available through the local press distribution or through the Foreign Trade Enterprise ARS Polona, 00-068 Warszawa, Krakowskie Przedmieście 7, Poland.

Ruch S. A. fulfils foregin customers orders, starting from any issue in the calendar year: tel.: (48)(22) 620 10 39; fax (48)(22) 620 17 62.

Cena 5 zł / 50 000 zł

Kwartalnik Elektroniki i Telekomunikacji

WARUNKI PRENUMERATY

Wpłaty na prenumeratę przyjmowane są na okresy roczne

Na teren kraju prenumeratę przyjmują jednostki kolportażowe RUCH S.A., urzędy pocztowe oddawcze właściwe dla miejsca zamieszkania lub siedziby prenumeratora oraz doręczyciele w miejscowościach, gdzie dostęp do urzędu jest utrudniony, a także Zakład Kolportażu Wydawnictwa SIGMA-NOT

Terminy przyjmowania przez RUCH S.A. prenumeraty krajowej i na zagranicę oraz przez Pocztę Polską (tylko prenumerata krajowa) są kwartalne: do 20 XI na I kw., do 20 II na II kw., do 20 V na III kw., i do 20 VIII na IV kw.
Zakład Kolportażu Wydawnictwa SIGMA-NOT przyjmuje prenumeratę do 5 XII roku poprzedzającego

Dostawa zamówionej prasy następuje w sposób uzgodniony z zamawiającym, pod wskazanym adresem, w ramach opłaconej prenumeraty.

Adresy przedsiębiorstw kolportażowych i ich konta bankowe:

Zakład Kolportażu Wydawnictwa SIGMA-NOT, 00-716 Warszawa, ul. Bartycka 20, skr. poczt. 1004
Konto PBK III Oddział Warszawa nr 370015-1573-139-11
Prenumerata ze zleceniem wysyłki za granicę jest dwukrotnie wyższa od ceny krajowej. Zlecający powinien podać dokładny adres odbiorcy. Dodatkowe informacje można otrzymać pod nr telefonu 49-30-86 lub 40-00-21 wew. 249, 295, 299.

Ruch S.A. Oddział Warszawa, 00-958 Warszawa, ul. Towarowa 28
Konto PBK XIII Oddział Warszawa nr 370044-1195-139-11
Dostawa odbywa się przesyłką zwykłą w ramach opłaconej prenumeraty, z wyjątkiem zlecenia dostawy pocztą lotniczą, której koszt w pełni pokrywa zleceniodawca. Prenumerata ze zleceniem dostawy za granicę jest o 100% wyższa od krajowej. Od osób lub instytucji, zamieszkałych lub mieszczących się w miejscowościach, w których nie ma jednostek kolportażowych RUCH prenumeratę Kwartalnika można zamówić w Oddziale Warszawskim RUCH na konto: PBK XIII Oddział Warszawa, nr 370044-1195-139-11.

Bieżące numery można nabyć w **Księgarni Wydawnictwa Naukowego PWN,** ul. Miodowa 10, 00-251 Warszawa. Również można je nabyć, a także zamówić (przesyłka za zaliczeniem pocztowym) we Wzorcowni Ośrodka Rozpowszechniania Wydawnictw Naukowych PAN, Pałac Kultury i Nauki, 00-901 Warszawa w cenie podanej na okładce kwartalnika.

W roku 1996	
przewidywana cena 1 egz. wyniesie	7,00 zł (70.000 zł)
prenumerata roczna w kraju	28,00 zł (280.000 zł)
prenumerata roczna za granicę	90 $ USA

POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

Indeks 363189
ISSN 0867-6747

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

TOM 41 — ZESZYT 3

WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1995

Kwartalnik jest wydawany z funduszu Komitetu Badań Naukowych i dofinansowany przez instytucje:

— Politechnika Warszawska, Wydział Elektroniki i Technik Informacyjnych
— Politechnika Gdańska, Wydział Elektryczny
— Politechnika Szczecińska, Instytut Elektroniki i Informatyki
— Instytut Łączności w Warszawie
— Spółka „Telekomunikacja Polska S.A."

POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

TOM 41 — ZESZYT 3

WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1995

ADRES REDAKCJI
00-665 Warszawa, ul. Nowowiejska 15/19 Politechnika, pok. 470
Instytut Telekomunikacji, Gmach im. prof. JANUSZA GROSZKOWSKIEGO

Dyżury Redakcji: środy i piątki, godz. 14—16
tel. 660 77 37; 628 89 81; 25 29 18+aut. sekr.

Telefony domowe: Redaktora Naczelnego: 12 17 65
Zast. Red. Naczelnego: 26 83 41
Sekretarza Odpowiedzialnego: 25 29 18

WYDAWNICTWO NAUKOWE PWN
Warszawa, ul. Miodowa 10

Ark. wyd. 9.25 Ark. druk. 7.75	Podpisano do druku w listopadzie 1995 r.
Papier offsetowy kl. III 70 g. B-1	Druk ukończono w grudniu 1995 r.

Skład: Agnieszka Chmielewska Warszawa ul. Husarii 12
Druk i oprawa: Drukarnia Braci Grodzickich, Żabieniec, ul. Przelotowa 7

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 3, ss. 293–303

New efficient implementation of Brayton–Gustavson–Hatchel method for solving of stiff ordinary differential equation systems

MAREK STABROWSKI

Instytut Elektrotechniki Teoretycznej i Miernictwa Elektrycznego, Politechnika Warszawska

Otrzymano 1995.06.21

Autoryzowano do druku 1995.07.19

Brayton–Gustavson–Hatchel (BGH) method for solving of stiff ordinary differential equations belongs to the group of backward difference formulas methods. Basic details of BGH method have been presented. New implementation with original modifications has been described. Special attention has been paid to reduction of operation count and improvement of error control. Experimental tests (to be reported in another paper) have proven spectacular superiority of BGH method with respect to classic Gear method. BGH method will be incorporated as prefered method for stiff equations in AMIL simulation language.

Key words: differential equation systems, errors control, simulation languages.

1. WHY NEW SOFTWARE FOR STIFF ODE?

Gear, or rather Gear-Nordsieck, method [4, 5, 6] is at present a classic established tool for solving of stiff ordinary differential equations (ODE). It is a time-proven method with satisfying stability and efficiency (speed). Popular MATLAB software package (and other tools basing on MATLAB) uses Gear method as a standard method for stiff problems.

The history of software and hardware engineering has shown however that suspiciously frequently inferior solutions gain wide popularity. This leads to the conclusion that it is advisable to carry out periodic revisions in the field of obvious, classic and established items. It is worth to note that two complete Fortran implementations of Gear method have been published quite early [4], helping to spread the application of this method. Critical conclusion, drawn from the last fact, has been the main driving force in the search of competitive method.

At the outset, this search has been confined to backward differentiation formulas (BDF) methods, encompassing also Gear method. It seems that potential advantages of some other competitive BDF methods have been neither fully investigated nor exploited. Special attention has been aroused by a method developed by Brayton, Gustavson and Hatchel [2]. In further course this method will be called BGH method — an obvious acronym of authors names. The authors of BGH method maintained that their method is more stable than Gear method, especially if frequent changes of internal integration step are necessary. Also operation count, according to authors of BGH method, is substantially reduced and error control may be more flexible. The BGH method and Gear-Nordsieck method are not equivalent algebraically, however appropriate modification [2] of Gear-Nordsieck method transforms it into equivalent of the BGH method. All these considerations supported the decision to select the BGH method as a potential challenger of Gear-Nordsieck method.

Known implementations of BGH method in Fortran and C [1] have very specific character. They have all features of "ad hoc" solution and no traces of true library style [7]. C implementation is just "working" software with several "go to" statements left. The method order in this implementation is constant (=2), error control simplified and not all operation count reductions fully exploited.

This paper will present comprehensively program BGH stiff, developed by the author. It is original implementation of BGH method in true library style along commonly accepted guidelines for ODE solvers development [7]. Some new features have been introduced into BGH algorithm. They include efficient flow of integration, versatile and thorough error control and moreover extensive operation count reduction.

2. BASIC PRINCIPLES OF BRAYTON–GUSTAVSON–HATCHEL (BGH) METHOD

A problem to be solved, i.e. ODE system, may be written in the form:

$$f(x,\dot{x},t)=0, \quad 0\leqslant t\leqslant T, \tag{1}$$

where f is a vector (a set of functions). Integration interval [0, T] is divided by nonequispaced (in general case) points $t_0, t_1, t_2,\ldots,t_{n-k}, t_{n+1-k},\ldots,t_N$ (integration nodes) into integral integration steps $h=\Delta t_n=t_{n+1}-t_n$.

The differential value $\dot{x}_n$ for the current integration node in BDF method is approximated with the aid of $k+1$ previous (hence "backward differentiation") values of x_{n-j} $(0\leqslant j\leqslant k)$. This leads to general linear multistep method equations

$$\sum_{j=0}^{k} a_j x_{n-j}-hb_k\dot{x}_{n-k}=0, \quad \alpha_k=1. \tag{2}$$

For $k=1$ this equation reduces to simple Euler method with backward difference.

The Gear method [4, 5] uses Nordsieck vector components

$$(x_n, h_n\dot{x}_n, 1/2h_n^2\ddot{x}_n,\ldots,(1/k!)h_n^k y_n^{(k)}) \tag{3}$$

as basic backward information.

Brayton, Gustavson and Hatchel [2, 6] have forwarded the thesis that usage of backward information in the form

$$x_{n-j}, \quad j=0,1,\ldots,k \tag{4}$$

is more efficient and leads to stable formulas, even for rapidly changing step size h.

The first step in BGH method is the predictor computation from the explicit polynomial formula

$$x_n^P = P(k,n-1,t_n) = \sum_{j=1}^{k+1} \gamma_j x_{n-j-1} \quad 1 \leqslant k \leqslant 6 \tag{5}$$

based on x_i values in previous $k+1$ nodes. The coefficients γ_j ($j=1, 2,\ldots,k+1$) depend on the step size values h_{n-j} ($j=0, 1,\ldots,k$).

This predicted values x_n^P are used as starting approximation in Newton-Raphson iterative solving of non-linear equations determining the corrector. General formula for corrector determination has also polynomial form

$$\dot{x}_n = P(k,n,\dot{x}_n) = -\frac{1}{h}\sum_{j=0}^{k} \alpha_j x_{n-j}, \quad 1 \leqslant k \leqslant 6. \tag{6}$$

It should be noted that x_n appears on the right hand side of eqs. (6) and thus it is implicit formula, i.e. a non-linear equation or rather a set of equations. The coefficients α_j in (6) depend on the step size values h_{n-j}. For constant h, there is $\alpha_j = -a_j/b_j$.

For every new step, reaching integration node n, the coefficients γ_j and α_j in (5) and (6) can be determined from the condition that the approximations in eq. (5) and (6) are fulfilled exactly for all polynomials $x=p(t)$ of degree not higher than k. This leads to a system of $k+1$ linear algebraic equations for predictor coefficients

$$H_k(n-1,t_n)\gamma = e_1, \tag{7}$$

and corrector coefficients

$$H_k(n,t_n)\alpha = e_2, \tag{8}$$

where

$$\gamma = \mathrm{col}(\gamma_1,\gamma_2,\ldots,\gamma_{k+1}) \tag{9}$$

$$e_1 = \mathrm{col}(1,0,0,\ldots,0) \tag{10}$$

and

$$\alpha = \mathrm{col}(\alpha_0,\alpha_1,\ldots,a_k) \tag{11}$$

$$e_2 = \mathrm{col}(0,1,0,\ldots,0). \tag{12}$$

Matrix H_k is a Vandermonde matrix of the form

$$H_k(j,t)=(\beta^0,\beta^1,\beta^2,\ldots,\beta^k)^T \tag{13}$$

with the elements

$$\beta^i=\mathrm{col}\left(\left(\frac{t-t_j}{h}\right)^i,\left(\frac{t-t_{j-1}}{h}\right)^i,\ldots,\left(\frac{t-t_{j-k}}{h}\right)^i\right). \tag{14}$$

It has been proved [2, 6] that main component of local truncation error for k-th order method depends on the difference of predictor and corrector

$$\varepsilon_i=h(\dot{x}_{true}(t_n)+\frac{1}{h}\sum_{i=0}^{k}\alpha_i x(t_{n-i})\equiv\frac{h}{t_n-t_{n-k-1}}(x_{n+1}-x_n^P)+0(h^{k+2}). \tag{15}$$

Here $\dot{x}_{true}$ is "*true*" (i.e. exact) value of derivative.

Algorithm of BGH method for single internal integration step, terminating with time t_n, may be summed up as follows [2, 6]:

BGH-1) determine coefficients γ from (7) and predictor from (5);

BGH-2) determine coefficients α from (8) and solve corrector equations (6) using Newton-Raphson non-linear equation solver; as initial approximation predictor values should be used;

BGH-3) compute truncation error ε_i from (15); if it does not comply with some preset value, integration step h or order k should be changed; in such case backtrack and repeat the step with changed h value;

BGH-4) complete step t_n and go to next integration node t_{n+h}, go to BGH-1).

3. FUNDAMENTAL IMPLEMENTATION DETAILS

Implementation of Zimmerman and Baker [1] follows very closely already outlined flow of computations and uses the formulas (5) to (15) in most straightforward way. Original paper by Brayton, Gustavson and Hatchel [2] recommends more efficient computation of predictor γ_i and corrector α_i coefficients. According to this paper, direct solving of equation system (7) and (8) is not necessary. It can be proved that the values of $\gamma_{i+1}(n+1,k)$ and $\alpha_{i+1}(n+1,k)$ depend on the values $\gamma_i(n,k)$ and $\alpha_i(n,k)$, i.e. for the new time point t_{n+1} they can be computed basing on the values for previous time point t_n. Index n referes here to the time point t_n and index k refers to the method order. Final updating formulas for predictor coefficients are

$$\gamma_{n+1}(n+1,k)=G1\left(\frac{t_n-t_{n-j}}{t_{n+1}-t_{n-j}}\right)\alpha_j(n,k),\quad j=1,\ldots,k \tag{16}$$

where

$$G1=-[\delta(n,k)]^{-1}\prod_{v}^{k}\left(\frac{t_{n+1}-t_{n-v}}{t_n-t_{n-1-v}}\right) \tag{17}$$

and

$$\delta(n,k)=\frac{t_n-t_{n-1}}{t_n-t_{n-k-1}}. \tag{18}$$

Analogous formulas for corrector coefficients are

$$\alpha_j(n+1),k)=\delta(n+1,k)\left(\frac{t_{n+1-j}-t_{n-k}}{t_{n+1}-t_{n+1-j}}\right)\gamma_j(n+1,k). \tag{19}$$

Moreover there is also

$$\alpha_0(n,k)=-\sum_{v=1}^{k}\alpha_v(n,k) \tag{20}$$

and

$$\gamma_1(n,k)=1-\sum_{v=2}^{k+1}\gamma_v(n,k). \tag{21}$$

Equations (16) to (21) from a set of updating (actualisation) formulas for predictor and corrector coefficients. Similar formulas can be derived [2] for the method order k change. In the implementation of BGH method described in this paper, all these formulas have been used. They reduce total operation count and they help to avoid conditioning difficulties in linear equations systems solving.

BGH method, similarly as other predictor-corrector methods, is not a self-starting one. In some implementations of this class of methods, at the start the order is set to $k=1$. After 6 steps and incrementing of method order, is fully possible to vary tghe method order in full range of stiff stability between 1 and 6. More conservative approach uses explicit method, during the start, for making several initial steps. Program BGHstiff starts with the aid of Euler predictor-corrector method, unifying to some degree both approaches.

After performing several starting steps, program BGHstiff switches to BGH method. In order to implement efficient computation of α_i and γ_i coefficients the one-dimensional arrays of internal time points t_i and internal time steps h_i are used for storing of $k+1$ values. Additionally two-dimensional array Aji stores $(k+2)^2$ values of time points differences

$$A_{ij}^{n}=t_{n-i}-t_{n-j}, \quad i,j=0,1,...,k+1. \tag{22}$$

The necessity to use these values stems from closer inspection of equations (16) to (19).

Every internal integration step in program BGHstiff starts with explicit shifting of **x** (solution) and **t** (time-independent variable) arrays. Similar explicit shifting is performed in h and A_{ij} arrays. Brayton, Gustavson and Hatchel advocate [2] treating of one-dimensional arrays **x**, **t** and **h** as some sort of circular buffers. Such buffers are only updated one element at the time and time-consuming shifting is eliminated. It is elegant solution but it incurs also some overhead of pointer manipulation. Moreover, changing of the method order k and thus changing of circular buffer size increases the

complexity of this solution. In the case of program BGHstiff the antisymmetry of A_{ij} array has been fully exploited and the shifts are performed in upper right half (including main diagonal) of the array only.

Computation of predictor — at first γ_i values and next predicted x values — follows this prologue of internal integration step. Coefficients γ_i are computed from the actualisation formulas (16) to (18). Predicted values of x variables are computed from explicit formulas (5).

Next operation is the corrector computation. It is iterative refinement procedure centred around solving of non-linear equation system with the aid of Newton-Raphson method. At first corrector coefficients α_i are computed from actualisation formula (19). Jacobian of the system of differential equation can be computed in two ways. In the first, Jacobian is computed with the aid of user supplied procedure (function). According to the versatile library style software development rules [4, 7] the Jacobian can be computed alternatively through internal finite difference approximation. In order to reduce computational operations count, Jacobian is computed only once for every call of Newton-Raphson procedure. Only right hand sides are updated in subsequent iterations. If Newton-Raphson procedure fails to converge in a prescribed iteration limit, the step size is reduced and whole corrector computation starts again. It has been found for a set of real life stiff problems that the limit of three Newton iterations suffices in most cases. As alternative to Newton-Raphson method, Brown method [3] of local linearization has been tried. It proved to be adequate in the numerical sense but its efficiency (i.e. computation time) has been significantly inferior to Newton-Raphson algorithm.

After determination of corrector, error control using formula (15) is carried out. In the error control block, current internal step is accepted or rejected. Generally the decisions about step change are made with eventual repetition of computations. In this block further decisions on method order change are also made.

Sometimes, if the internal integration node is located beyond communication node, interpolation procedure is called in order to compute output values.

4. SECURING SMOOTH INTEGRATION FLOW IN PROGRAM BGHSTIFF

Typical ODE solver clearly distinguishes internal integration steps (nodes) and communication steps. Internal integration steps determine the pace of integration and the step size depends on the decisions inside error control block. Communications steps are set by the user and the values of output variables for these time points should be computed and output in some way. Internal integration steps may vary in wide range but in contrast, communication step in typical cases remains constant. Some very versatile ODE solvers offer the possibility to hit exactly the communication time points. Proper criterion in error control block is responsible for this exact time point hitting. In such solvers, or rather in this mode of computations, the internal integration step is always lower or equal the communication step. Thus integration pace may be sometimes markedly slower and accuracy clearly overdone.

Program BGHstiff follows another [4, 5] strategy of step size control with smoother and faster integration flow. Principally BGHstiff integrates ODE system using basic accuracy criteria without any limitations of internal step size. Strictly speaking there is a possibility to set maximum limit of step size and a reasonable value should be lower than 10 communication steps. After one internal integration step an overshoot may occur, i.e. current time may exceed current communication time. Exact hit, owing to finite accuracy of floating point arithmetic, is rather an exception. In such situation the values of integrated variables are computed at communication time point t_c, using standard Lagrange interpolation formula

$$x_i(t_c) = \sum_{j=0}^{k} x_{n-j} \frac{(t_c - t_n)(t_c - t_{n-1})\ldots(t_c - t_{n-j+1})(t_c - t_{n-j-1})\ldots(t_c - t_{n-j-k})}{(t_{n-j} - t_n)(t_{n-j} - t_{n-1})\ldots(t_{n-j} - t_{n-j+1})(t_{n-j} - t_{n-j-1})\ldots(t_{n-j} - t_{n-j-k})} . \quad (23)$$

The time differences stored in A_{ij} array come very handy in this interpolation formula. The interpolation order is always co-ordinated with the current order of BGH method. Interpolated x_i(kt) values are used for output only. Integration is continued starting with the last time point reached internally. If last internal step reached beyond several communication steps, interpolation is performed several times before real integration is resumed. Thus no effective backtrack of integration takes place. Along with speeding up the integration pace, this approach improves overall numerical accuracy. It follows from the fact that the number of internal time steps is kept at minimum value and lower operation count means lower value of overall arithmetic truncation error. The interpolation errors influence only locally the variables values at communication time points. These values are not used in further integration and thus interpolation errors do not influence integration accuracy.

5. ERROR CONTROL IN PROGRAM BGHSTIFF

Error control algorithm in ODE solver influences both the computations accuracy and the computations time. It should lead to short computations time and acceptable error level. Standard rules of error control are as follows:

— if error value is lower than a preset limit, step size h is increased in order to speed up computations;

— if error value is higher than a preset limit, step size h is reduced in order to reduce local computation error,

— according to some special rules [2], necessity of method order change should be periodically checked.

Efficiency of ODE solver depends very strongly on several fine points of these general error control rules. It seems that error control strategy is of the same or at least comparable importance as basic integration algorithm. The rules controlling step size may be of the "brute force" type reducing or expanding h by the factor of 10 [4]. More sophisticated methods are based on the ratio η_k of preset acceptable error limit and actual error value ε_i. Method order change is considered in some solvers [4, 5] only

after performing integration step change. In other solvers [2] the advisability of order change is checked after some preset number of internal steps.

An error control strategy implemented in program BGHstiff is presented in simplified form below. It is a well balanced mixture of criteria found elsewhere [2] and of original ideas based on extensive experiments.

ERR-0) perform one internal integration step, i.e. computations BGH-1) to BGH-4);

ERR-1) if error ratio $\eta_k < 1$ and time step h is larger than smallest acceptable value, reduce h and repeat computations ERR-0) with new h value;

ERR-2) if error ratio η_k is significantly larger than 1 and step size h is lower than maximum preset value, increase appropriately step size h and go to step ERR-5);

ERR-3) if method order already reached maximum ($k=6$), step size h has minimum acceptable value and $\eta_k < 1$, issue warning about insufficient accuracy; nothing constructive can be done — go to step ERR-5);

ERR-4) this step is performed if current number of internal integration steps is an integer multiple of current method order k; compute the coefficients η_{k-1} and η_{k+1}, i.e. the ratio of preset and actual error values for method order equal to $k-1$ and $k+1$ respectively; if $\eta_{k+1} > 1+\Delta$ — increment method order, if $\eta_{k-1} > 1+\Delta$ — decrement order, otherwise do not change order;

ERR-5) end current internal integration step h; go to ERR-0), if next communicatoin time point not reached, otherwise perform interpolation for appropriate number of communication time points.

Some details should be presented a bit more comprehensively. During computation of error ratio η_k, eventual division by zero must be avoided. If actual error ε_i is inumerically equal to zero (e.g. lower than machine epsilon) then some small value commensurate with machine epsilon replaces ε_i.

In stage ERR-1) the time step reduction coefficient has been equal to η_k at first [1, 2]. It has been found experimentally that more conservative value of η_k^2 leads to fewer step adjustments.

Similar modification of recommendations [1, 2] regarding step size h expansion in step ERR-2) has been made. Standard value of step expansion is equal to η_k but the square root of η_k value proved to be more efficient during extensive experiments. The qualitative term "significantly larger" in ERR-2 should be translated into the value of 4, also found experimentally.

There appears a deadspace 2 in step ERR-4). Selection of small value will result in frequent changes of method order but otherwise large value of will block the changes. It seems that the value in the range betwen 0.15 and 0.4 is most appropriate. Asymmetry in the favour of order reduction is advisable.

At last the problem of order change checking in stage ERR-4) should be considered more comprehensively. Exact implementation of the introductory condition in step ERR-4) [2] may lead to rather strange randomness in the frequency of order change checking. Lets illustrate this phenomenon with a small example. Below, in the first global internal steps numbers are given. The second row contains the values of k (method order) and in the third row an asterisk "*" marks the steps with

operation ERR-4). These steps number are the integer multiples of previous method order.

59	60	61	62	63	64	65	66	67	68	69	70	71	72
3	3	3	3	4	5	5	4	4	3	2	3	3	4
	*			*	*	*			*	*	*		*

We can observe that in this example for 14 internal steps, 8 order checking operations have been performed. For the average value of method order $k=3.5$ it is far to large frequency. On the other side it is intuitively obvious that for high method order this check should be performed rather seldom, as high order is equivalent to inherent high accuracy. For low method order this check should be performed more frequently. In the case considered, three checks will be appropriate. In order to solve this problem, a local counter of time steps taken has been introduced. Order checking operation is performed if this local counter reaches the value of current method order. If it becomes once again equal to method order, order checking is performed, etc. An example analogous to the previous one will illustrate this idea. There has been inserted additional row with local counter just below global steps counter.

59	60	61	62	63	64	65	66	67	68	69	70	71	72
2	0	1	2	0	1	2	3	0	1	2	0	1	2
3	3	3	3	4	4	4	4	3	3	3	43	4	4
	*			*				*			*		

It can be observed that for 14 internal steps, 4 order checking operations have been performed. Average method order is 3.5, which leads to the conclusion about perfect tuning with order checking frequency.

6. METHOD INDEPENDENT SOFTWARE OPTIMISATION

Some steps, optimising program BGHstiff have been already presented. They are closely coupled with the details of BGH method. To this group belongs, in the first place, explicit computation (actualisation) of γ_i and α_i coefficients used by predictor and corrector. Next, Jacobian matrix is computed once only for every internal integration step. Antisymmetry of Aij matrix storing time differences $t_i - t_j$ is also fully exploited. Only right upper part of this matrix is computed, updated and shifted.

It is worth to note that also some method independent optimisations have been performed. They are concentrated on speeding-up the operations with two-dimensional arrays. At the start of program development these arrays have been declared (rather dynamically allocated) as one-dimensional. In order to make easier the traversing of such arrays in two dimensions, a handy macrodefinition has been used:

```
#define INDEXC(i,j,ndim)  [j+i*ndim]
```
(24)

First argument of this macrodefinition is equivalent to row index, second to column index and the third to total column number in C language array. This macro notation is easy to read but far from optimum. In the final form of program BGHstiff, after debugging, this macro has been replaced with explicit computation of the indexes of one-dimensional arrays. In a typical situation before a loop, scanning an array, initial index is computed from a formula similar to (24), i.e. involving multiplication and addition. Inside a loop, the index is updated through addition/subtraction — single and simple operation consuming less time than addition and multiplication.

7. SOME TESTS AND SOME CONCLUSIONS

Comprehensive presentation of the tests derived from real life will be a subject of another paper [9]. It should suffice for the present that Brayton–Gustavson–Hatchel method in new implementation has shown spectacularly better performance than Gear-Nordsieck method. The efficiency — not only computation speed but also stability and accuracy — is especially marked in the field of stiff problems, i.e. in the field being the main target of both methods.

Several improvements and modifications have been introduced in program BGHstiff — a new implementation of BGH method. All operation count reductions recommended in original paper on BGH method [2] have been realised. Next, antisymmetry of one of matrices has been taken into account. All arrays have been indexed in an efficient way. But most important is the development of fine-tuned strategy of error control influencing directly step size and method order selection. BGH method will replace Gear-Nordsieck method in AMIL language [8], an efficient simulation tool, developed by the author of this paper.

REFERENCES

1. L. Baker: *C Tools for Scientists and Engineers.* MacGraw Hill, New York 1989
2. R.K. Brayton, F.G. Gustavson, G.D. Hatchel: *A New Efficient Algorithm for Solving Differential-Algebraic Systems Using Implicit Backward Differentiation Formulas.* Proceedings of the IEEE, vol. 60, no. 1, January 1972, pp. 98–108
3. G.D. Byrne, C.A. Hall: *Numerical Solution of Systems of Nonlinear Algebraic Equations.* Academic Press, New York 1973
4. G.D. Byrne, A.C. Hindmarsh: *A Polyalgorithms for the Numerical Solution of Ordinary Differential Equations.* ACM Trans. Math. Software 1975, nr 1, pp. 71–96
5. C.W. Gear: *The Automatic Integration of Ordinary Differential Equations.* In: Information Processing 68, ed.: Morrell A.J.H., North Holland, 1968, pp. 187–193
6. G. Hall, J.M. Watt: *Modern Numerical Methods for Ordinary Differential Equations.* Clarendon Press, Oxford 1976
7. T.E. Hull: *The Validation and Comparison of Programs for Stiff Systems.* In: Willoughby R.A. (ed.): Stiff Differential Systems. Plenum Press, New York 1974

8. M.M. Stabrowski: *Execution Technology in Dynamic Systems Simulation.* Archiwum Informatyki Teoretycznej i Stosowanej, 1995 (in print)
9. M.M. Stabrowski: *Numerical Experiments with Stiff Ordinary Differential Equations.* Kwartalnik Elektroniki i Telekomunikacji (to be submitted)

M.M. STABROWSKI

NOWA EFEKTYWNA IMPLEMENTACJA METODY BRAYTONA – GUSTAVSONA – HATCHELA ROZWIĄZYWANIA SZTYWNYCH RÓWNAŃ RÓŻNICZKOWYCH ZWYCZAJNYCH

Streszczenie

Metoda rozwiązywania sztywnych równań różniczkowych zwyczajnych opracowana przez Braytona, Gustavsona i Hatchela (metoda BGH) należy do grupy metod korzystających z wzorów różnic wstecznych. Przedstawiono podstawowe zasady metody BGH. Opisano nową implementację tej metody, zawierającą wiele oryginalnych rozwiązań. Zwrócono szczególną uwagę na redukcję pracy obliczeniowej i poprawę kontroli błędów. Eksperymenty numeryczne (będą one przedmiotem innego artykułu) wykazują wyraźną wyższość metody BGH w porównaniu z klasyczną metodą Geara. Metoda BGH zostanie włączona jako metoda zalecana dla równań sztywnych do języka symulacyjnego AMIL.

Słowa kluczowe: systemy równań różniczkowych, kontrola błędow, jezyki symulacyjne.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 3, ss. 305–319

Identyfikacja szczegółowej struktury wielowymiarowych systemów stacjonarnych[c)]

GRZEGORZ CIESIELSKI

Instytut Elektrotechniki Teoretycznej, Metrologii i Materiałoznawstwa, Politechnika Łódzka

Otrzymano 1995.01.25

Autoryzowano do druku 1995.06.08

Artykuł dotyczy zagadnień związanych z identyfikacją strukturalną wielowymiarowych systemów stacjonarnych opisanych za pomocą operatorów, które w ogólnym przypadku mogą być nieliniowe. Założono, że wartości sygnałów wejściowych i wyjściowych rozważanych systemów mogą przyjmować wartości rzeczywiste lub zespolone, a zbiór sygnałów wejściowych ma strukturę przestrzeni Hilberta. Udowodniono dwa twierdzenia, które pozwalają wniknąć w szczegółową stukturę operatorów stacjonarnych oraz operatorów stacjonarnych przyczynowych. Otrzymane wyniki pozwalają na lepsze zrozumienie istoty tak złożonych odzworowań jakimi są operatory, a co za tym idzie, istoty systemów opisanych za ich pomocą.

Słowa kluczowe: identyfikacja strukturalna, szczegółowa struktura systemów, modelowanie systemów, systemy wielowymiarowe, systemy stacjonarne, operatory, sploty, przestrzeń stacjonarna, przestrzeń splotowa, przestrzeń przyczynowa

1. WSTĘP

Artykuł dotyczy identyfikacji strukturalnej wielowymiarowych systemów stacjonarnych, które dają się opisać za pomocą odwzorowań nazywanych operatorami ([9]). Systemy rozważane w artykule, jak również związane z nimi operatory, mogą być w ogólnym przypadku nieliniowe, a jedną z możliwych form ich reprezentacji jest układ równań różniczkowych, być może nieliniowych. Systemy te są wielowymiarowe, czyli mają dowolną skończoną liczbę wejść i wyjść. Sygnały na poszczególnych wejściach i wyjściach mogą przyjmować wartości rzeczywiste lub zespolone, a zbiór sygnałów wejściowych jest przestrzenią Hilberta. Przestrzeń ta może zawierać wszystkie funkcje typu wykładniczego ([9]), które jak wiadomo, są bezwzględnie transformowalne w sensie jednostronnego przekształcenia Laplace'a.

[c)] Pracę wykonano w ramach projektu badawczego Nr 8 T10C 031 09 finansowanego przez Komitet Badań Naukowych w latach 1995–1997.

Ponieważ w światowej literaturze naukowej nie podjęto udanej próby w pełni formalnego, a zarazem jednolitego podejścia do zagadnienia modelowania systemów opisanych za pomocą operatorów ([11]–[12], [15]–[16] oraz [18]–[24]), to głównym celem tego artykułu jest przedstawienie ogólnej teorii modelowania systemów, obejmującej między innymi, teorię Volterry oraz Wienera ([2]–[6]). Ponadto artykuł ten jest kontynuacją rozważań zawartych w pracy [8], w której omówione zostały ogólne struktury rozważanych tu systemów. Przypomnijmy, że struktury te wymagały, by zbiór wymuszeń był przestrzenią liniową. Obecnie, przy rozważaniu szczegółowych struktur systemów stacjonarnych, niezbędnym okaże się wprowadzenie do przestrzeni wymuszeń pewnych konstrukcji topologicznych. Na koniec warto dodać, że przedstawione tu rozważania, są uogólnieniem części rozważań zawartych w artykule [9], który dotyczył zagadnienia modelowania wielowymiarowych liniowych systemów stacjonarnych.

2. PRELIMINARIA MATEMATYCZNE

Dowolny system można traktować jako odwzorowanie zbioru U sygnałów wejściowych zwanych *wymuszeniami* w zbiór V sygnałów wyjściowych zwanych *odpowiedziami*. Będziemy dalej zakładać, że oba te zbiory wyposażone są w strukturę przestrzeni liniowej, a zatem określona jest w nich operacja dodawania sygnałów i mnożenia sygnału przez liczbę z ciała liczbowego F[1].

Niech R[2] będzie zbiorem chwil czasowych, dla których określone są sygnały z U oraz V. Załóżmy ponadto, że $m \in N$[3] jest liczbą wejść, a $n \in N$ liczbą wyjść systemu. Wartości sygnałów wejściowych należą zatem do F^m[4], a wartości sygnałów wyjściowych do F^n. Wówczas przestrzeń sygnałów wejściowych U_F[5] jest podprzestrzenią liniową przestrzeni $L_F^m(R)$[6]. Podobnie przestrzeń sygnałów wyjściowych V_F jest podprzestrzenią liniową przestrzeni $L_F^n(R)$. Ponieważ U oraz V są przestrzeniami funkcyjnymi, których elementy określone są na tej samej dziedzinie R, a dowolne odwzorowanie ϕ takich przestrzeni nazywane jest *operatorem*, to każdy system można opisać za pomocą operatora. Zauważmy, że w ogólnym przypadku operator ϕ nie musi być liniowy, a jedną z możliwych form jego reprezentacji jest układ równań różniczkowych. By uprościć dalszą dyskusję, pojęcia: system opisany za pomocą operatora ϕ i operator ϕ, będą więc traktowane jako równoważne.

---

[1] W całym artykule przez F oznaczamy ciało liczbowe. W szczególności ciało F oznacza ciało liczb rzeczywistych R lub ciało liczb zespolonych C.

[2] Zbiór X wyposażony w dowolną strukturę algebraiczą i/lub topologiczną będziemy oznaczać literą X, a zatem przez R oznaczamy zbiór liczb rzeczywistych, natomiast przez R ciało liczb rzeczywistych.

[3] Przez N oznaczamy zbiór liczb naturalnych.

[4] Dla dowolnego zbioru X przez X^n oznacamy n-krotny iloczyn kartezjański (produkt) zbioru X.

[5] Przez X_F oznaczamy przestrzeń liniową nad ciałem skalarów F. Jeżeli F wynika z kontekstu, to będzie ono pomijane w jej oznaczeniu.

[6] Przez $L_F^n(X)$ oznaczamy przestrzeń odwzorowań określonych na X o wartościach z F^n. Oznaczenia $L_F^1(X)$ oraz $L_F(X)$ będziemy traktować jako równoważne, a gdy zbiór X wynika z kontekstu, to piszemy w skrócie L_F^n lub L_F. Przestrzeń L_F^n będziemy również traktować jako n-krotny produkt przestrzeni L_F.

W modelowaniu systemów wykorzystuje się trzy elementarne operatory ([2]–[4] i [6]–[9]).

Przesunięciem funkcyjnym o $\sigma \in R^k$ nazywamy operator, który oznaczamy przez $\nabla_\sigma : L_F^m(R^k) \to L_F^m(R^k)$ oraz $\forall u \in L_F^m(R^k)$ i $\forall t \in R^k$

$$[\nabla_\sigma(u)](t) := u(t+\sigma).$$

Funkcję $\nabla_\sigma(u)$ nazywamy *przesunięciem funkcji u o* σ.

Obrotem funkcyjnym nazywamy operator, który oznaczamy przez $\nabla^* : L_F^m(R^k) \to L_F^m(R^k)$ oraz $\forall u \in L_F^m(R^k)$ i $\forall t \in R^k$

$$[\nabla^*(u)](t) := u(-t).$$

Funkcję $\nabla^*(u)$ nazywamy *obrotem funkcji u.*

Splotowym przesunięciem funkcyjnym o $\sigma \in R^k$ nazywamy operator, który oznaczamy przez $\nabla_\sigma^* : L_F^m(R^k) \to L_F^m(R^k)$ oraz $\forall u \in L_F^m(R^k)$ i $\forall t \in R^k$

$$[\nabla_\sigma^*(u)](t) := u(\sigma - t).$$

Funkcję $\nabla_\sigma^*(u)$ nazywamy *przesunięciem splotowym funkcji u o* σ.

Właściwości zdefiniowanych operatorów omówione zostały w artykule [8]. Przejdźmy teraz do definicji podstawowych właściwości operatorów, a więc i systemów opisanych za ich pomocą.

Operator ϕ na $U_F \subset L_F^m(R)$ nazywamy *stacjonarnym*, jeżeli $\forall u \in U$ i $\forall \sigma \in R$

$$\nabla_\sigma(u) \in U \quad \text{oraz} \quad \nabla_\sigma(\phi(u)) = \phi\big(\nabla_\sigma(u)\big).$$

Innymi słowy, operator jest stacjonarny, gdy odpowiedź operatora na wymuszenie u przesunięte w czasie o σ jest równa przesuniętej w czasie o σ odpowiedzi operatora na wymuszenie u. Zauważmy, że przestrzeń liniowa U musi być zbiorem zamkniętym ze względu na przesunięcie funkcyjne, a zatem zawierać musi wszystkie przesunięcia funkcyjne swoich elementów.

Operator ϕ na $U_F \subset L_F^m(R)$ nazywamy *przyczynowym*, jeżeli $\forall u,v \in U$ i $\forall t \in R$ z warunku, że

$$\forall \tau \in R \ \big((\tau \leqslant t) \Rightarrow \big(u(\tau) = v(\tau)\big)\big).$$

wynika, że

$$[\phi(u)](t) = [\phi(v)](t).$$

Operator jest więc przyczynowy, gdy wartość sygnału wyjściowego w chwili t zależy jedynie od wartości sygnału wejściowego do chwili t.

W modelowaniu systemów bardzo często napotykamy przestrzenie $L_{Fp\mu}^n$ i l_{Fpw}. Dlatego też podamy teraz definicje tych przestrzeni oraz pojęć z nimi związanych.

Załóżmy, że μ jest miarą na σ-ciele w X.

Jeżeli $1 \leqslant p < \infty$, to *przestrzenią* $L_{Fp\mu}^n(X)$ nazywamy zbiór klas równoważności ze względu na relację $\mu - Æ$[7], funkcji mierzalnych (μ-mierzalnych) $u : X \to F^n$ takich, że

[7] Przez $\mu - Æ$ oznaczamy zwrot „μ-prawie wszędzie" lub „dla μ-prawie każdego".

$$\int_X |u|^p \, du < \infty.$$

Normą w $L^n_{Fp\mu}(X)$ nazywamy funkcję, którą $\forall u \in L^n_{Fp\mu}(X)$ oznaczamy przez

$$\| u \|_p := \left(\int_X |u|^p \, d\mu \right)^{1/p}.$$

Iloczynem skalarnym w przestrzeni $L^n_{Fp\mu}(X)$ nazywamy funkcję, którą $\forall u,v \in L^n_{Fp\mu}(X)$ oznaczamy przez

$$\langle u,v \rangle := \int_X v^* u \, d\mu^{(8)}.$$

Funkcje należące do $L^n_{F1\mu}(X)$ nazywamy *całkowalnymi* (w sensie Lebesgue'a). Ponadto $\forall u := (u_i)^n_1 \in L^n_{F1\mu}(X)$

$$\int_X u \, d\mu := \left(\int_X u_i \, d\mu \right)^n_1.$$

Przestrzenią $L^n_{F\infty\mu}(X)$ nazywać będziemy zbiór klas równoważności ze względu na relację $\mu - Æ$, funkcji mierzalnych (μ-mierzalnych) $u : X \to F^n$, dla których

$$\operatorname*{supess}_{x \in X} |u(x)| < \infty,$$

gdzie

$$\operatorname*{supess}_{x \in X} |u(x)| := \inf \left\{ M \in \bar{R}_+ : \mu - Æ \, x \in X \ (|u(x)| \leqslant M) \right\}^{(9)},$$

nazywamy *istotnym kresem górnym* funkcji $|u|$. *Normą* w $L^n_{F\infty\mu}(X)$ nazywamy funkcję, którą $\forall u \in L^n_{F\infty\mu}(X)$ oznaczamy przez

$$\| u \|_\infty := \operatorname*{supess}_{x \in X} |u(x)|.$$

Funkcje należące do przestrzeni $L^n_{F\infty\mu}(X)$ nazywamy *istotnie* (μ-istotnie) *ograniczonymi* ([14], s. 53 i 59).

Załóżmy dalej, że $1 \leqslant p \leqslant \infty$. Przez $C^n_{Fp\mu}(X)$ będziemy oznaczać podprzestrzeń funkcji ciągłych z przestrzeni $L^n_{Fp\mu}(X)$.

Jeżeli zbiór X wynika z kontekstu, to piszemy w skrócie $L^n_{Fp\mu}$ lub $C^n_{Fp\mu}$. Oznaczenia $L^1_{Fp\mu}$ i $C^1_{Fp\mu}$ oraz odpowiednio $L_{Fp\mu}$ i $C_{Fp\mu}$ są równoważne. W przypadku gdy miara μ jest

---

[8] Dla dowolnej macierzy A przez A^* oznaczamy jej transpozycję ze sprzężeniem, a zatem

$$A^* := \bar{A}^T.$$

[9] Przez $\bar{R}_+$ oznaczamy zbiór liczb rzeczywistych nieujemnych, który z kolei jest oznaczany przez R_+, uzupełniony o liczbę ∞.

miarą Lebesgue'a na R^k, którą oznaczamy przez λ_k, w skrócie λ, to jej symbol jest pomijany w oznaczeniach zdefiniowanych przestrzeni, a oznaczenia: $\lambda-Æ$ oraz Æ są równoważne.

Niech $1 \leqslant p < \infty$. Przez $l_{Fp\omega}$ będziemy oznaczać *przestrzeń* szeregów o wyrazach z F zbieżnych z p-tą potęgą i wagami

$$\mathscr{W} := (w_i \in R_+)_N,$$

w której $\forall x := (x_i)_N \in l_{Fpw}$ *norma* oznaczana przez

$$\| x \|_p := \left(\sum_{i \in N} w_i |x_i|^p \right)^{1/p}.$$

Jeżeli $\forall i \in N$ wagi $w_i = 1$, to przestrzeń taką będziemy oznaczać przez l_{Fp} ([17], s. 73). Sploty funkcji są odwzorowaniami, które są elementami modeli systemów stacjonarnych.

Niech λ_k będzie miarą Lebesgue'a na R^k. *Splotem funkcji* λ_k-mierzalnych u, $v: R^k \to F^m$ nazywamy funkcję oznaczaną przez $u * v: R^k \to F$, która o ile istnieje, dla λ_k prawie każdego $t \in R^k$ spełnia warunek:

$$[u * v](t) := \int_{R^k} u^T(t-\tau)\, v(\tau)\, \mathrm{d}\tau.$$

Zwykle splot definiuje się dla funkcji o wartościach z F, a nie jak tutaj z F^m. Takie rozszerzenie pojęcia splotu wynika z jego związku z iloczynem skalarnym, o czym była mowa w artykule [9]. W artykule tym udowodniono również dwa twierdzenia dotyczące najistotniejszych właściwości zdefiniowanego tu splotu funkcji.

Przestrzenie stacjonarne, splotowe i przyczynowe odgrywają bardzo ważną rolę w modelowaniu systemów stacjonarnych.

Załóżmy, że μ jest miarą na R^k oraz $1 \leqslant p \leqslant \infty$.

Przestrzenią stacjonarną oznaczaną przez $L^m_{Fp\mu\nabla}(R^k)$, nazywamy podzbiór klas równoważności z przestrzeni $L^m_{Fp\mu}(R^k)$ ze względu na relację $\mu-Æ$ o postaci:

$$L^m_{Fp\mu\nabla}(R^k) := \left\{ u \in L^m_{Fp\mu}(R^k) : \forall t \in R \left((\nabla_t(u) \in L^m_{Fp\mu}(R)) \right) \right\}.$$

Przestrzenią splotową oznaczaną przez $L^m_{Fp\mu *}(R^k)$, nazywamy podzbiór klas równoważności przestrzeni $L^m_{Fp\mu}(R^k)$, ze względu na relację $\mu-Æ$ o postaci:

$$L^m_{Fp\mu *}(R^k) := \left\{ u \in L^m_{Fp\mu}(R^k) : \forall t \in R^k \, (\nabla^*_t(u) \in L^m_{Fp\mu}(R^k) \right\}.$$

Przestrzenią przyczynową oznaczaną przez $L^m_{Fp\mu +}(R^k)$, nazywamy podzbiór klas równoważności z przestrzeni $L^m_{Fp\mu}(R^k)$ ze względu na relację $\mu-Æ$ o postaci:

$$L^m_{Fp\mu+}(R^k) := \left\{u_{-1}u : u \in L^m_{Fp\mu}(R^k)\right\}^{(10)}.$$

Jeżeli zbior R^k wynika z kontekstu, to piszemy w skrócie $L^m_{Fp\mu\nabla}$, $L^m_{Fp\mu *}$ oraz $L^m_{Fp\mu+}$. Oznaczenia $L^1_{Fp\mu\nabla}$, $L^1_{Fp\mu *}$ i $L^1_{Fp\mu+}$ oraz odpowiednio $L_{Fp\mu\nabla}$, $L_{Fp\mu *}$ i $L_{Fp\mu+}$ są równoważne. W przypadku gdy miarą na R^k jest miara Lebesgue'a λ, to jej symbol jest pomijany w oznaczeniach zdefiniowanych przestrzeni. Właściwości przestrzeni stacjonarnej, splotowej i przyczynowej można znaleźć w artykule [9].

3. SZCZEGÓŁOWA STRUKTURA SYSTEMÓW STACJONARNYCH

Można wykazać ([8]), że dowolny operator ϕ na $L^m_{F2\times\nabla}(R)$ jest stacjonarny, wtedy i tylko wtedy, gdy

$$\phi = \varphi \circ \nabla_{\cdot},$$

gdzie przekształcenie[11]

$$\varphi := [\phi(\cdot)](0).$$

Z kolei, $\forall t \in R$ obrazem przestrzeni $L^m_{F2\times\nabla}(R)$ w odwzorowaniu ∇_t jest ta sama przestrzeń $L^m_{F2\times\nabla}(R)$, a zatem obrazem produktu $L^m_{F2\times\nabla}(R) \times R$ w odwzorowaniu $\nabla_{\cdot}$ jest również przestrzeń $L^m_{F2\times\nabla}(R)$. Jest to bardzo ważny wynik, który pozwala ograniczyć analizę struktury dowolnego stacjonarnego operatora ϕ na $L^m_{F2\times\nabla}$, do analizy struktury odpowiadającego mu przekształcenia φ również na $L^m_{F2\times\nabla}$. Przyjrzymy się zatem bliżej właściwościom tego rodzaju przekształceń.

Twierdzenie 1. *Jeżeli* $\upsilon := (\upsilon_i)_N$ *jest bazą ortonormalną*, a $\langle\cdot,\cdot\rangle_\varkappa$ *iloczynem* skalarnym w $L^m_{F2\varkappa}(R)$, μ *jest miarą* na σ-ciele *zbiorów borelowskich*[12] w $L^m_{F2\varkappa}(R)$, ν jest miarą na l_{F2},

$$\phi := \langle\cdot,\upsilon\rangle_\varkappa{}^{(13)}, \quad \nu = \mu \circ \Phi^{-1}{}^{(14)}$$

---

[10] Przez u_{-1} oznaczamy *skok jednostkowy*, który $\forall t := (t_i)_1^k \in R^k$ zdefiniowany jest przez wzór:

$$u_{-1}(t) := \begin{cases} 1, & \forall i \in \{1,\ldots,k\} \quad (t_i \geqslant 0) \\ 0 & \exists i \in \{1,\ldots,k\} \quad (t_i < 0) \end{cases}.$$

Jeżeli $f: X \to F$ oraz $g: X \to \mathrm{F}^n$, to $\forall x \in X$ *iloczyn odwzorowań* oznaczany przez fg, zdefiniowany jest za pomocą wzoru:

$$[fg](x) := f(x)\,g(x).$$

Sygnały należące do przestrzeni przyczynowej $L^m_{Fp\mu+}(R)$ nazywane są *przyczynowymi* ([25], s. 48).

[11] Jeżeli X oraz Y są zbiorami z dowolną strukturą algebraiczną, to każde odwzorowanie $f: X \to Y$ nazywamy *przekształceniem*.

[12] *Zbiorami borelowskimi* w przestrzeni topologicznej X nazywamy zbiory należące do σ-ciała generowanego przez topologię w X ([17], s. 21).

[13] Jeżeli $\langle\cdot,\cdot\rangle$ jest iloczynem skalarnym w X, $f \in X$ oraz $g := (g_i \in X)_I$, to dla uproszczenia zapisu

$$\langle f,g\rangle := (\langle f,g_i\rangle)_I \quad \text{oraz} \quad \langle g,f\rangle := (\langle g_i,f\rangle)_I$$

[14] Jeżeli μ jest miarą na X oraz $f: X \to Y$, to przez $\mu \circ f^{-1}$ będziemy oznaczać miarę wyznaczoną w Y przez odwzorowanie f i miarę μ ([1], s. 185 i [17], s. 21).

oraz $1 \leqslant p \leqslant \infty$, *to dla każdego przekształcenia* $\varphi \in C^n_{Fp\mu}(L^m_{Fpx}(R))$ *istnieje taka funkcja* $f \in C^n_{Fpv}(l_{F2})$, *dla której*

$$\varphi = f \circ \Phi.$$

Dowód. Ponieważ każdy układ ortogonalny jest liniowo niezależny, a każdy układ ortogonalny maksymalny, nazywany też bazą ortogonalną, jest pełny, to $\forall u \in L^m_{F2x}$ $\exists \alpha := (\alpha_i \in \mathrm{F})_N$, dla którego

$$\| u - \alpha^T \upsilon \|_x = 0,$$

gdzie $\| \cdot \|_x$ jest normą wyznaczoną przez iloczyn skalarny w L^m_{F2x} oraz

$$\alpha = \Phi(u)$$

([17], s. 92 i 95). Z twierdzenia Pitagorasa mamy, że

$$\| u \|^2_x = \| \alpha^T \upsilon \|^2_x = \left((|\alpha_i|^2)_N\right)^T \| \upsilon \|^2_x = \sum_{i \in N} |\alpha_i|^2 = \| \alpha \|^2 < \infty^{(15)},$$

gdzie $\| \cdot \|$ jest normą w l_{F2}, a zatem obrazem przestrzeni L^m_{F2x} w odwzorowaniu Φ jest przestrzeń l_{F2}. Widać więc, że miara ν może spełnić przyjęte założenie:

$$\nu = \mu \circ \Phi^{-1}.$$

Ponieważ L^m_{F2x} jest przestrzenią Hilberta ([14], s. 53), to z twierdzenia Riesza – Fischera ([14], s. 76) i z założenia, iż υ jest bazą ortogonalną wynika, że $\forall \alpha \in l_{F2}$ element $\alpha^T \upsilon \in L^m_{F2x}$. Możemy zatem $\forall \alpha \in l_{F2}$ zdefiniować funkcję f w następujący sposób:

$$f(\alpha) := \varphi(\alpha^T \upsilon).$$

Pokazaliśmy już, że $\forall u \in L^m_{F2x}$ $\exists \alpha \in l_{F2}$, dla którego u oraz α^T należą do tej samej klasy równoważności, a z definicji przekształcenia Φ wynika, że współczynniki α są jednakowe na każdej klasie równoważności. Z kolei φ jest przekształceniem ciągłym na przestrzeni L^m_{F2x}, czyli dla każdego $u \in L^m_{F2x}$ oraz dla każdego ciągu $(u_i \in L^m_{F2x})_N$, dla którego

$$\lim_{i \to \infty} \| u - u_i \|_x = 0,$$

granica

$$\lim_{i \to \infty} |\varphi(u) - \varphi(u_i)| = 0,$$

([14], s. 25). Przekształcenie φ oraz Φ są zatem stałe na każdej klasie równoważności, a stąd wynika, że

(15) Jeżeli $\| \cdot \|$ jest normą w X oraz $f := (f_i \in X)_I$, to dla uproszczenia zapisu

$$\| f \| := (\| f_i \|)_I.$$

$$\varphi = f \circ \Phi.$$

Pokażemy teraz, że funkcja f jest ciągła. W tym celu zauważmy, że Φ jest przekształceniem liniowym. Ponieważ wykazaliśmy na wstępie, że $\forall u \in L^m_{F2x}$

$$\| \Phi(u) \| = \| u \|_x$$

to przekształcenie Φ jest również ograniczone. Stąd i z twierdzenia Banacha o przekształceniu ciągłym ([14], s. 139) wynika, że Φ jest przekształceniem ciągłym. Można z kolei dowieść, żę przestrzenie L^m_{F2x} i l_{F2} są przestrzeniami Banacha ([14], s. 53 i 51). Z twierdzenia o odwzorowaniu otwartym ([14], s. 147) wynika zatem, że Φ jest odzworowaniem otwartym. Załóżmy teraz, że τ jest topologią w F^n, τ_v jest topologią w l_{F2} oraz τ_μ jest topologią w L^m_{F2x}. Z ciągłości przekształcenia φ mamy, że przeciwobrazy

$$[f \circ \Phi]^{-1}(\tau) \subset \tau_\mu \text{ (16)}.$$

Ponieważ Φ jest odwzorowaniem otwartym, to obrazy

$$\Phi(\tau_\mu) \subset \tau_v.$$

Z wykazanej ciągłosci przekształcenia Φ mamy, że przeciwobrazy

$$\Phi^{-1}(\tau_v) \subset \tau_\mu.$$

Zauważmy, że Φ jest bijekcją klas równoważności. Istnieje zatem odwzorowanie odwrotne Φ^{-1}, które $\forall \alpha \in l_{F2}$ określone jest przez wzór:

$$\Phi^{-1}(\alpha) = \alpha^T v,$$

a stąd

$$\Phi(\Phi^{-1}(\tau_v)) = \tau_v \subset \Phi(\tau_\mu).$$

Z wzajemnego zawierania się rodzin $\Phi(\tau_\mu)$ oraz τ_v mamy, że

$$\Phi(\tau_\mu) = \tau_v \quad \text{oraz} \quad \Phi^{-1}(\tau_v) = \tau_\mu.$$

Wynika stąd, że przeciwobrazy

$$f^{-1}(\tau) \subset \Phi(\tau_\mu) = \tau_v,$$

a zatem funkcja f jest ciągła.

Załóżmy, że $\mathcal{M}$ jest σ-ciałem w L^m_{F2x} oraz $\mathcal{N}$ jest σ-ciałem w l_{F2}. Pokażemy teraz, że funkcja f jest $\mathcal{N}$-mierzalna. Przypomnijmy, że przekształcenie Φ jest ciągłe oraz zgodnie z założeniem $\mathcal{M}$ jest σ-ciałem borelowskim. Można wowczas wykazać, że $\mathcal{N}$ jest również σ-ciałem borelowskim, a stąd i z dowiedzionej ciągłości funkcji f wynika jej $\mathcal{N}$-mierzalność ([1], s. 182).

Przejdźmy teraz do wykazania, że funkcje $f \in C^n_{Fpv}$. Jeżeli $p = \infty$, to dowód jest oczywisty. Załóżmy zatem, że $1 \leqslant p < \infty$. Ponieważ wykorzystamy dalej twierdzenie

(16) Jeżeli f jest odwzorowaniem na X oraz $\mathcal{F}$ jest rodzina podzbiorów zbioru X, to przez $f(\mathcal{F})$ oznaczamy rodzinę zbiorów $\{f(A) : A \in \mathcal{F}\}$, gdzie $f(A)$ jest obrazem zbioru A w odwzorowaniu f.

o zamianie zmiennych w całce Lebesgue'a ([1], s. 214), to należy pokazać, że przekształcenie Φ jest $\mathcal{M}/\mathcal{N}$-mierzalne. Jest tak rzeczywiście, co wynika z założenia, że ν jest miarą wyznaczoną na σ-ciele $\mathcal{N}$ w przestrzeni l_{F2} przez przekształcenie Φ i miarę μ na σ-ciele $\mathcal{M}$ w przestrzeni $L^m_{F2\varkappa}$ ([1], s. 185).

Ponieważ $\varphi \in C^n_{Fp\mu}(L^m_{F2\varkappa}) \subset L^n_{Fp\mu}(L^m_{F2\varkappa})$ oraz

$$\varphi = f \circ \Phi,$$

to spełnione są wszystkie założenia twierdzenia o zamianie zmiennych w całce Lebesgue'a ([1], s. 214), z którego wynika, że $kf \in L^n_{Fp\nu}$. Z dowiedzionej już ciągłości funkcji f mamy więc, że $f \in C^n_{Fp\nu} \subset L^n_{Fp\nu}$, co kończy dowód. ■

Udowodnione twierdzenie pozwala nam wniknąć w szczegółową strukturę operatorów stacjonarnych. By to pokazać załóżmy, że $\varkappa$ jest taką miarą na R, dla której

$$L^m_{F2\varkappa\nabla} = L^m_{F2\varkappa}.$$

Wynika wówczas ([8], że dla każdego stacjonarnego operatora $\varphi \circ \nabla.$ na $L^m_{F2\varkappa}$, dla którego $\varphi \in C^n_{F2\mu}(L^m_{F2\varkappa})$ jest przekształceniem, o którym mowa w ostatnio udowodnionym twierdzeniu, istnieje taka funkcja $f \in C^n_{Fp\nu}(l_{F2})$, że

$$\varphi(\nabla_t(u)) = f(\langle \nabla_t(u), \upsilon \rangle_\varkappa),$$

gdzie $\upsilon := (\upsilon_i)_N$ jest bazą ortonormalną w $L^m_{F2\varkappa}$.

4. SZCZEGÓŁOWA STRUKTURA SYSTEMÓW STACJONARNYCH PRZYCZYNOWYCH

Wśród operatorów, najistotniejszą grupę stanowią operatory stacjonarne przyczynowe. Zauważmy, że struktura operatora $\varphi \circ \nabla.$, która wynika z twierdzenia 1, nie jest w ogólnym przypadku strukturą operatora przyczynowego ([8]). Odpowiedź operatora $\varphi \circ \nabla.$ w chwili t na wymuszeniu u może bowiem zależeć również od wartości wymuszenia u po chwili t. Zajmiemy się zatem dalej strukturą operatorów stacjonarnych przyczynowych.

Twierdzenie 2. *Jeżeli $\varkappa$ jest miarą symetryczną na R*[17] *taką, że*

$$L^m_{F2\varkappa\nabla}(R) = L^m_{F2\varkappa}(R),$$

$\upsilon := (\upsilon_t)_N$ jest bazą ortonormalną w $L^m_{F2\varkappa+}(R)$, $\langle .,. \rangle_\varkappa$ jest iloczynem skalarnym w $L^m_{F2\varkappa}(R)$, μ jest miarą na σ-ciele zbiorów borelowskich w $L^m_{F2\varkappa}(R)$, ν jest miarą skończoną na l_{F2}.

---

[17] Niech μ będzie miarą na R^n. *Miarę* μ nazywamy *symetryczną*, jeżeli dla każdego mierzalnego zbioru $A \subset R^n$ zachodzi równość:

$$\mu(A) = \mu(-A),$$

gdzie

$$-A := \{-\alpha : \alpha \in A\}.$$

$$\omega := (\upsilon^T, \nabla^*(\upsilon^T))^T, \quad \Phi := \langle \cdot, \omega \rangle_\varkappa, \quad \nu \times \nu = \mu \circ \Phi^{-1}{}^{(18)},$$

$$\Gamma := \langle \cdot, \upsilon \rangle_\varkappa$$

oraz $1 \leqslant p \leqslant \infty$, *to dla każdego przekształcenia,* $\varphi \in C^n_{F p\mu}(L^m_{F2\varkappa}(R))$, *dla którego* $\varphi \circ \nabla$. jest *operatorem stacjonarnym przyczynowym, istnieje taka funkcja* $g \in C^n_{F p\nu}(l_{F2})$, *dla której*

$$\varphi = g \circ \Gamma \circ \nabla^*.$$

Dowód. Zauważmy na wstępie, że z założenia

$$L^m_{F2\varkappa\nabla} = L^m_{F2\varkappa}$$

wynika, że

$$L^m_{F2\varkappa\nabla} = L^m_{F2\varkappa *} = L^m_{F2\varkappa}$$

([9]). Z kolei $\forall u \in L^m_{F2\varkappa}$ i $\forall \tau \in R$ funkcję $\nabla_\tau(u)$ możemy rozłożyć na dwie składowe

$$u_{p\tau} := \nabla^*(u_{-1})\nabla(u) \quad \text{oraz} \quad u_{f\tau} := \nabla(u) - u_{p\tau}.$$

Składowe te spełniają równość

$$\nabla(u) = u_{p\tau} + u_{f\tau}.$$

Funkcja $\nabla^*(u_{p\tau}) \in L^m_{F2\varkappa+}$ oraz $\forall t > 0$

$$u_{p\tau}(t) = 0,$$

a zatem jest to składowa przeszła sygnału $\nabla(u)$ względem chwili 0. Podobnie funkcja $u_{f\tau} \in L^m_{F2\varkappa+}$ oraz $\forall t \leqslant 0$

$$u_{f\tau}(t) = 0,$$

i jest to składowa przyszła sygnału $\nabla(u)$ względem chwili 0. Zgodnie z założeniem układ υ jest bazą ortogonalną w $L^m_{F2\varkappa+}$. Wynika stąd, że $\nabla^*(\upsilon)$ jest bazą ortogonalną w przestrzeni

$$\nabla^* \circ L^m_{F2\varkappa+} := \left\{ \nabla^*(u) : u \in L^m_{F2\varkappa+} \right\},$$

a ponieważ miara $\varkappa$ jest symetryczna, to można również zauważyć, że ω jest bazą ortogonalną w $L^m_{F2\varkappa}$. Mamy zatem, że $\forall u \in L^m_{F2\varkappa}$ i $\forall \tau \in R$, $\exists \alpha_\tau \in l_{F2} \times l_{F2}$ i $\exists \alpha_{f\tau}, \alpha_{p\tau} \in l_{F2}$, dla których spełnione są równania:

$$\| \nabla_\tau(u) - \alpha_\tau^T \omega \|_\varkappa = 0,$$

$$\| u_{f\tau} - \alpha_{f\tau}^T \upsilon \|_\varkappa = 0 \quad \text{oraz} \quad \| u_{p\tau} - \alpha_{p\tau}^T \nabla^*(\upsilon) \|_\varkappa = 0,$$

gdzie $\| \cdot \|_\varkappa$ jest normą wyznaczoną przez iloczyn skalarny w $L^m_{F2\varkappa}$,

(18) Jeżeli μ jest miarą na X, ν jest miarą na Y oraz miary te są σ-skończone, to przez $\mu \times \nu$ oznaczamy miarę produktową na $X \times Y$ ([17], s. 152).

$$\alpha_\tau = \Phi(u), \quad \alpha_{f\tau} = \Gamma(u) \quad \text{oraz} \quad \alpha_{p\tau} = \Gamma(\nabla^*(u))$$

([17], s. 92 i 95). Ponieważ

$$\{v_i : i \in N\} \perp \{\nabla^*(v_i) : i \in N\},$$

to

$$\alpha_\tau = (\alpha_{f\tau}^T, \alpha_{p\tau}^T)^T.$$

Z twierdzenia 1 wynika, że $\exists f \in C^n_{Fpv\times v}(l_{F2} \times l_{F2})$, dla której

$$\varphi = f \circ \Phi,$$

gdyż obrazem przestrzeni L^m_{F2x} w odwzorowaniu Φ jest przestrzeń $l_{F2} \times l_{F2}$. Zgodnie z założeniem, $\varphi \circ \nabla$. jest operatorem stacjonarnym przyczynowym, czyli $\forall u \in L^m_{F2x}$ oraz $\forall t \in R$

$$\varphi(\nabla_\tau(u)) = \varphi(\nabla^*(u_{-1})\nabla_\tau(u))$$

([8]). Wynika stąd, że wartość funkcji f zależy jedynie od składowych wektora $\alpha_{p\tau}$, a zatem $\forall \alpha_{f\tau}, \alpha_{p\tau} \in l_{F2}$ oraz $0 \in l_{F2}$

$$f((\alpha_{f\tau}^T, \alpha_{p\tau}^T)^T) = f((\mathbf{0}^T, \alpha_{p\tau}^T)^T).$$

Istnieje więc funkcja $g : l_{F2} \to F^n$ taka, że $\forall \alpha_{p\tau} \in l_{F2}$ oraz $0 \in l_{F2}$

$$g(\alpha_{p\tau}) := f((\mathbf{0}^T, \alpha_{p\tau}^T)^T),$$

dla której

$$g \circ \Gamma \circ \nabla^* = f \circ \Phi.$$

Spełniony jest więc warunek

$$\varphi = g \circ \Gamma \circ \nabla^*.$$

Pokażemy teraz, że funkcja $g \in C^n_{Fpv}$. W tym celu zauważmy, że funkcja f zgodnie z twierdzeniem 1 jest ciągła oraz $v \times v$-mierzalna. Wynika stąd, że funkcja g będąca przekrojem funkcji f, też jest ciągła ([13], s. 152) i v-mierzalna ([17], s. 151). Dalszy dowód dla $p = \infty$ jest oczywisty. Niech więc $1 \leqslant p < \infty$. Funkcja $f \in C^n_{Fpv\times v}(l_{F2} \times l_{F2})$, a zatem

$$\|f\|_{v\times v} := \int_{l_{F2}\times l_{F2}} |f|^2 \mathrm{d}[v \times v] = \int_{l_{F2}} \mathrm{d}v \int_{l_{F2}} |f|^2 \mathrm{d}v < \infty,$$

gdzie $\|\cdot\|_{v\times v}$ jest normą w $L^n_{Fpv\times v}(l_{F2} \times l_{F2})$. Ponieważ wartość funkcji f zależy jedynie od składowych wektora $\alpha_{p\tau}$, to

$$\|f\|_{v\times v} = v(l_{F2}) \int_{l_{F2}} |f|^2 \mathrm{d}v = v(l_{F2}) \int_{l_{F2}} |g|^2 \mathrm{d}v = v(l_{F2}) \|g\|_v,$$

gdzie $\|\cdot\|_v$ jest normą w L^n_{Fpv}. Zgodnie z założeniem miara v jest skończona oraz $\|f\|_{v\times v} < \infty$, a więc $\|g\|_v < \infty$, czyli $g \in C^n_{Fpv} \subset L^n_{Fpv}$, co kończy dowód. ■

Zauważmy, że z przyjętego w ostatnim twierdzeniu założenia

$$L^m_{F2\varkappa\nabla} = L^m_{F2\varkappa}$$

wynika, iż

$$L^m_{F2\varkappa\nabla} = L^m_{F2\varkappa *} = L^m_{F2\varkappa}$$

([9]). Ponadto każdy stacjonarny i przyczynowy operator $\varphi \circ \nabla_\cdot$ na $L^m_{F2\varkappa}$, dla którego $\varphi \in C^n_{Fp\mu}(L^m_{F2\varkappa})$ jest przekształceniem spełniającym założenia udowodnionego twierdzenia, ma postać:

$$\varphi(\nabla_t(u)) = g(\langle \nabla^*(\nabla_t(u)), \upsilon \rangle_\varkappa) = g(\langle \nabla^*_t(u), \upsilon \rangle_\varkappa),$$

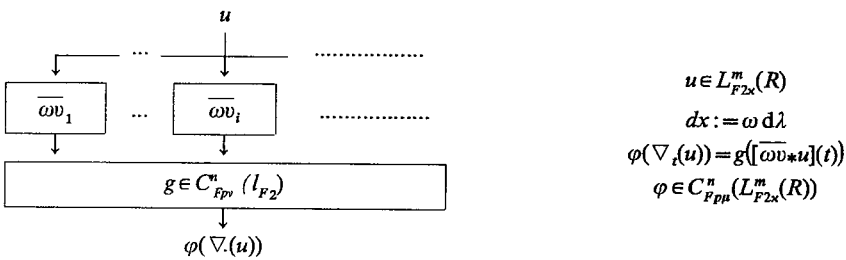
gdzie $g \in C^n_{Fpv}(l_{F_2})$, a $\upsilon := (\upsilon_i)_N$ jest bazą ortonormalną w $L^m_{F2\varkappa +}$. Ponieważ można wykazać ([9]), że

$$\langle \nabla^*_t(u), \upsilon \rangle_\varkappa = [\overline{w\upsilon} * u](t)^{(19)},$$

gdzie w jest gęstością miary $\varkappa$ względem miary Lebesgue'a na R, to

$$\varphi(\nabla_t(u)) = g([\overline{w\upsilon} * u](t)).$$

Tak otrzymaną strukturę operatora $\varphi \circ \nabla_\cdot$ pokazano na rys. 1.



Rys. 1. Struktura operatorów stacjonarnych przyczynowych

Warto jeszcze zauważyć, że funkcja g jest ciągła, a z twierdzeń dotyczących właściwości splotu funkcji ([9]) wynika, że sygnały $\overline{w\upsilon} * u$ są również ciągłe. Z otrzymanej struktury operatora $\varphi \circ \nabla_\cdot$ widać więc, że każda jego odpowiedź jest ciągła. Otrzymana właściwość jest zgodna z rzeczywistością, gdyż wszystkie fizycznie realizowalne systemy mają ograniczone pasmo przenoszenia, a stąd wynika ciągłość ich odpowiedzi.

---

[19] Jeżeli $u := (u_i)_I$, to dla uproszczenia zapisu

$$u * \upsilon := (u_i * \upsilon)_I.$$

PODSUMOWANIE

W artykule omówione zostały zagadnienia związane z identyfikacją strukturalną wielowymiarowych systemów stacjonarnych opisanych za pomocą operatorów, które w ogólnym przypadku mogą być nieliniowe. Założono, że wartości sygnałów wejściowych i wyjściowych rozważanych systemów mogą przyjmować wartości rzeczywiste lub zespolone, a zbiór sygnałów wejściowych ma strukturę przestrzeni Hilberta. Udowodniono twierdzenie 1, które pozwala wniknąć w szczegółową strukturę operatorów stacjonarnych oraz twierdzenie 2, dotyczące szczegółowej struktury operatorów stacjonarnych przyczynowych, których strukturę pokazano na rys. 1. Otrzymane wyniki pozwalają na lepsze zrozumienie istoty tak złożonych odwzorowań, jakimi są operatory, a co za tym idzie, istoty systemów opisanych za ich pomocą. Ponadto wyniki te są niezbędnym elementem ogólnej teorii modelowania systemów, obejmujacej między innymi, teorię Volterry oraz Wienera ([2]–[4] i [6]).

Zauważmy na koniec, że problem wyznaczenia modelu systemu opisanego za pomocą operatora $\varphi \circ \nabla$. można sprowadzić do problemu aproksymacji średniokwadratowej przekształcenia φ ([2]–[4] i [6]). Z udowodnionego twierdzenia 2 wynika, że przekształcenie to należy do przestrzeni $L^n_{Fp\mu}(L^m_{F2\infty})$, w której iloczyn skalarny dla $p=2$ zdefiniowany jest $\forall \varphi, \gamma \in L^n_{Fp\mu}$ przez wzór:

$$\langle \varphi, \gamma \rangle_\mu := \int_{L^m_{F2\infty}} \gamma^*(u)\varphi(u)\,\mathrm{d}\mu(u).$$

Z twierdzenia o aproksymacji średniokwadratowej ([10], s. 91) wynika, że dla wyznaczenia modelu systemu $f \circ \nabla$., niezbędnym jest obliczenie tego typu całek, co wydaje się dość trudne. Okazuje się jednak, że sygnały stochastyczne ergodyczne pozwalają wyznaczać całki na przestrzeniach funkcyjnych ([2]–[6]).

BIBLIOGRAFIA

1. P. Billingsley (tłum. z ang. K. Kizeweter, J.E. Rogulski): *Prawdopodobieństwo i miara*, PWN, Warszawa 1987
2. G. Ciesielski, S. Derlecki, Z. Marks-Wojciechowska: *Identyfikacja wielowymiarowych systemów pomiarowych opisanych operatorem nieliniowym.* Praca badawcza nr I12/5/90 BP, Politechnika Łódzka, Łódź 1990
3. G. Ciesielski, S. Derlecki, Z. Marks-Wojciechowska: *Identyfikacja i korekcja nieliniowa wielowymiarowych systemów pomiarowych opisanych operatorem nieliniowym.* Praca badawcza nr MEN I12/6/90 BP, Politechnika Łódzka, Łódź 1990
4. G. Ciesielski: *Aproksymacja średniokwadratowa operatorów nieliniowych opisujących wielowymiarowe systemy pomiarowe.* XXIII Międzyuczelniana Konferencja Metrologów, Politechnika Warszawska, Warszawa 1991 (referat nie opublikowany)
5. G. Ciesielski: *Modele Wienera wielowymiarowych systemów pomiarowych opisanych za pomocą operatorów nieliniowych.* Modelowanie i symulacja systemów pomiarowych. Materiały Sympozjum, AGH, Kraków 1992, s. 41–50

6. G. Ciesielski: *Operatorowe modele wielowymiarowych systemów pomiarowych* Seminarium Naukowe Komisji Kształcenia Komitetu Metrologii i Aparatury Naukowej Polskiej Akademii Nauk, 1994 (referat nie opublikowany)
7. G. Ciesielski: *Identyfikacja ogólnej struktury wielowymiarowych systemów pomiarowych.* Modelowanie i symulacja systemów pomiarowych, Materiały Sympozjum, AGH, Kraków, 1994, s. 17–23
8. G. Ciesielski: *Identyfikacja ogólnej struktury wielowymiarowych systemów stacjonarnych.* Kwart. Elektr i Telekom. 1994, t. 40, z. 3, s. 419–428
9. G. Ciesielski: *Modelowanie wielowymiarowych liniowych systemów stacjonarnych za pomocą splotów.* Kwart. Elektr. i Telekom., 1994, t. 40, z. 3, s. 429–448
10. G. Dahlquist, Å. Björck (tłum. z ang. S. Paszkowski): *Metody numeryczne,* PWN, Warszawa 1983
11. R.J.P. de Figueiredo, T.A.W. Dwyer: *A best approximation framework and implementation for simulation of large-scale nonlinear systems.* IEEE Trans. Circuits Syst., 1980, vol. CAS-27, nr 11, s. 1005–1014
12. R.J.P. de Figueiredo: *Nonlinear circuits and systems applications of functional splines.* Proc. of IEEE, 1982, s. 1245–1247
13. K. Kuratowski: *Wstęp do teorii mnogości i topologii.* Biblioteka Matematyczna, t. 9, PWN, Warszawa 1980
14. J. Musielak: *Wstęp do analizy funkcjonalnej.* PWN, Warszawa 1989
15. J. Park, I.W. Sandberg: *Criteria for the approximation of nonlinear systems.* IEEE Trnas. Circuts Syst.-I: Fundamental Theory and Applications, 1992, vol. 39, nr 8, s. 673–676
16. K.A. Pupkow, W.I. Kapalin, A.S. Juszczenko: *Funkcjonalnyje rjady w tjeorii nieliniejnych sistjem.* Nauka, Moskwa 1976
17. W. Rudin (tłum. z ang. A. Pierzchalski, P. Walczak): *Analiza rzeczywista i zespolona.* PWN, Warszawa 1986
18. I.W. Sandberg: *Criteria for the response of nonlinear systems to be L-asymptotic periodic.* Bell Syst. Tech. J., 1981, vol. 60, nr 10, s. 2359–2371
19. I.W. Sandberg: *Series expansion for nonlinear systems.* Proc. IEEE, 1982, s. 110–113
20. I.W. Sandberg: *Expansion for nonlinear systems.* Bell Syst. Tech. J., 1982, vol. 61, nr 2, s. 159–199
21. I.W. Sandberg: *Volterra expansion for time-varying nonlinear systems.* Bell Syst. Tech. J., 1982, vol. 61, nr 2, s. 201–225
22. I.W. Sandberg: *Multidimensional nonlinear systems and structure theorems.* J. Circuits, Syst. and Computers, 1992, vol. 2, nr 4, s. 383–388
23. I.W. Sandberg: *Approximately-finite memory and input-output maps.* IEEE Trans. Circuits Syst.–I: Fundamental Theory and Applications, 1992, vol. 39, nr 7, s. 549–556
24. M. Schetzen: *The Volterra and Wiener Theories of Nonlinear Systems.* John Wiley & Sons, New York 1980
25. A. Wojnar: *Teoria sygnałów.* Podręczniki akademickie EIT, WNT, Warszawa 1980

G. CIESIELSKI

DETAILED STRUCTURE IDENTIFICATION OF MULTIDIMENSIONAL TIME-INVARIANT SYSTEMS

Summary

This paper concern some problems related to the structure identification of multidimensional time-invariant systems governed by operators, which can be nonlinear in general case. It is assumed that the inputs and responses of considered systems are real or complex valued, and the set of inputs have the Hilbert

space structure. There are proven two theorems, which allow to get an insight into the detailed structure of time-invariant operators and causal time-invariant operators. Received results afford possibilities for better understanding of the essence of such complicated maps as operators and, which involves the essence of systems governed by them.

Key words: structure identification, detailed structure of systems, system modelling, multidimensional systems, time-invariant systems, operators, convolution, stationary space, convolution space, causal space.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 3, ss. 321–335

Modelowanie wielowymiarowych systemów stacjonarnych przyczynowych[c)]

GRZEGORZ CIESIELSKI

Instytut Elektroniki Teoretycznej, Metrologii i Materiałoznawstwa, Politechnika Łódzka

Otrzymano 1995.01.25

Autoryzowano do druku 1995.06.08

Artykuł dotyczy zagadnień związanych z modelowaniem wielowymiarowych systemów stacjonarnych przyczynowych opisanych za pomocą operatorów, które w ogólnym przypadku mogą być nieliniowe. Założono, że wartości sygnałów wejściowych i wyjściowych tych systemów mogą przyjmować wartości rzeczywiste lub zespolone, a zbiór sygnałów wejściowych ma strukturę przestrzeni Hilberta. Udowodniono twierdzenie, które wykorzystuje teorię aproksymacji średniokwadratowej pozwala rozwiązać zadanie modelowania rozważanych systemów. Otrzymane wyniki pozwalają lepiej zrozumieć istotę tak złożonych odwzorowań, jakimi są operatory opisujące wielowymiarowe systemy stacjonarne przyczynowe.

Słowa kluczowe: modelowanie, identyfikacja strukturalna, identyfikacja parametryczna, systemy wielowymiarowe, systemy stacjonarne, systemy przyczynowe, operatory, sploty, przestrzeń stacjonarna, przestrzeń splotowa, przestrzeń przyczynowa.

1. WSTĘP

Artykuł dotyczy modelowania ([17]) wielowymiarowych systemów stacjonarnych przyczynowych, które dają się opisać za pomocą odwzorowań nazywanych operatorami. Systemy rozważane w artykule, jak też związane z nimi operatory, mogą być nieliniowe, a jedną z możliwych form ich reprezentacji jest układ równań różniczkowych, być może nieliniowych. Systemy te są wielowymiarowe, czyli mają dowolną skończoną liczbę wejść i wyjść. Sygnały na poszczególnych wejściach i wyjściach mogą przyjmować wartości rzeczywiste lub zespolone. Ponadto przyjęto, że zbiór sygnałów wejściowych ma strukturę przestrzeni Hilberta i może zawierać wszystkie funkcje typu

[c)] Pracę wykonano w ramach projektu badawczego Nr 8 T10C 031 09 finansowanego przez Komitet Badań Naukowych w latach 1995–1997.

wykładniczego ([10]) które jak wiadomo, są bezwzględnie transformowalne w sensie jednostronnego przekształcenia Laplace'a ([21], s. 88).

Ponieważ w światowej literaturze naukowej ([14]–[15], [23], [25] i [27]–[33]) nie podjęto udanej próby w pełni formalnego, a zarazem jednolitego podejścia do zagadnienia modelowania i korekcji systemów opisanych za pomocą operatorów nieliniowych, to głównym celem artykułu jest sformułowanie ogólnej teorii modelowania systemów, które obejmowałaby między innymi teorię Volterry oraz Wienera ([3]–[7]). W artykule będą wykorzystywane preliminaria matematyczne sformułowane w [11].

2. STABILNOŚĆ SYSTEMÓW

Jedną z ważnych właściwości operatorów oraz opisanych za ich pomocą systemów jest stabilność. By ją zdefiniować załóżmy, że $U_F \subset L_F^m(R)$, $V_F \subset L_F^p(R)$, $\Sigma_U : U \to \bar{R}_+$ oraz $\Sigma_V : V \to \bar{R}_+$.

Operator $\phi : U \to V$ będziemy nazywać *stabilnym względem funkcjonałów* Σ_U oraz Σ_y, jeżeli

$$\forall u \in U((\Sigma_U(u) < \infty) \Rightarrow (\Sigma_V(\phi(u)) < \infty))).$$

Załóżmy dalej, że $\| \cdot \|_U$ jest normą w U, $\| \cdot \|_V$ jest normą w V,

$$\Sigma_U = \| \cdot \|_U \quad \text{oraz} \quad \Sigma_V = \| \cdot \|_V.$$

Z podanej definicji stabilności mamy wówczas, że operator ϕ jest zawsze stabilny na U względem funkcjonałów Σ_U oraz Σ_V, gdyż każda norma przyjmuje wartości ze zbioru R_+. Jeżeli normy $\| \cdot \|_U$ oraz $\| \cdot \|_V$ są jednostajne[1], to *operator* ϕ na U nazywamy *jednostajnie stabilnym* (*stabilnym względem normy jednostajnej*), a tak określony warunek stabilności nazywamy *warunkiem BIBO* (ang. *bounded input–bounded output*). Jeżeli $1 \leqslant p \leqslant \infty$, $U \subset L_{fp\mu}^m(R)$ oraz $V \subset L_{Fp\mu}^n(R)$, to *operator* ϕ na U nazywać będziemy *μ-stabilnym* (*stabilnym względem normy* $\| \cdot \|_p$). *Operator* stabilny względem normy $\| \cdot \|_\infty$ nazywamy też *istotnie* (*μ-istotnie*) *stabilnym* na U.

Z przedstawionych definicji wynika, że operator ϕ jednostajnie stabilny jest również istotnie stabilny. Ponadto dla skończonej miary μ, operator istotnie stabilny jest również μ-stabilny ([20], s. 54). Zobaczymy, że warunek jednostajnej stabilności jest często zbyt silny, dlatego też zwykle stosować będziemy słabsze warunki stabilności.

---

[1] Jeżeli $\| \cdot \|_Y$ jest normą w Y, to

$$\forall f \in \{g : X \to Y : \sup\{ \| g(x) \|_Y : x \in X\} < \infty\}$$

norma jednostajna (*Czebyszewa*) oznaczana przez

$$\| f \|_u := \sup\{ \| f(x) \|_Y : x \in X\}.$$

3. SYGNAŁY ERGODYCZNE

Sygnały ergodyczne pozwalają wyznaczać całki złożonych odwzorowań. Wykorzystywane jest przy tym pojęcie wartości średniej sygnału.

Załóżmy, że funkcja $x: R \to F^n$ jest sygnałem oraz $T \in R$. *Wartość średnią sygnału* x oznaczaną przez

$$A(x) := \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)\,\mathrm{d}t.$$

Załóżmy dalej, że μ jest miarą na F^m.

Sygnał $x: R \to F^m$ nazywamy *ergodycznym względem funkcji* $g \in L^n_{F1\mu}(F^m)$, gdy

$$\int_{F^m} g\,\mathrm{d}\mu = A(g(x)).$$

Jeżeli

$$g := (g_i \in L^n_{F1\mu}(F^m))_I,$$

to *sygnał* x nazywamy *ergodycznym względem układu funkcji* g, gdy jest ergodyczny względem każdej funkcji wchodzącej w skład układu g. Z kolei *sygnał* x nazywamy *ergodycznym*, gdy $\forall g \in L^n_{F1\mu}(F^m)$

$$\int_{F^m} g\,\mathrm{d}\mu = A(g(x)).$$

Ergodyczność sygnałów rozważana jest też w [1], [22] oraz [24].

4. ROZWIĄZANIE ZADANIA MODELOWANIA SYSTEMÓW

Zwykle teoria aproksymacji stosowana jest do wyznaczania przybliżeń prostych funkcji. Okazuje się, że teoria ta obejmuje również bardziej złożone odwzorowania, jakimi są niewątpliwie operatory nieliniowe. Ponadto, zadanie modelowania systemu opisanego za pomocą operatora jest równoważne zadaniu aproksymacji liniowej tego operatora. Podamy teraz twierdzenie rozwiązujące zadanie modelowania systemów stacjonarnych przyczynowych.

Twierdzenie. *Załóżmy, że* $\varkappa$ jest taką gęstością miary $\varkappa$ *względem miary Lebesgue'a na* R[2], *że*

[2] Będziemy zakładać, że gęstość $\varkappa$ miary μ na X względem miary ν również na X jest funkcją ν-mierzalną o wartościach z $\bar{R}_+$. Piszemy wówczas

$$\mathrm{d}\mu := \varkappa \mathrm{d}\nu$$

([26], s. 32).

$$L^m_{F2x^*}(R) = L^m_{F2x}(R),$$

$\langle \cdot, \cdot \rangle_x$ jest iloczynem skalarnym w $L^m_{F2x}(R)$, μ jest miarą na σ-ciele zbiorów borelowskich[3] *w $L^m_{F2x}(R)$ niezmienniczą względem obrotu funkcyjnego*[4],

$$\upsilon_{[k]} := (\upsilon_i)_1^{k(5)}$$

jest układem funkcji liniowo niezależnych w $L^m_{F2x+}(R)$,

$$\Phi_k := \bar{G}^{-1}(\upsilon_{[k]})\langle \cdot, \upsilon_{[k]} \rangle_x^{(6)}, \quad \nu_k = \mu \circ \phi_k^{-1}$$

jest miarą wyznaczoną w F^k przez przekształcenie[7] *ϕ_k i miarę μ,*

$$\chi_{[l]} := (\chi_i)_1^l$$

jest układem liniowo niezależnym w przestrzeni $L^n_{F2\nu_k}(F^k)$ oraz

$$\vartheta_{k[l]} := (\vartheta_{k,i})_1^l := \chi_{[l]} \circ \Phi_k \circ \nabla^*.$$

Prawdziwe są wówczas następujące twierdzenia:

(1) *$\vartheta_{k[l]}$ jest układem przekształceń liniowo niezależnych w przestrzeni $L^n_{F2\mu}(L^m_{F2x}(R))$.*

(2) *Jeżeli $\langle \cdot, \cdot \rangle_\mu$ jest iloczynem skalarnym w przestrzeni $L^n_{F2\mu}(L^m_{F2x}(R))$, to dla każdego przekształcenia $\varphi \in L^n_{F2\mu}(L^m_{F2x}(R))$ istnieje dokładnie jedno przybliżenie średniokwadratowe $\gamma_{k,l} \in \operatorname{span}\vartheta_{k[l]}$*[8] *i ma ono postać:*

$$\gamma_{k,l} = (G^{-1}(\chi_{[l]}) < \vartheta_{k[l]}, \varphi >_\mu)^* \vartheta_{k[l]}. \tag{2.1}$$

---

[3] *Zbiorami borelowskimi* w przestrzeni topologicznej X nazywamy zbiory należące do σ-ciała generowanego przez topologię w X ([26], s. 21).

[4] Jeżeli μ jest miarą na X oraz $f: X \to Y$, to przez $\mu \circ f^{-1}$ oznaczamy miarę wyznaczoną w Y przez odwzorowanie f i miarę μ ([2], s. 185).

Niech teraz μ będzie miarą na X oraz $f: X \to X$. *Miarę* μ nazywamy *niezmienniczą względem odwzorowania* f, jeżeli

$$\mu = \mu \circ f^{-1}.$$

[5] Wektor n-elementowy α oznaczamy też symbolem $\alpha_{[n]}$.

[6] Niech

$$I := \{1, \ldots, k\}, \quad \text{lub} \quad I := N,$$

$$\upsilon := (\upsilon_i)_I$$

będzie dowolnym układem elementów z przestrzeni unitarnej X z iloczynem skalarnym $\langle \cdot, \cdot \rangle$. Przez $\mathbf{G}(\upsilon)$ oznaczamy dalej *macierz Grama układu* υ, która ma postać:

$$\mathbf{G}(\upsilon) := (\langle \upsilon_i, \upsilon_j \rangle)_{i,j \in I}.$$

Jeżeli układ υ wynika z kontekstu, to macierz tą oznaczamy w skrócie przez $\mathbf{G}$. Ponadto dla dowolnego $x \in X$

$$\langle x, \upsilon \rangle := (\langle x, \upsilon_i \rangle)_I \quad \text{oraz} \quad \langle \upsilon, x \rangle := (\langle \upsilon_i, x \rangle)_I.$$

[7] Jeżeli X oraz Y są zbiorami z pewną strukturą algebraiczną, to dowolne odwzorowanie $f: X \to Y$ nazywamy *przekształceniem*.

[8] Niech x będzie dowolnym układem elementów z przestrzeni liniowej X. Przez $\operatorname{span} x$ oznaczamy przestrzeń liniową generowaną przez układ x, która jest zbiorem wszystkich liniowych kombinacji elementów tego układu.

Załóżmy dalej, że

$$L^m_{F2\varkappa\nabla}(R)=L^m_{F2\varkappa}(R).$$

(3) *Elementy układu* $\vartheta_{k[l]}\circ\nabla$ *są liniowo niezależnymi operatorami stacjonarnymi i przyczynowymi.*

(4) *Jeżeli* $\varkappa$ *jest miarą skonczoną oraz funkcje* $(\chi_i)_1^l$ *są lokalnie ograniczone, to elementy układu* $\vartheta_{k[l]}\circ\nabla$ *są operatorami lokalnie ograniczonymi,* $\varkappa$*-stabilnymi względem normy* $\|\cdot\|_2$ *oraz* $\forall u\in L^m_{F2\varkappa}(R)$ *i* $\forall i=1,\ldots,l$ *funkcja* $\vartheta_{k,i}(\nabla.(u))\in L^n_{F2\varkappa}(R)$.

Załóżmy ponadto dalej, że $\varkappa$ *jest miarą symetryczną*[9],

$$\upsilon:=(\upsilon_i)_N$$

jest bazą ortonormalną w $L^m_{F2\varkappa+}(R)$, ν *jest miarą skończoną na* l_{F2}, *układ*

$$\omega:=(\upsilon^T,\nabla^*(\upsilon^T))^T,\quad \Gamma:=\langle\cdot,\omega\rangle_\varkappa \quad\text{oraz}\quad \nu\times\nu:=\mu\circ\Gamma^{-1}{}^{(10)}.$$

(5) Jeżeli

$$\chi:=(\chi_i)_N$$

jest układem liniowo niezależnym pełnym[11] *w przestrzeni* $L^n_{F2\nu_k}(F^k)$ *oraz* $\|\cdot\|_\mu$ *jest normą wyznaczoną przez iloczyn skalarny w przestrzeni* $L^n_{F2\mu}(L^m_{F2\varkappa}(R))$, *to dla każdego przekształcenia* $\varphi\in C^n_{F\infty\mu}(L^m_{F2\varkappa}(R))$, *dla którego złożenie* $\varphi\circ\nabla.$ *jest operatorem stacjonarnym przyczynowym, przybliżenie średniokwadratowe* $\gamma_{k,l}$ *określone przez wzór* (2.1), *spełnia warunek:*

$$\lim_{k,l\to\infty}\|\varphi-\gamma_{k,l}\|_\mu=0.$$

---

[9] Niech μ będzie miarą na R^n. *Miarę* μ nazywamy *symetryczną*, jeżeli dla każdego mierzalnego zbioru $A\subset R^n$ zachodzi równość:

$$\mu(A)=\mu(-A),$$

gdzie

$$-A:=\{-\alpha:\alpha\in A)\}.$$

[10] Jeżeli μ jest miarą na X, ν jest miarą na Y oraz miary te są σ-skończone, to przez $\mu\times\nu$ oznaczamy miarę produktową na $X\times Y$ ([26], s. 151).

[11] Niech υ będzie układem elementów z przestrzeni X_F z normą $\|\cdot\|$. *Układ* υ nazywamy *pełnym* w X, jeżeli zbiór span υ jest gęsty w X. Niech ponadto

$$\upsilon:=(\upsilon_i)_N,\quad \upsilon_n:=(\upsilon_i)_1^n,\quad \alpha:=(\alpha_i\in F)_N \quad\text{oraz}\quad \alpha_n:=(\alpha_i\in F)_1^n.$$

Jeżeli dla każdego $x\in X$ istnieje jeden i tylko jeden wektor $\mathbf{a}$, dla którego

$$\lim_{n\to\infty}\|x-a_n^T\upsilon_n\|=0,$$

to układ υ nazywamy *bazą Schaudera* ([20], s. 125). Piszemy wtedy

$$x=a^T\upsilon:=\sum_{i\in N}a_i\upsilon_i:=\sum_{i=1}^{\infty}a_i\upsilon_i.$$

(6) *Jeżeli* $f \in L^n_{F2v}(l_{F2})$,

$$\Phi := \langle \cdot, v \rangle_\varkappa, \quad \varphi := f \circ \Phi \circ \nabla^*,$$

$\forall x := (x_i)_N \in 1_{F2}$ *odwzorowanie*

$$\pi_k(x) := (x_i)_1^k$$

oraz ξ *jest takim wymuszeniem, że sygnały* $\overline{\varkappa v} * \xi$ *są ergodyczne względem układu funkcji* $f^* \chi_{[l]} \circ \pi_k$ *z przestrzeni* $L^n_{F1v}(l_{F2})$, *to przybliżenie średniokwadratowe* $\gamma_{k,l}$, *określone przez wzór* (2.1), *spełnia równość:*

$$\gamma_{k,l} = \left(G^{-1}(\chi_{[l]}) A\left(\varphi^*(\nabla_\cdot(\xi)) \vartheta_{[l]}(\nabla_\cdot(\xi))\right)\right)^* \vartheta_{k[l]}{}^{(12)}.$$

Dowód

(1) Przypomnijmy, że dowolny układ elementów z przestrzeni unitarnej jest liniowo niezależny wtedy i tylko wtedy, gdy macierz Grama tego ukladu jest odwracalna ([16], s. 216). By wykazać zatem, że $\vartheta_{[l]}$ jest układem liniowo niezależnym wystarczy pokazać, że macierz Grama

$$\mathbf{G}(\vartheta_{[l]}) = (\langle \vartheta_{k,i}, \vartheta_{k,j} \rangle_\mu)_{i,j=1}^l$$

jest macierzą nieosobliwą. W tym celu zauważmy, że z niezmienniczości miary μ względem obrotu funkcyjnego oraz twierdzenia o zamianie zmiennych w całce Lebesgue'a ([2], s. 214) wynika, iż $\forall i,j = 1,\ldots,l$ iloczyn skalarny

$$\langle \vartheta_{k,i}, \vartheta_{k,j} \rangle_\mu = \int\limits_{L^m_{F2\varkappa}} \vartheta^*_{k,j} \vartheta_{k,i} \, d\mu = \int\limits_{F^k} \chi^*_j \chi_i \, d\nu_k = \langle \chi_i, \chi_j \rangle_{\nu_k},$$

gdzie $\langle \cdot, \cdot \rangle_{\nu_k}$ jest iloczynem skalarnym w $L^n_{F2\nu_k}(F^k)$. Widać więc, że macierz

$$\mathbf{G}(\vartheta_{[l]}) = \mathbf{G}(\chi_{[l]}),$$

gdzie $\mathbf{G}(\chi_{[l]})$ oznacza macierz Grama względem iloczynu skalarnego $\langle \cdot, \cdot \rangle_{\nu_k}$, a zatem z liniowej niezależności układu $\chi_{[l]}$ wynika liniowa niezależność układu $\vartheta_{k[l]}$.

(2) Przestrzeń $L^n_{F2\mu}(L^m_{F2\varkappa}(R))$ z iloczynem skalarnym $\langle \cdot, \cdot \rangle_\mu$ jest przestrzenią unitarną, w której przekształcenia $(\vartheta_{k,i})_1^l$ tworzą układ liniowo niezależny. Z teorii aproksymacji średniokwadratowej ([12], s. 91) wynika, że $\forall \varphi \in L^n_{F2\mu}(L^m_{F2\varkappa}(R))$ istnieje dokladnie jedno przybliżenie średniokwadratowe $\gamma_{k,l}$ w przestrzeni span $\vartheta_{k[l]}$ i ma ono postać:

$$\gamma_{k,l} = (G^{-1}(\vartheta_{k[l]}) \langle \vartheta_{k[l]}, \varphi \rangle_\mu)^* \vartheta_{k[l]}.$$

W dowodzie części (1) tego twierdzenia wykazaliśmy już, że macierz Grama

$$\mathbf{G}(\vartheta_{k[l]}) = \mathbf{G}(\chi_{[l]}),$$

a stąd wynika teza.

---

(12) Jeżeli $f: X \to F^n$ oraz $g := (g_i: X \to F^n)_I$, to dla uproszczenia zapisu $\forall x \in X$

$$[F^T g](x) := (f^T(x) g_i(x))_I.$$

(3) Zauważmy na wstępie, że z dokonanych założeń przestrzenie $L^m_{F2\varkappa\nabla}$, $L^m_{F2\varkappa*}$ oraz $L^m_{F2\varkappa}$ są równoważne ([10]). Dowód liniowej niezależności elementów należących do układu $\vartheta_{k[l]} \circ \nabla.$ przeprowadzimy nie wprost. Załóżmy zatem, że elementy układu $\vartheta_{k[l]} \circ \nabla.$ są operatorami liniowo zależnymi. Obrazem produktu kartezjańskiego $L^m_{F2\varkappa} \times R$ w odwzorowaniu $\nabla.$ jest przestrzeń $L^m_{F2\varkappa}$ ([11]). Wynikałoby stąd, że układ $\vartheta_{k[l]}$ jest również liniowo zależny, co przeczy dowiedzionej już części (1) tego twierdzenia. Z kolei elementy układu $\vartheta_{k[l]}$ są przekształceniami przestrzeni $L^m_{F2\varkappa}$ w F^n, a zatem elementy układu $\vartheta_{k[l]} \circ \nabla.$ są operatorami stacjonarnymi ([9]).

Ponieważ $\forall u \in L^m_{F2\varkappa}$ spełniony jest warunek:

$$\vartheta_{k[l]}(u) = \vartheta_{k[l]}(\nabla^*(u_{-1})u),$$

gdyż $\upsilon_{[k]}$ jest układem liniowo niezależnym w przestrzeni $L^m_{F2\varkappa+}$, to wynika stąd, iż elementy układu $\vartheta_{k[l]} \circ \nabla.$ są operatorami przyczynowymi ([9]).

(4) Zauważmy, że $\forall u \in L^m_{F2\varkappa}$ oraz $\forall t \in R$ mamy:

$$[\vartheta_{k[l]} \circ \nabla_t](u) = \chi_{[l]}(\bar{G}^{-1}(\upsilon_{[k]}) \langle \nabla^*_t(u)\upsilon_{[k]} \rangle_\varkappa) =$$

$$= \chi_{[l]}(\bar{G}^{-1}(\upsilon_{[k]})\,\overline{[\omega\upsilon_{[k]} * u]}(t)).$$

Wykażemy na wstępie, że $\forall u \in L^m_{F2\varkappa}$ elementy wchodzące w skład układu $[\vartheta_{k[l]} \circ \nabla.](u)$ są $\varkappa$-mierzalne. W tym celu zdefiniujmy przekształcenie $\psi : L^m_{F2\varkappa} \times R \to F^k$ w następujący sposób:

$$\psi(u,t) := \bar{G}^{-1}(\upsilon_{[k]})\,\overline{[\sqrt{\mathscr{W}}\upsilon_{[k]} * u]}(t).$$

Wystarczy teraz pokazać, że odwzorowanie $\psi(u, \cdot)$ jest $\varkappa$-mierzalne, a elementy układu $\chi_{[l]}$ są funkcjami borelowskimi[13] ([2], s. 182).

Odwzorowanie $\psi(u, \cdot)$ jest $\forall u \in L^m_{F2\varkappa}$ $\varkappa$-mierzalne na R, co wynika z literatury ([20], s. 325 i [10]) oraz spostrzeżenia, że $\forall t \in R$ funkcje $\sqrt{\mathscr{W}}(\nabla^*_t(u)) \in L^m_{F2}$ i $(\sqrt{\mathscr{W}}\upsilon_i \in L^m_{F2})^k_1$, gdyż przestrzenie $L^m_{F2\varkappa\nabla}$ oraz $L^m_{F2\varkappa}$ są równoważne. Z kolei iloczyn skalarny jest odwzorowaniem ciągłym ([20], s. 69), czyli przekształcenie Φ_k jest również ciagłe na przestrzeni $L^m_{F2\varkappa}$, a więc jest μ-mierzalne, gdyż miara μ jest określona na σ-ciele zbiorów borelowskich ([26], s. 21). Miara

$$\nu_k = \mu \circ \Phi_k^{-1}$$

jest wyznaczona w F^k przez przekształcenie Φ_k i miarę μ, a zatem określona jest na σ-ciele zbiorów borelowskich ([2], s. 182 i 185). Elementy układu $\chi_{[l]}$ są więc funkcjami borelowskimi, gdyż należą do przestrzeni $L^n_{F2\nu_k}$. Z rozumowania tego wynika ostatecznie, że $\forall u \in L^m_{F2\varkappa}$ elementy układu

$$[\vartheta_{k[l]} \circ \nabla.](u) = [\chi_{[l]} \circ \psi](u, \cdot)$$

są $\varkappa$-mierzalne ([2], s. 182).

[13] *Odwzorowanie* mierzalne względem σ-ciała zbiorów borelowskich będziemy nazywać *borelowskim* ([26], s. 21).

Zauważmy teraz, że $\forall i=1,...,l$ oraz $\forall u \in L^m_{F2\varkappa}$ całka

$$\int_R |\vartheta_{k,i}(\nabla_t(u))|^2 \mathrm{d}\varkappa(t) = \int_R |\chi_i(\psi(u,t))|^2 \mathrm{d}\varkappa(t).$$

Ponieważ dla każdego $u \in L^m_{F2\varkappa}$ odwzorowanie $\psi(u,\cdot)$ jest ograniczone na R ([20], s. 325 i [10]), to $\exists M_u \in R_+ \forall t \in R$, że moduł

$$|\psi(u,t)| \leqslant M_u.$$

Niech

$$A_u := \{x \in F^k \; A_u := \{x \in F^k\colon |x| \leqslant M_u\} \; |x| \leqslant M_u\}.$$

Ponieważ zbiór A_u jest ograniczony, to z założenia, że $(\chi_i)_1^l$ są operatorami lokalnie ograniczonymi wynika, że obrazy $(\chi_i(A_u))_1^l$ są również ograniczone, czyli $\forall i=1,...,l \; \exists M_i \in R_+$, że

$$|\chi_i(A_u)| \leqslant M_i.$$

Mamy zatem, że $\forall i=1,...,l$ oraz $\forall u \in L^m_{F2\varkappa} \; \exists M_i \in R_+$, że całka

$$\int_R |\vartheta_{k,i}(\nabla_t(u))|^2 \mathrm{d}\varkappa(t) \leqslant M_i \int_R \mathrm{d}\varkappa = M_i \varkappa(R),$$

a stąd oraz założenia, że $\varkappa$ jest miarą skończoną wynika, że $\forall u \in L^m_{F2\varkappa}$ oraz $\forall i=1,...,l$ funkcja $\vartheta_{k,i}(\nabla.(u)) \in L^n_{F2\varkappa}$.

Z przedstawionego rozumowania mamy również, że operatory $(\vartheta_{k,i} \circ \nabla.)_1^l$ są lokalnie ograniczone oraz $\varkappa$-stabilne względem normy $L_{F2\varkappa}$.

(5) Można wykazać ([11]), iż dla każdego przekształcenia $\varphi \in C^n_{F\infty\mu}(L^m_{F2\varkappa}(R))$, dla którego założenie $\varphi \circ \nabla.$ jest operatorem stacjonarnym przyczynowym, istnieje taka funkcja $f \in C^n_{F\infty\nu}(l_{F2})$, że

$$\varphi = f \circ \Phi \circ \nabla^* \quad \text{oraz} \quad \Phi := \langle \cdot, \upsilon \rangle_\varkappa.$$

Przestrzeń $C^n_{F\infty\mu}(L^m_{F2\varkappa}(R)) \subset L^n_{F\infty\mu}(L^m_{F2\varkappa}(R))$. Ponieważ założyliśmy, że ν jest miarą skończoną na l_{F2} oraz

$$\nu \times \nu = \mu \circ \Gamma^{-1},$$

to miara μ jest również skończona ([2], s. 185) i wówczas mamy, że przestrzeń $L^n_{F\infty\mu}(L^m_{F2\varkappa}(R) \subset L^n_{F2\mu}(L^m_{F2\varkappa}(R))$ ([20], s. 54), czyli $C^n_{F\infty\mu}(L^m_{F2\varkappa}(R)) \subset L^n_{F2\mu}(L^m_{F2\varkappa}(R))$. Z kolei przestrzeń $C^n_{F\infty\nu} \subset L^n_{F\infty\nu}$, a z założenia, że ν jest miarą skończoną na l_{F2} wynika, że $L^n_{F\infty\nu} \subset L^n_{F2\nu}$ ([20], s. 54), czyli $C^n_{F\infty\nu} \subset L^n_{F2\nu}$. Mamy zatem, że przekształcenie $\varphi \in L^n_{F2\mu}(L^m_{F2\varkappa}(R))$, a funkcja $f \in L^n_{F2\nu}$.

Niech $\Pi_k : F^k \to l_{F2}$ będzie takim odwzorowaniem, że $\forall x_{[k]} := (x_i \in F)_1^k$

$$\Pi_k(x_{[k]}) := \left(\xi_i := \begin{cases} x_i, & i \leqslant k \\ 0, & i > k \end{cases}\right)_N.$$

Niech ponadto $\forall x := (x_i)_N \in l_{F2}$ odwzorowanie $\pi_k : l_{F2} \to F^k$ ma postać:

$$\pi_k(x) := (x_i)_1^k.$$

Ponieważ można wykazać ([18], s. 120), że $\forall x \in l_{F2}$

$$\lim_{k \to \infty} \| x - \Pi_k(\pi_k(x)) \|_\infty = 0,$$

gdzie $\| \cdot \|_\infty$ jest normą w l_{F2}, to z ciągłości funkcji f wynika, iż

$$\lim_{k \to \infty} \| f(x) - f(\Pi_k(\pi_k(x))) \| = 0,$$

gdzie $\| \cdot \|$ jest normą w F^n. Zdefiniujmy teraz funkcję $f_k : F^k \to F^n$ w następujący sposób:

$$f_k := f \circ \Pi_k.$$

Wówczas

$$\lim_{k \to \infty} \| f(x) - f_k(\pi_k(x)) \| = 0,$$

czyli

$$f \circ \Phi \circ \nabla^* = \lim_{k \to \infty} f_k \circ \pi_k \circ \Phi \circ \nabla^*.$$

Ponieważ υ jest układem ortogonalnym, to macierz $\overline{\mathbf{G}}^{-1}(\upsilon)$ jest macierzą jedynkową i przekształcenie

$$\Phi_k = \pi_k \circ \Phi.$$

Wynika stąd, że przekształcenie $f_k \circ \Phi_k \circ \nabla^*$ spełnia warunek:

$$f \circ \Phi \circ \nabla^* = \lim_{k \to \infty} f_k \circ \Phi_k \circ \nabla^*,$$

a zatem

$$\varphi = \lim_{k \to \infty} f_k \circ \Phi_k \circ \nabla^*.$$

By móc dalej wykorzystać twierdzenie Lebesgue'a o zbieżności zmajoryzowanej ([2], s. 209) wykażemy, że $\forall k \in N$ przekształcenie $f_k \circ \Phi_k \circ \nabla^*$ jest μ-mierzalne, a nawet należy do $L^n_{F2\mu}(L^n_{F2\varkappa}(R))$.

Można łatwo sprawdzić, że

$$F^k \times l_{F2} = l_{F2}.$$

Jeżeli $\mathbf{0} \in l_{F2}$, to $\forall x_k \in F^k$ odwzorowanie

$$\Pi_k(x_k)=(x_k^T, \mathbf{0}^T)^T.$$

Dowodzi się, że odwzorowanie takie jest ciągłe, a nawet jest homeomorfizmem ([19], s. 152). W poprzedniej części tego twierdzenia wykazaliśmy, że miara ν_k jest określona na σ-ciele zbiorów borelowskich w F^k. Odwzorowanie Π_k jest zatem ν_k-mierzalne ([2], s. 182). Podobnie jak poprzednio wykazaliśmy, że miara ν_k jest określona na σ-ciele zbiorów borelowskich w F^k, można wykazać, iż ν jest miarą na σ-ciele zbiorów borelowskich w l_{F2}. Ponieważ funkcja f jest ν-mierzalna, to wynika stąd, że $\forall k \in N$ funkcja

$$f_k := f \circ \Pi_k$$

jest ν_k-mierzalna ([2], s. 182). Z wykazanej w poprzedniej części μ-mierzalności przekształcenia Φ_k mamy więc, że złożenie $f_k \circ \Phi_k$ jest również μ-mierzalne ([2], s. 182). Z kolei założonej niezmienniczości miary μ względem obrotu funkcyjnego dostajemy, że $\forall k \in N$ założenie $f_k \circ \Phi_k \circ \nabla^*$ jest μ-mierzalne. Ponieważ miara μ jest skończona, a przekształcenie φ ograniczone, to $\forall k \in N$ założenie $f_k \circ \Phi_k \circ \nabla^*$ jest również ograniczone i tym samym należy do $L^n_{F2\mu}(L^n_{F2\varkappa}(R))$.

Z części (2) twierdzenia mamy, że istnieje przybliżenie średniokwadratowe $\gamma_{k,l} \in \operatorname{span} \vartheta_{k[l]}$ przekształcenia φ, dla którego błąd

$$\rho_{k,l} = \| \varphi - \gamma_{k,l} \|_\mu .$$

Zbadamy teraz właściwości tego błędu.

Przybliżenie średniokwadratowe $\tilde{\gamma}_{k,l}$, które należy do przestrzeni $\operatorname{span} \vartheta_{k[l]}$, przekształcenia $f_k \circ \Phi_k \circ \nabla^*$, jak wynika z części (2), ma postać:

$$\tilde{\gamma}_{k,l} = (G^{-1}(\chi_{[l]}) \langle \vartheta_{k[l]}, f_k \circ \Phi_k \circ \nabla^* \rangle_\mu)^* \vartheta_{k[l]}.$$

Przybliżenie to na ogół nie jest przybliżeniem średniokwadratowym przekształcenia φ, a zatem błąd

$$\rho_{k,l} \leqslant \| \varphi - \tilde{\gamma}_{k,l} \|_\mu = \| \varphi - f_k \circ \Phi_k \circ \nabla^* + f_k \circ \Phi_k \circ \nabla^* - \tilde{\gamma}_{k,l} \|_\mu \leqslant$$

$$\leqslant \| \varphi - f_k \circ \Phi_k \circ \nabla^* \|_\mu + \| f_k \circ \Phi_k \circ \nabla^* - \tilde{\gamma}_{k,l} \|_\mu .$$

Niech

$$\rho_k := \| \varphi - f_k \circ \Phi_k \circ \nabla^* \|_\mu \quad \text{oraz} \quad \rho_l := \| f_k \circ \Phi_k \circ \nabla^* - \tilde{\gamma}_{k,l} \|_\mu .$$

Wówczas błąd

$$\rho_{k,l} \leqslant \rho_k + \rho_l .$$

Zauważmy, że z założenia, miara μ jest skończona, a przekształcenie φ, ograniczone, a zatem dla każdego $k \in N$ istnieje całkowalna majoranta odwzorowania $| f_k \circ \Phi_k \circ \nabla^* |$. Z twierdzenia Lebesgue'a o zbieżności zmajoryzowanej ([2], s. 209) mamy, że granica

$$\lim_{k \to \infty} \| \varphi - f_k \circ \Phi_k \circ \nabla^* \|_\mu = 0 ,$$

a stąd

$$\lim_{k\to\infty} \rho_k = 0.$$

Zajmijmy się teraz drugim składnikiem błędu średniokwadratowego. W tym celu pokażemy, że $\forall k \in N$ funkcja $f_k \in L^n_{F2\nu_k}$. Wykazaliśmy już wcześniej, że funkcja f_k jest ν_k-mierzalna, a zatem z niezmienniczości miary μ względem obrotu funkcyjnego i z twierdzenia o zamianie zmiennych w całce Lebesgue'a ([2], s. 214) wynika, że $\forall k \in N$, spełniona jest równość:

$$\| f_k \circ \Phi_k \circ \nabla^* \|_\mu = \| f_k \circ \Phi_k \|_\mu = \| f_k \|_{\nu_k},$$

gdzie $\| \cdot \|_{\nu_k}$ jest normą wyznaczoną przez iloczyn skalarny w przestrzeni $L^n_{F2\nu_k}$. Ponieważ $\forall k \in N$ przekształcenie $f_k \circ \Phi_k \circ \nabla^* \in (L^n_{F2\nu}(R))$, to funkcja $f_k \in L^n_{F2\nu_k}$. Z niezmienniczości miary μ względem obrotu funkcyjnego oraz z twierdzenia o zamianie zmiennych w całce Lebesgue'a wynika, że

$$\tilde{\gamma}_{k,l} = (G^{-1}(\chi_{[l]})\langle \chi_{[l]} \circ \Phi_k \circ \nabla^*, f_k \circ \Phi_k \circ \nabla^* \rangle_\mu)^* \vartheta_{k[l]} =$$

$$= (G^{-1}(\chi_{[l]})\langle \chi_{[l]}, f_k \rangle_\nu)^* \vartheta_{k[l]} =$$

$$= (G^{-1}(\chi_{[l]})\langle \chi_{[l]}, f_k \rangle_{\nu_k})^* \chi_{[l]} \circ \Phi_k \circ \nabla^* = g_{k,l} \circ \Phi_k \circ \nabla^*.$$

Układ liniowo niezależny χ jest zgodnie z założeniem pełny w przestrzeni $L^n_{F2\nu_k}$. Stąd i z równości Parsevala ([12], s. 94) wynika, że przybliżenie średniokwadratowe $\tilde{\gamma}_{k,l}$ spełnia warunek:

$$\lim_{l\to\infty} \| f_k \circ \Phi_k \circ \nabla^* - \tilde{\gamma}_{k,l} \|_\mu = \lim_{l\to\infty} \| f_k - g_{k,l} \|_{\nu_k} = 0,$$

a zatem

$$\lim_{l\to\infty} \rho_l = 0.$$

Ponieważ

$$\rho_{k,l} = \| \varphi - \gamma_{k,l} \|_\mu \leqslant \rho_k + \rho_l,$$

to ostatecznie mamy, że

$$\lim_{k,l\to\infty} \| \varphi - \gamma_{k,l} \|_\mu = 0.$$

(6) Wykażemy na wstępie, że przekształcenie φ należy do przestrzeni $L^n_{F2\mu}(L^m_{F2\varkappa}(R))$. W tym celu zauważmy, że $\forall u \in L^m_{F2\varkappa}$ norma

$$\| \Phi(\nabla^*(u)) \|_\infty \leqslant \| \Gamma(\nabla^*(u)) \|_\infty = \| u \|_\varkappa,$$

gdzie $\|\cdot\|_{\varkappa}$ jest normą w $L^m_{F2\varkappa}$, gdyż υ jest bazą ortonormalną w przestrzeni $L^m_{R2\varkappa+}$. Wynika stąd, że $\Phi \circ \nabla^*$ jest liniowym przekształceniem ograniczonym na $L^m_{F2\varkappa}$, a zatem jest przekształceniem ciągłym, co wnioskujemy z twierdzenia Banacha o przekształceniu ciągłym ([20], s. 139). Ponieważ wykazaliśmy wcześniej, że ν jest miarą na σ-ciele zbiorów borelowskich w l_{F2}, to odwzorowanie $f \in L^n_{F2\nu}$ jest borelowskie, zatem operator φ jest ν-mierzalny ([2], s. 182). Z kolei z niezmienniczości miary μ względem obrotu funkcyjnego oraz z twierdzenia o zamianie zmiennych w całce Lebesgue'a ([2], s. 214), mamy że

$$\|\varphi\|_{\mu} = \|f \circ \Phi \circ \nabla^*\|_{\mu} = \|f \circ \Phi\|_{\mu} = \nu(l_{F2})\,\|f\|_{\nu} < \infty,$$

gdzie $\|\cdot\|_{\nu}$ jest normą w przestrzeni $L^n_{F2\nu}$, a stąd wynika, że $\varphi \in L^n_{F2\mu}(L^m_{F2\varkappa}(R))$.

Z części (2) tego twierdzenia wynika, że wystarczy wykazać równość:

$$\langle \vartheta_{k[l]}, \varphi \rangle_{\mu} = A(\varphi^*(\nabla_{\cdot}(\xi))\vartheta_{k[l]}(\nabla_{\cdot}(\xi))).$$

Iloczyn skalarny

$$\langle \vartheta_{k[l]}, \varphi \rangle_{\mu} = \langle \chi_{[l]} \circ \Phi_{k[l]} \circ \nabla^*, f \circ \Phi \circ \nabla^* \rangle_{\mu}.$$

Ponieważ υ jest układem ortogonalnym, to macierz $\overline{\mathbf{G}}^1(\upsilon)$ jest diagonalna, czyli

$$\Phi_k = \pi_k \circ \Phi,$$

a zatem

$$\langle \vartheta_{k[l]}, \varphi \rangle_{\mu} = \langle \chi_{[l]} \circ \pi_k \circ \Phi \circ \nabla^*, f \circ \Phi \circ \nabla^* \rangle_{\mu}.$$

By wykorzystać teraz twierdzenie o zamianie zmiennych w całce Lebesgue'a ([2], s. 214) należy wykazać, że $\chi_{[l]} \circ \pi_k$ jest układem odzworowań ν-mierzalnych.

Zauważmy, że odwzorowanie $\pi_k : l_{F2} \to F^k$ jest ciągłe. Wynika stąd, że jest ono również ν-mierzalne, gdyż zgodnie z tym co powiedziano w części (5), ν jest miarą na σ-ciele zbiorów borelowskich w l_{F2}. W części (4) wykazaliśmy już, że ν_k jest miarą na σ-ciele zbiorów borelowskich w F^k. Z ν_k-mierzalności elementów układu $\chi_{[l]}$ wynika więc, że założenie $\chi_{[l]} \circ \pi_k$ jest rzeczywiście układem odwzorowań ν-mierzalnych ([2], s. 182).

Z niezmienniczości miary μ względem obrotu funkcyjnego i twierdzenia o zamianie zmiennych w całce Lebesgue'a ([2], s. 214) mamy, że iloczyn skalarny

$$\langle \vartheta_{k[l]}, \varphi \rangle_{\mu} = \langle \chi_{[l]} \circ \pi_k, f \rangle_{\nu},$$

gdzie $\langle \cdot, \cdot \rangle_{\nu}$ jest iloczynem skalarnym w $L^n_{F2\nu}$. Z kolei, z założenia ergodyczności sygnałów $\overline{\varkappa\upsilon} * \xi$ dostajemy, że

$$\langle \chi_{[l]} \circ \pi_k, f \rangle_{\nu} = \int_{l_{F2}} f^* \chi_{[l]} \circ \pi_k \, d\nu = A(f^*([\overline{\varkappa\upsilon} * \xi](\cdot))\chi_{[l]}(\pi_l([\overline{\varkappa\upsilon} * \xi](\cdot)))) =$$

$$= A(f^*([\overline{\varkappa\upsilon} * \xi](\cdot))\chi_{[l]}([\overline{\varkappa\upsilon}_k * \xi](\cdot))) = A(f^*(\Phi(\nabla^*_{\cdot}(\xi)))\vartheta_{k[l]}(\nabla_{\cdot}(\xi))) =$$

$$= A(\varphi^*(\nabla_\cdot(\xi))\vartheta_{k[l]}(\nabla_\cdot(\xi))),$$

co kończy dowód. ■

5. PODSUMOWANIE

Załóżmy, że

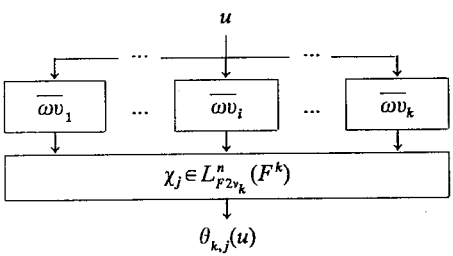
$$L^m_{F2x\nabla} = L^m_{F2x*} = L^m_{F2x}$$

oraz $\varphi \in C^n_{F\infty\mu}(L^m_{F2x}(R))$ jest przekształceniem ograniczonym, dla którego założenie $\varphi \circ \nabla$ jest operatorem stacjonarnym przyczynowym, o którym mowa w części (5) twierdzenia. Wynika wówczas, że każdy operator $\varphi \circ \nabla$ można dowolnie dokładnie przybliżyć za pomocą liniowej kombinacji operatorów o postaci:

$$[\theta_{k[l]}(u)](t) := ([\theta_{k,i}(u)](t)]_1^l := \vartheta_{k[l]}(\nabla_t(u)) = \chi_{[l]}([\overline{\mathscr{W}\upsilon}_{[k]} * u](t)),$$

gdzie $\upsilon_{[k]}$ jest układem ortonormalnym w L^m_{F2x+}. Strukturę tych operatorów pokazano na rys. 1. Wektor współczynników liniowej kombinacji

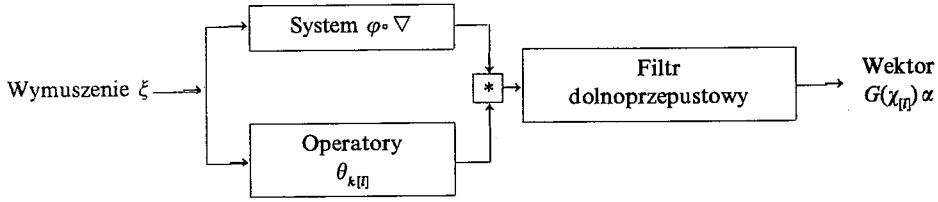
$$\alpha := G^{-1}(\chi_{[l]})A[\varphi^*(\nabla_\cdot(\xi))\vartheta_{k[l]}(\nabla_\cdot(\xi)))$$



$u \in L^m_{F2x}(R)$
$dx := \mathscr{W}\,\mathrm{d}\lambda$
$[\theta_{k,j}(u)](t) := \chi_j([\overline{\mathscr{W}\upsilon}_{[k]} * u](t))$

Rys. 1. Struktura operatora $\theta_{k,j}$

możemy wyznaczyć, zgodnie z częścią (6) twierdzenia, na podstawie znajomości wartości średniej iloczynu odpowiedzi operatorów $\varphi^* \circ \nabla_\cdot$ oraz $\theta_{k[l]}$ na wymuszenie ξ dla którego sygnały $\overline{\omega\upsilon} * \xi$ są ergodyczne. Współczynniki te mogą być zatem wyznaczone w bardzo prostym układzie przedstawionym na rys. 2.



Rys. 2. Schemat blokowy układu do eksperymentalnego wyznaczania wektora α

W jednym z następnych artykułów zobaczymy, że przybliżenia systemów nieliniowych, które wykorzystują układ operatorów $\vartheta_{k[l]} \circ \nabla$., obejmują klasę modeli Wienera oraz klasę modeli w postaci szeregów Volterry.

BIBLIOGRAFIA

1. J.S. Bendat, A.G. Piersol (tłum. z ang. J. Dudziewicz, R. Białobrzeski): *Metody analizy i pomiaru sygnałów losowych.* Biblioteka Naukowa Inżyniera, PWN, Warszawa 1976
2. P. Billingsley (tłum. z ang. K. Kizewetr, J.E. Rogulski): *Prawdopodobieństwo i miara.* PWN, Warszawa 1987
3. G. Ciesielski, S. Derlecki, Z. Marks-Wojciechowska: *Identyfikacja wielowymiarowych systemów pomiarowych opisanych operatorem nieliniowym.* Praca badawcza nr I12/5/90 BP, Politechnika Łódzka, Łódź 1990
4. G. Ciesielski, S. Derlecki, Z. Marks-Wojciechowska: *Identyfikacja i korekcja nieliniowa wielowymiarowych systemów pomiarowych opisanych operatorem nieliniowym.* Praca badawcza nr MEN I12/6/90 BP, Politechnika Łódzka, Łódź 1990
5. G. Ciesielski: *Aproksymacja średniokwadratowa operatorów nieliniowych opisujących wielowymiarowe systemy pomiarowe.* XXIII Międzyuczelniana Konferencja Metrologów, Politechnika Warszawska, Warszawa 1991 (referat nie opublikowany)
6. G. Ciesielski: *Modele Wienera wielowymiarowych systemów pomiarowych opisanych za pomocą operatorów nieliniowych.* Modelowanie i symulacja systemów pomiarowych, Materiały Sympozjum, AGH, Kraków 1992, s. 41–50
7. G. Ciesielski: *Operatorowe modele wielowymiarowych systemów pomiarowych.* Seminarium Naukowe Komisji Kształcenia Komitetu Metrologii i Aparatury Naukowej Polskiej Akademii Nauk, 1994 (referat nie opublikowany)
8. G. Ciesielski: *Identyfikacja ogólnej struktury wielowymiarowych systemów pomiarowych.* Modelowanie i symulacja systemów pomiarowych. Materiały Sympozjum, AGH, Kraków, 1994, s. 17–23
9. G. Ciesielski: *Identyfikacja ogólnej struktury wielowymiarowych systemów stacjonarnych.* Kwart. Elektr. i Telekom. 1994, t. 40, z. 3, s. 419–428
10. G. Ciesielski: *Modelowanie wielowymiarowych systemów stacjonarnych za pomocą splotów.* Kwart. Elektr. i Telekom. 1994, t. 40, z. 3, s. 429–448
11. G. Ciesielski: *Identyfikacja szczegółowej struktury wielowymiarowych systemów stacjonarnych.* Kwart. Elektr. i Telekom. 1995, t. 41, z. 3, s. 305–319
12. G. Dahlquist, A. Björck (tłum. z ang. S. Paszkowski): *Metody numeryczne.* PWN, Warszawa 1983
13. P. Eykhoff (tłum. z ang. A. Bauer): *Identyfikacja w układach dynamicznych.* Biblioteka Naukowa Inżyniera, PWN, Warszawa 1980
14. R.J.P. de Figueiredo, T.A.W. Dwyer: *A best approximation framework and implementation for simulation of large-scale nonlinear systems.* IEEE Trans. Circuits Syst., 1980, vol. CAS-27, nr 11, s. 1005–1014
15. R.J.P. de Figueiredo: *Nonlinear circuits and systems applications of functional splines.* Proc. IEEE, 1982, s. 1245–1247
16. V.A. Illyin, E.G. Poznyak (tłum. z ros. I. Aleksandrowa): *Linear Algebra.* Mir, Moscow 1986
17. J. Jaworski, R. Morawski, J. Olędzki: *Wstęp do metrologii i techniki eksperymentu,* WNT, Warszawa 1992
18. W. Kołodziej: *Analiza matematyczna,* Matematyka dla politechnik, PWN, Warszawa 1979
19. K. Kuratowski: *Wstęp do teorii mnogości i topologii.* Biblioteka Matematyczna, t. 9, PWN, Warszawa 1980
20. J. Musielak: *Wstęp do analizy funkcjonalnej.* PWN, Warszawa 1989

21. J. Osiowski: *Zarys rachunku operatorowego. Teoria i zastosowania w elektrotechnice.* WNT, Warszawa 1981
22. A. Papoulis (tłum. z ang. T. Gerstenkorn): *Prawdopodobieństwo, zmienne losowe i procesy stochastyczne.* WNT, Warszawa 1972
23. J. Park, I.W. Sandberg: *Criteria for the approximation of nonlinear systems.* IEEE Trans. Circuits Syst. — I: Fundamental Theory and Applications, vol. 39, nr 8, 1992, s. 673–676
24. A. Plucińska, E. Pluciński: *Elementy probabilistyki,* Matematyka dla politechnik, PWN, Warszawa 1979
25. K.A. Pupkow, W.L. Kapalin, A.S. Juszczenko: *Funkcjonalnyje rjady w tjeorii nieliniejnych sistjem.* Nauka, Moskwa, 1976
26. W. Rudin (tłum. z ang. A. Pierzchalski, P. Walczak): *Analiza rzeczywista i zespolona.* PWN, Warszawa 1986
27. I.W. Sandberg: *Criteria for the response of nonlinear systems to be L-asymptotic periodic.* Bell Syst. Tech. J., 1981, vol. 60, nr 10, s. 2359–2371
28. I.W. Sandberg: *Series expansion for nonlinear systems.* Proc. IEEE, 1982, s. 110–113
29. I.W. Sandberg: *Expansion for nonlinear systems.* Bell Syst. Tech. J., 1982, vol. 61, nr 2, s. 159–199
30. I.W. Sandberg: *Volterra expansion for time-varying nonlinear systems.* Bell Syst. Tech. J., 1982, vol. 61, nr 2, s. 201–225
31. I.W. Sandberg: *Multidimensional nonlinear systems and structure theorems.* J. Circuits, Syst. and Computers, 1992, vol. 2, nr 4, s. 383–388
32. I.W. Sandberg: *Approximately-finite memory and input-output maps.* IEEE Trans. Circuits Syst. — I: Fundamental Theory and Applications, 1992, vol. 39, nr 7, s. 549–556
33. M. Schetzen: *The Volterra and Wiener Theories of Nonlinear Systems.* John Wiley & Sons, New York 1980

G. CIESIELSKI

MODELLING OF MULTIDIMENSIONAL TIME-INVARIANT AND CAUSAL SYSTEMS

Summary

This paper concern some problems related to the modelling of multidimensional time-invariant and causal systems governed by operators, which can be nonlinear in general case. It is assumed that the inputs and responses of considered systems are real or complex valued, and the set of inputs have the Hilbert space structure. There is proven theorem, which using theory of least mean square approximation allows to solve the modelling task of considered systems. Received results afford possibilities for better understanding of the essence of such complicated maps as operators, which govern multidimensional time-invariant and causal systems.

Key words: modelling, structure identification, parametric identification, multidimensional systems, time-invariant systems, causal systems, operators, convolution, stationary space, convolution space, causal space.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 3, ss. 337–347

Projektowanie scalonych wzmacniaczy operacyjnych w technologii CMOS

ZYGMUNT CIOTA

Instytut Elektroniki, Politechnika Łódzka

Otrzymano 1995.06.15

Autoryzowano do druku 1995.09.30

W artykule przedstawiona została metoda projektowania wzmacniaczy operacyjnych CMOS do filtrów analogowych sterowanych impulsami zegarowymi. Prędkość i dokładność analogowych układów scalonych jest ściśle związana z jakością wzmacniaczy operacyjnych. Proponowana metoda projektowania uwzględnia kilka prostych analitycznych równań, wykorzystuje reguły wynikające z technologii układów scalonych oraz symulacje komputerowe do końcowego wyznaczenia parametrów. Projektowany wzmacniacz wykonany został w technologii CMOS z kanałem 2 μm. Wyniki badań laboratoryjnych potwierdziły poprawność symulacji komputerowych. Przedstawione podejście pozwala projektować wzmacniacze operacyjne dla różnorodnych układów analogowych.

Słowa kluczowe: wzmacniacze operacyjne, układy scalone, symulacja komputerowa

1. WSTĘP

Układy analogowe stanowią nieodłączną część większości systemów wielkiej skali integracji CMOS. Ich rozpowszechnienie wiąże się w dużej mierze z wprowadzeniem nowych dziedzin elektroniki takich jak układy z przełączanymi pojemnościami, układy pracujące w trybie prądowym oraz systemy neuronopodobne. Najważniejszymi zaletami podejścia analogowego są duże prędkości i dokładność. W większości układów analogowych występują wzmacniacze operacyjne, których parametry, a szczególnie czas ustalania napięcia wyjściowego ograniczają dokładność oraz maksymalny zakres częstotliwości. Wymagania dużej dokładności i prędkości są w przypadku wzmacniacza operacyjnego wymaganiami przeciwstawnymi — polepszenie jednego z nich powoduje pogorszenie drugiego. Poprawę parametrów częstotliwościowych czyli wzrost częstotliwości przy której wzmocnienie jest równe jedności, można osiągnąć projektując wzmacniacz z małą liczbą stopni pośrednich, stosując

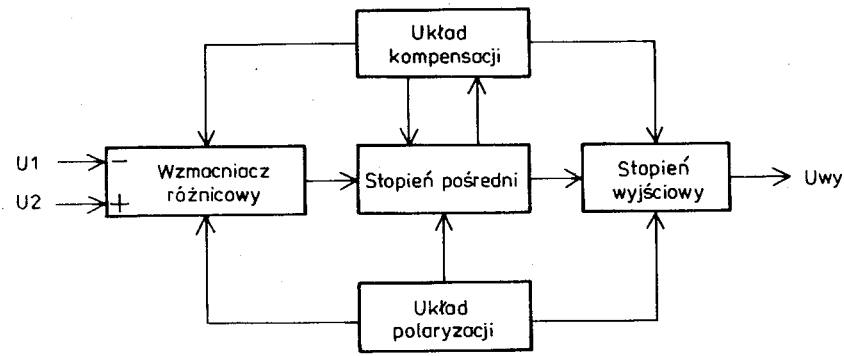
tranzystory o krótkich kanałach pracujących przy dużym prądzie polaryzacji. Z kolei maksymalne wzmocnienie stałoprądowe osiąga się we wzmacniaczach wielostopniowych z tranzystorami o długich kanałach i przy małych prądach polaryzacji.

Istnieje wiele rozwiązań układowych, których celem jest osiągnięcie jak najlepszego kompromisu między wymaganiami prędkości i wzmocnienia. Najważniejszym jest zastosowanie układów kaskody, która zmniejsza obciążenie pojemnościowe poprzedniego stopnia wzmacniacza oraz charakteryzuje się większą impedancją wyjściową w porównaniu z inwerterem [2]. Właściwości te pozwalają zwiększyć wzmocnienie stałoprądowe wzmacniacza bez istotnego pogorszenia jego własności częstotliwościowych. Inna metoda polega na stosowaniu pętli dodatniego sprzężenia zwrotnego [3], powoduje ona jednak wzrost prawdopodobieństwa wystąpienia oscylacji oraz niestabilności. Jeżeli wzmacniacz ma pracować w trybie przełączającym, można stosować również układy sterujące prądem polaryzacji w funkcji okresu przełączającego oraz w funkcji zmian amplitudy napięcia wyjściowego wzmacniacza, osiągając w ten sposób lepszy kompromis między zakresem częstotliwości a wzmocnieniem stałoprądowym [3]. Podejście to komplikuje proces projektowania, ponadto aby metoda była efektywna wymagana jest dobra znajomość sygnałów na wejściu i wyjściu wzmacniacza dla wyznaczenia funkcji sterującej prądem polaryzacji. Wzrasta również możliwość wystąpienia dodatkowych oscylacji w czasie przełączania.

Z przedstawionych rozważań wynika jasno, że zaprojektowanie dobrego wzmacniacza operacyjnego CMOS (np. o wzmocnieniu stałoprądowym większym od 80 dB z częstotliwością wzmocnienia jednostkowego większą od 200 MHz) przeznaczonego do zastosowań uniwersalnych, jest sprawą skomplikowaną. Ponieważ każdego roku pojawiają się nowe technologie CMOS z nowymi parametrami tranzystorów, koncepcja każdorazowego projektowania takiego wzmacniacza zastępowana jest często projektem ściśle związanym z konkretnym zastosowaniem [5]. Po dokładnym określeniu warunków pracy wzmacniacza, można go zaprojektować optymalnie, rezygnując z parametrów nieistotnych dla danego zastosowania, przy czym prowadzi to najczęściej do uproszczenia struktury zmniejszając tym samym powierzchnię układu scalonego. Dużą pomoc stanowi projektowanie wspomagane komputerem pozwalające wyeliminować błędy powstające we wstępnym procesie projektowania oraz umożliwiające poprzez wielokrotne symulacje ustalenie ostatecznych optymalnych wymiarów tranzystorów.

2. BLOKI FUNKCJONALNE WZMACNIACZA OPERACYJNEGO

Rysunek 1 pokazuje ogólny schemat blokowy wzmacniacza operacyjnego. Głównym zadaniem stopnia różnicowego jest wzmocnienie napięcia różnicowego czyli różnicy dwóch potencjałów mierzonych względem masy stałoprądowej DC (Direct Current) i zamiana tego napięcia na potencjał odniesiony do masy zmiennoprądowej AC (Alternate Current).



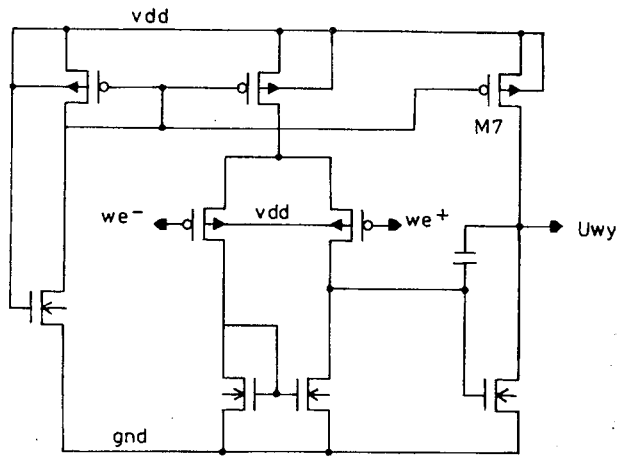
Rys. 1. Schemat blokowy wzmacniacza operacyjnego

Wzmacniacz różnicowy CMOS może być zbudowany z kilku stopni, aby na wyjściu uzyskać poziom napięcia wystarczająco duży do wysterowania następnego stopnia tzn. wzmacniacza pośredniego. Stopień pośredni charakteryzuje się największym wzmocnieniem napięciowym i spełnia dwa podstawowe zadania. Po pierwsze, wzmacnia sygnał wejściowy do poziomu umożliwiającego poprawne wysterowanie stopnia wyjściowego i po drugie zawiera obwody sprzężenia zwrotnego zapewniające stabilną pracę. Stopień wyjściowy decyduje o mocy wyjściowej wzmacniacza operacyjnego. Charakteryzuje się on wzmocnieniem bliskim jedności, dużą impedancją wejściową i bardzo małą impedancją wyjściową. Obwód polaryzacji wraz z obwodem kompensacyjnym, wyodrębnione na rysunku 1 jako bloki funkcjonalne, są zintegrowane z poprzednimi stopniami. Pierwszy z nich powinien być jak najbardziej niezależny od procesu technologicznego, czyli charakterystyka stałoprądowa wzmacniacza operacyjnego powinna być niewrażliwa na zmiany napięcia progowego V_T tranzystorów MOS. Obwód prądu polaryzacji powinien również uniezależniać parametry stałoprądowe od zmian napięcia zasilania i zmian temperatury. Obwód kompensacyjny składa się zazwyczaj z pojemności włączonej między stopień wyjściowy i pośredni. Wzmacniacz powinien być tak projektowany, aby pojemność ta była jak najmniejsza (pojedyncze pikofarady) nie zwiększając niepotrzebnie powierzchni układu scalonego.

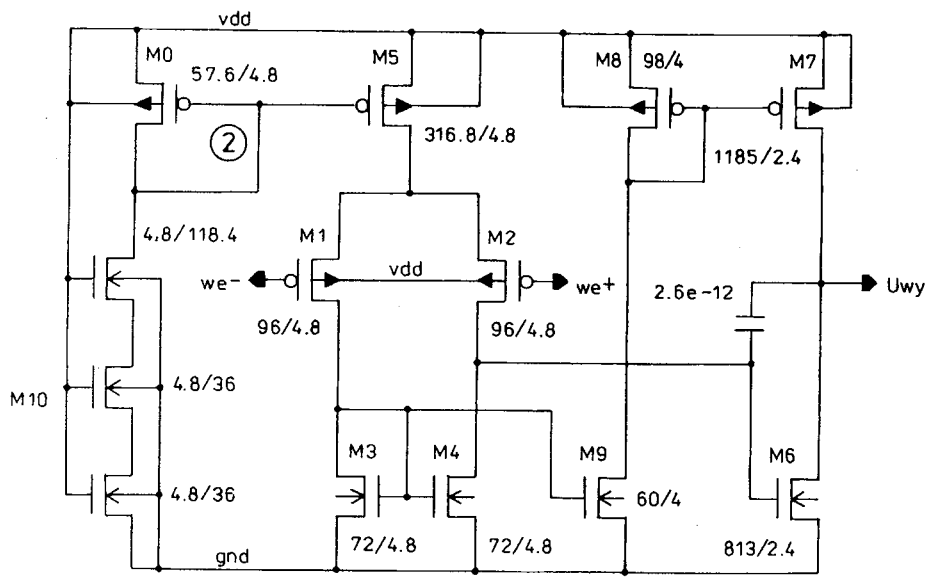
Metoda projektowania scalonego wzmacniacza operacyjnego w technologii CMOS polegająca na wyznaczeniu wymiarów wszystkich tranzystorów z analitycznych równań opisujących zjawiska fizyczne zachodzące w tych tranzystorach jest metodą bardzo pracochłonną. Źle dobrana, w stosunku do koniecznych do osiągnięcia parametrów, początkowa struktura obwodu może prowadzić do wykonywania wielu zbędnych iteracji przy próbach znalezienia nie istniejącego rozwiązania. Dlatego, przy praktycznym projektowaniu, dobre wyniki uzyskuje się korzystając z uproszczonych równań analitycznych dla wstępnego wyznaczenia wymiarów tranzystorów, a następnie poprzez symulację komputerową nastepuje dokładne wyznaczenie wymiarów wszystkich tranzystorów aż do osiągnięcia założonych charakterystyk i parametrów.

3. PRZYKŁAD PROJEKTOWANIA WZMACNIACZA OPERACYJNEGO

Punktem wyjścia do zaprojektowania wzmacniacza do układów z przełączanymi pojemnościami (układów analogowych sterowanych zegarami cyfrowymi) jest prosty wzmacniacz klasy A przedstawiony na rysunku 2. Dodając dodatkową gałąź sterującą tranzystorem M_7 otrzymujemy wzmacniacz pokazany na rysunku 3 [7]. Zwiększona została w ten sposób symetria stopnia wyjściowego, który pracuje teraz w pobliżu klasy AB. Oznacza to poprawę dynamiki charakterystyki przejściowej dla dużych sygnałów bez pogorszenia charakterystyki małosygnałowej.



Rys. 2. Wzmacniacza klasy A



Rys. 3. Schemat wzmacniacza operacyjnego

W dalszym etapie projektowania uwzględnione zostały nastepujące reguły wynikające z technologii CMOS [5, 7, 10]:

1. Dla stopnia wejściowego wybrana została polaryzacja kanałów typu P, ponieważ tranzystory te charakteryzują się mniejszym poziomem szumów.

2. Tranzystory stopnia wyjściowego pracujące przy stosunkowo dużych prądach i przyłączane najczęściej do innych obwodów znajdujących się na zewnątrz projektowanego układu scalonego, powinny być zlokalizowane jak najbliżej standardowych końcówek wejścia/wyjścia.

3. Nominalny prąd spoczynkowy stopnia wyjściowego powinien mieć wartość około 1 mA dla zapewnienia prędkości narastania napięcia wyjściowego (parametr slew rate) większej od 30 V/μs przy obciążeniu pojemnościowym 30 pF.

4. Wymiary (długość i szerokość kanału) tranzystorów stopnia wejściowego powinny być znacznie większe od minimalnych wymiarów dopuszczalnych dla danej technologii w celu zapewnienia jak najlepszej symetrii stopnia różnicowego i minimalizacji szumów.

5. Tranzystory wejściowe powinny być przystosowane do pracy przy małym napięciu bramka–źródło umożliwiające szeroki zakres pracy wzmacniacza różnicowego.

6. Stopień pośredni powinien mieć pośrednie długości kanałów, aby podobnie jak w układach cyfrowych, spełniać „prawo progresywnego buforowania" zapewniające minimalizację całkowitego czasu propagacji sygnału.

Do wyznaczenia wymiarów tranzystorów (parametrów L i W czyli długości i szerokości kanału tranzystora) wykorzystane zostały dwie zależności: pierwsza opisująca bilans prądowy oraz druga wyznaczająca równowagę prądów spoczynkowych.

Bilans prądowy

Przy zerowym napięciu wejściowym tranzystory M_1 i M_2 mają te same napięcia i prądy drenów, zatem M_6 i M_3 mają te same napięcia bramek. Wynika stąd, że prąd I_6 tranzystora M_6 jest określony następującą zależnością:

$$I_6 = I_3[(W_6/L_6)/(W_3/L_3)]. \tag{1}$$

Prąd płynący przez tranzystor M_1 jest kopiowany do tranzystora M_7 przez dwa źródła prądowe: $M_3 - M_9$ oraz $M_8 - M_7$; można zatem zapisać w postaci:

$$I_7 = I_1[(W_9/L_9)/(W_3/L_3)][(W_7/L_7)/(W_8/L_8)]. \tag{2}$$

Przyjmując zerowy prąd na wyjściu wzmacniacza otrzymuje się z równań (1) i (2) następującą zależność:

$$(W_6/L_6)(W_8/L_8) = (W_9/L_9)(W_7/L_7). \tag{3}$$

Równanie prądu spoczynkowego

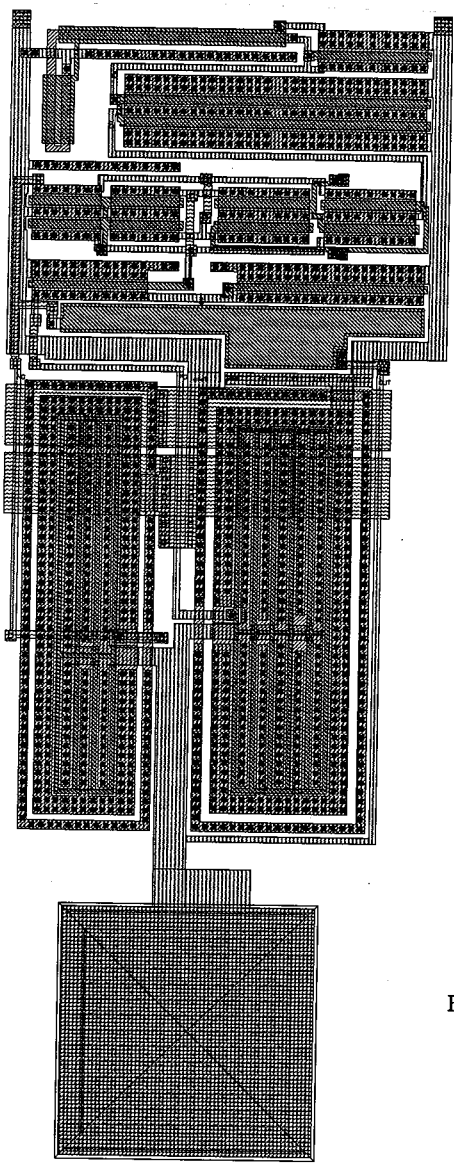
Prąd w tranzystorze M_1 jest dwukrotnie mniejszy od prądu tranzystora M_5. Z drugiej strony prąd I_5 można zapisać w postaci następującego iloczynu:

$$I_5 = I_0[(W_5/L_5)/(W_0/L_0)]. \quad (4)$$

Zatem wartość prądu spoczynkowego tranzystora M_6 jest następująca:

$$I_6 = 0{,}5I_0[(W_6/L_6)/(W_3/L_3)]\,[(W_5/L_5)/(W_0/L_0)] \quad (5)$$

W pierwszej kolejności przyjęte zostały wymiary wszystkich tranzystorów z wyjątkiem M_0, M_9 i M_{10}. Następnie, korzystając z równania (3) wyznaczono wymiary tranzystora M_9. Parametry tranzystora M_0 wyznaczone zostały z równania (5) przy założeniu, że wartość prądu polaryzacji jest równa 16 μA, a następnie określono



Rys. 4. Komplet masek scalonego wzmacniacza operacyjnego CMOS

wymiar tranzystora M_{10} przyjmując ten sam prąd polaryzacji i napięcie zasilania równe 5 V.

Kolejny krok polegał na wykonaniu szeregu symulacji komputerowych z wykorzystaniem pakietu programów SPICE, podczas których nastąpiła ostateczna korekta wymiarów tranzystorów zapewniająca bilans prądów i napięć, oraz sprawdzenie charakterystyk stałoprądowych wzmacniacza (wzmocnienie napięciowe, wartości prądów spoczynkowych, zakresy napięcia wejściowego i wyjściowego). Końcowe symulacje tak zaprojektowanego wzmacniacza pozwoliły wyznaczyć jego parametry;

Tablica 1

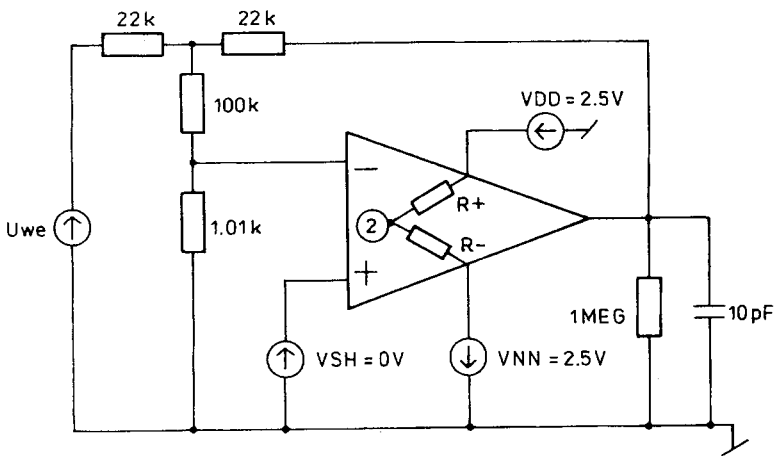
Parametry wzmacniacza operacyjnego

wzmocnienie stałoprądowe przy obciążeniu 1 MΩ	3000 V/V
wzmocnienie stałoprądowe przy obciążeniu 10 kΩ	1600 V/V
rezystancja wyjściowa	8,9 kΩ
częstotliwość graniczna przy obciążeniu 10 kΩ/30 pF	25 MHz

ich wartości zamieszczone są w Tablicy 1. Zaprojektowany wzmacniacz operacyjny wykonany został w technologii CMOS z kanałem 2 μm z podwójną warstwą metalizacji [10]. Schemat masek tego wzmacniacza przedstawia rysunek 4.

4. BADANIA LABORATORYJNE

Obwód pomiarowy do wyznaczania wzmocnienia napięciowego w otwartej pętli sprzężenia zwrotnego przedstawiony jest na rysunku 5 [8]. Przy projektowaniu wzmacniacza operacyjnego zapewniono dostęp do węzła Nr 2 pokazanego na rysunkach 2 i 5, przez przyłączenie go do końcówki zewnętrznej układu scalonego.



Rys. 5. Schemat układu do pomiaru wzmocnienia w otwartej pętli sprzężenia zwrotnego

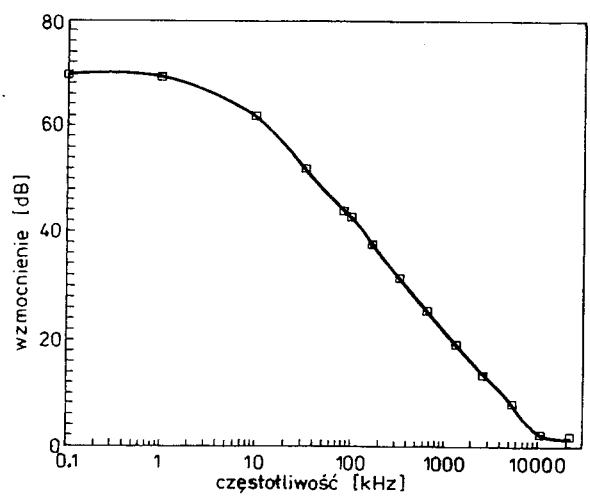
Tablica 2

Wzmocnienie, napięcie niezrównoważenia oraz prędkość narastania napięcia wyjściowego wzmacniacza operacyjnego

R^- [kΩ]	∞	∞	1000	470	220	100
R^+ [kΩ]	470	∞	∞	∞	∞	∞
wzmocnienie [V/V]	3600	3000	2500	2000	1550	1130
offset [mV]	−1,69	−1,70	−1,76	−1,80	−1,84	–
slew rate [V/μs]	14.6	17.5	29,2	35.0	58.3	87.5

Uzyskano w ten sposób możliwość regulacji prądu polaryzacji przez przyłączanie zewnętrznego rezystora między ten węzeł i źródło napięcia zasilającego (rezystory R^+ i R^- na rysunku 5). Wyniki pomiarów wzmocnienia stałoprądowego oraz napięcia niezrównoważenia (*offset voltage*) przy różnych prądach polaryzacji przy znamionowym napięciu zasilania przedstawione są w Tablicy 2. Bardzo małe, bo nie przekraczające 2 mV napięcie niezrównoważenia osiągnięte zostało poprzez wspólne centrowanie pary tranzystorów wzmacniacza różnicowego oraz nadanie im wymiarów L i W przekraczających minimalną dopuszczalną wartość 2 μm.

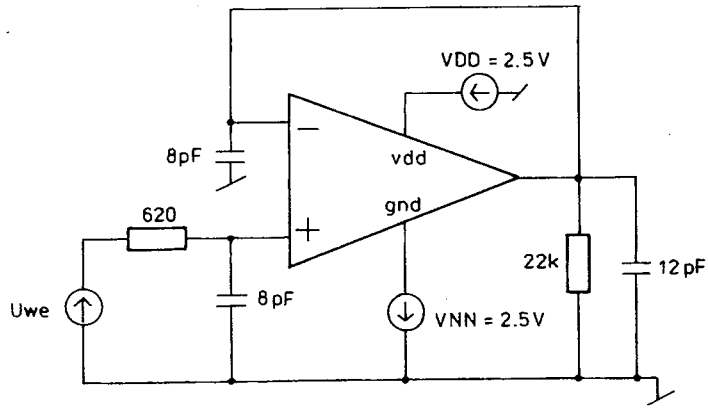
W tym samym obwodzie pomiarowym (rysunek 5) zmierzona została charakterystyka amplitudowa wzmacniacza — wyniki pomiarów pokazane są na rysunku 6.



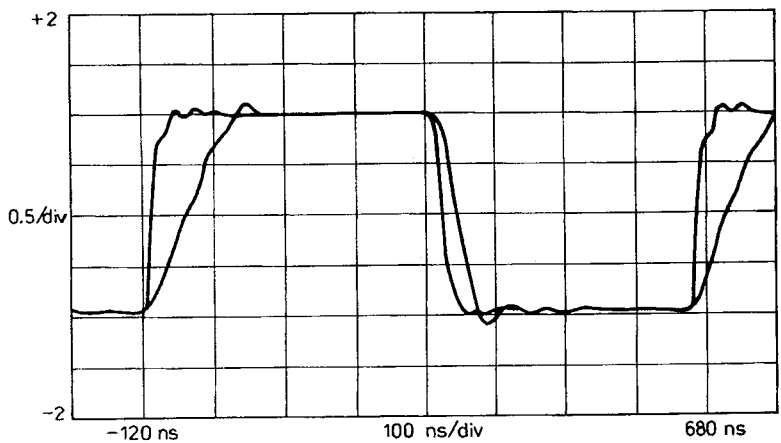
Rys. 6. Charakterystyka amplitudowa wzmacniacza

Wartość częstotliwości granicznej jest zgodna z wynikiem symulacji komputerowej przedstawionym w poprzednim rozdziale.

Schemat połączeń do badania dynamiki wzmacniacza w układzie ze zwartą pętlą sprzężenia zwrotnego pokazany jest na rysunku 7. Przykładowy wynik pomiaru



Rys. 7. Schemat układu do badania wzmacniacza ze zwartą pętlą sprzężenia zwrotnego



Rys. 8. Odpowiedź wzmacniacza jednostkowego na wymuszenie impulsem trapezowym

przedstawiony na rysunku 8 pokazuje przebieg napięcia wyjściowego w odpowiedzi na trapezowy impuls wejściowy.

Wpływ prądu polaryzacji na dynamiczne własności wzmacniacza operacyjnego jest uwidoczniony w Tablicy 2, gdzie zamieszczone zostały wyniki pomiarów czasu narastania napięcia wyjściowego (*parametr slew rate*) przy różnych wartościach rezystora polaryzującego.

PODSUMOWANIE

Projektowanie analogowych układów scalonych jest niewątpliwie trudniejsze od projektowania scalonych układów cyfrowych, przy czym można wyodrębnić dwie główne przyczyny wywołujące tą różnicę. Po pierwsze, dwustanowa logika układów

cyfrowych pozwala w sposób bardziej jednoznaczny dokonywać podziału systemów cyfrowych na standardowe, powtarzalne bloki podukładów. Po drugie, ponieważ do niedawna ogromna większość układów scalonych była układami czysto cyfrowymi, powstała duża liczba bibliotek standardowych komórek i większych od nich podukładów, optymalnych pod względem czasu propagacji oraz zajmowanej powierzchni płytki krzemowej, znakomicie ułatwiających proces projektowania. W przypadku układów analogowych biblioteki te są znacznie mniejsze, ponadto istnieją większe trudności przy przenoszeniu ich do nowo powstających technologii, właśnie z uwagi na analogowy charakter prądów i napięć nie poddający się łatwo procesowi standaryzacji.

Obserwowany od kilku lat wzrost produkcji układów analogowych a szczególnie analogowo-cyfrowych spowodował konieczność podjęcia nowych poszukiwań w celu ułatwienia procesu ich projektowania. Oznacza to przede wszystkim rozbudowę programów komputerowych pozwalających na symulację tych układów z zastosowaniem dokładnych modeli przyrządów półprzewodnikowych na poziomie zjawisk fizycznych z uwzględnieniem parametrów i ograniczeń konkretnej technologii, wybranej do produkcji projektowanego systemu. Dobrym przykładem jest program HSPICE współpracujący z pakietem programów CADENCE do projektowania scalonych układów analogowych [5].

Porównanie symulacji komputerowych z badaniami laboratoryjnymi wzmacniacza operacyjnego potwierdziło słuszność przyjętej koncepcji metody projektowania. Wykonując symulacje komputerowe programem PSPICE zaobserwowano jednak w niektórych przypadkach, np. przy wyznaczaniu wzmocnienia stałoprądowego, błędy przekraczające 10%. Przyczyną były niezbyt dokładne modele tranzystorow wbudowanych w ten program oraz brak możliwości uwzględnienia w tym programie elementów pasożytniczych. Powtórzenie symulacji programem HSPICE z uwzględnieniem parametrów występujących w dwumikrometrowej technologii CMOS dało wyniki zgodne z pomiarami.

PODZIĘKOWANIE

Scalony wzmacniacz operacyjny wykonany został w Institut National des Sciences Appliquées de Toulouse we Francji w ramach współpracy z Instytutem Elektroniki Politechniki Łódzkiej.

BIBLIOGRAFIA

1. M.T. Abuelmaatti: *Prediction of RFI demodulation in BiFET, BiMOS and CMOS operational amplifiers.* IEE Proceedings-G 1991, vol. 138, nr 1, p. 56
2. S. Aggarwal, A.B Bhattacharyya: *Low-frequency gain-enhanced CMOS operational amplifier.* IEE Proceedings-G 1991, vol. 138, nr 2, p. 170
3. K. Bult, G.J. Geelen: *A fast-settling CMOS op amp for SC circuits with 90-dB DC gain.* IEEE J. Solid-State Circuits 1990, vol. 25, nr 6, p. 1379

4. Z. Ciota, A. Napieralski, J.L. Noullet: *Theoretical analysis of two different SC-FIR realizations*. International Symposium on Signals, Systems and Electronics. Paris (France), September 1992, p. 218
5. S.L. Hurst: *Custom VLSI microelectronics*. Prentice-Hall International, New York 1992
6. H. Kwakernaak, R. Sivan: *Modern signal and systems*. Prentice-Hall International, New Jersey 1991
7. A. Napieralski, Z. Ciota, M. Napieralska, J.L. Noullet: *Design methodology of CMOS operational amplifiers*. Advanced Training Course: Mixed Design of VLSI Circuits, Dębe (Poland), April 1994, p. 130
8. A. Napieralski, J.L. Noullet, Z. Ciota: *CAD of CMOS operational amplifier (ASIC) using SPICE software*. XV Krajowa Konferencja Teorii Obwodów i Układów Elektronicznych, Szczyrk (Poland), October 1992, p. 209
9. G. Nicollini, D. Senderowicz: *A CMOS bandgap reference for differential signal processing*. IEEE J. Solid-State Circuitsf 1990, vol. 26, nr 1, p. 41
10. J.L. Noullet, A. Napieralski, Z. Ciota: *Switched-capacitor FIR filters: a unified approach*. VIII Congress of the Sociedade Brasileira de Microeletrônica, Campinas (Brazil), September 1993, p. II.11
11. J.L. Noullet, A. Napieralski, Z. Ciota: *ASIC realization of multiphase switched-capacitor decimating filter*. Bulletin of the Polish Academy of Sciences, Technical Sciences 1993, vol. 41, nr 1, p. 56
12. J. Vital, J.E. Franca, F. Maloberti: *Integrated mixed-mode digital-analog filter converters*. IEEE J. Solid-State Circuits 1990, vol. 25, nr 3, p. 660

Z. CIOTA

DESIGN OF INTEGRATED OPERATIONAL AMPLIFIERS IN CMOS TECHNOLOGY

Summary

Computer aided design of an operational amplifier in CMOS technology for sampled data filters is presented. Speed and accuracy of analog integrated circuits are closely dependent on performances of operational amplifiers. The proposed design method takes into account a few simple analytic equations, some qualitative trends and iterative computer simulations. The designed amplifier was made in 2 μm CMOS technology with the possibility to control of the bias current. The computer simulations and the experimentally obtained results are in a good agreement. The presented approach permits to design operational amplifiers for different analog circuits according to the individual requirements.

Key words: operational amplifiers, integrated circuits, computer simulation.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 3, ss. 349–360

Metody symulacji komputerowej analogowych filtrów scalonych

ZYGMUNT CIOTA

Instytut Elektroniki, Politechnika Łódzka

Otrzymano 1995.06.16

Autoryzowano do druku 1995.09.30

W artykule przedstawione zostały najważniejsze metody symulacji komputerowych wielofazowych filtrów z przełączanymi pojemnościami. Porównane zostały symulacje przeprowadzone różnymi programami analizy układów elektronicznych. Jeżeli wszystkie elementy sieci traktowane są jako idealne, zmodyfikowana metoda potencjałów węzłowych jest najbardziej efektywna do przeprowadzenia takiej analizy w dziedzinie czasu i częstotliwości. Dla wykonania bardziej precyzyjnych symulacji z uwzględnieniem niezerowych rezystancji kluczy w stanie załączenia oraz skończonej rezystancji w stanie otwarcia, gdy występują niezerowe stałe czasowe, konieczne jest stosowanie metod całkowania numerycznego (dobre wyniki dają programy NAP2 i SPICE). Przedstawiona została również metoda precyzyjnego projektowania pojemności scalonych w technologii CMOS z podwójną warstwą polikrzemu.

Słowa kluczowe: symulacja komputerowa, analogowe układy CMOS, pojemności scalone

1. WSTĘP

W rozwoju układów scalonych dąży się obecnie do minimalizacji wymiarów tranzystorów oraz do umieszczania w jednym układzie scalonym jak największej ilości elementów. W rezultacie gęstość upakowania monolitycznych systemów scalonych stała się parametrem krytycznym, powodującym wzrost gęstości mocy wydzielanej w układzie scalonym. Jednym ze sposobów zmniejszenia powierzchni takiego układu okazało się zastępowanie niektórych bloków wykonywanych dotychczas w technice cyfrowej ich odpowiednikami analogowymi. W większości systemów scalonych wielkiej skali integracji niektóre bloki, głównie systemy filtrów, wykonywane są w postaci analogowej. Wzrost zainteresowania techniką analogową oprócz wyżej wymienionych aspektów technologicznych, spowodowany został również rozwojem

nowych dziedzin elektroniki, takich jak układy z przełączanymi pojemnościami, filtry pracujące w trybie prądowym oraz sieci neuronopodobne, które już z definicji są systemami analogowymi, wykonującymi równoległe operacje arytmetyczne, w przeciwieństwie do obliczeń szeregowych realizowanych w technice logiki dwustanowej przez procesor współczesnego komputera.

Analogowe filtry scalone mogą wykonywać operacje filtrowania albo w sposób czysto analogowy (*continuous time filters*) lub też przetwarzając sygnał wejściowy na ciąg impulsów z wykorzystaniem zegarów cyfrowych (*sampled-data filters*). Najważniejszymi reprezentantami pierwszej grupy są filtry ze wzmacniaczami transkonduktancyjnymi (*Operational Transconductance Amplifier* — OTA), a wśród nich filtry których obciążeniem wzmacniaczy są pojemności (OTA-C) [11]. Najważniejszą zaletą tych filtrów jest możliwość pracy przy bardzo dużych częstotliwościach dochodzących do kilkuset megaherców. Główną wadą jest mała dokładność rzędu kilkudziesięciu procentów i wynikająca stąd konieczność stosowania układów strojenia zintegrowanych z filtrem. Ich zastosowanie ogranicza się obecnie do układów wielkoczęstotliwościowych [1], są one wykonywane np. w nadajnikach telewizyjnych.

Wśród filtrów z próbkowaniem danych na uwagę zasługują układy z przełączanymi prądami oraz z przełączanymi pojemnościami. Pierwsze z nich można wyprodukować w standardowej technologii CMOS razem z układami cyfrowymi, ponieważ przełączanie prądów polega na przemieszczaniu ładunków elektrycznych zgromadzonych na pojemnościach bramka-kanał tranzystorów MOS [5]. Dokładność tych filtrów, przy starannym zaprojektowaniu zwierciadeł prądowych, jest rzędu kilku procentów, ponadto charakteryzują się one małą wydajnością prądową, szczególnie wówczas gdy wymagany jest mały poziom zniekształceń. Filtry z przełączanymi pojemnościami (*Switched-Capacitor* — SC) [12] wymagają wykonania dodatkowych operacji w standardowej cyfrowej technologii CMOS, potrzebnych do otrzymania pojemności scalonych. Pojemności te charakteryzują się jednak bardzo dobrą liniowością, ponadto można je wykonywać tak dokładnie, że błąd wartości określającej stosunek dwóch pojemności może być mniejszy niż 0,1%. Zastosowanie układów z przełączanymi pojemnościami ograniczało się początkowo do filtrów z zegarami dwufazowymi które symulowały aktywne filtry RC zastępując rezystor kombinacją złożoną z pojemności i przełączników. Wprowadzenie zegarów wielofazowych znacznie rozszerzyło zakres zastosowań, umożliwiając również otrzymywanie precyzyjnych pamięci analogowych przechowujących informację w postaci ładunków elektrycznych zgromadzonych na pojemnościach oraz w oparciu o te pamięci projektowanie filtrów o skończonej odpowiedzi impulsowej, wykonywanych dotychczas jako układy cyfrowe [3, 12].

Przedstawione w artykule metody symulacji komputerowych poświęcone są wielofazowym układom z przełączanymi pojemnościami z uwzględnieniem realizacji filtrów o skończonej odpowiedzi impulsowej. Brak możliwości ingerencji w układ scalony wykonany w technologii CMOS powoduje konieczność przeprowadzenia szeregu symulacji komputerowych projektowanego filtru w poszczególnych etapach projektu. Daje to możliwość wyeliminowania błędów na etapie opracowywania projektu przed oddaniem go do praktycznej realizacji.

2. METODY SYMULACJI KOMPUTEROWYCH

Rozpowszechnienie produkcji układów scalonych na zamówienie ASIC (*Application Specific Integrated Circuits*) pozwala na samodzielne projektowanie dowolnych układów scalonych i wykonywanie ich za pośrednictwem europejskich fabryk krzemu. Istotnym składnikiem kosztów produkcji są opłaty za korzystanie z komputerowych stacji roboczych wyposażonych w kosztowne licencjonowane oprogramowanie. Jest zatem rzeczą bardzo pożądaną wstępne wykonanie projektu wspomagane symulacjami komputerowymi z wykorzystaniem szeroko dostępnych komputerów PC (*Personal Computer*).

W przypadku filtrów, we wstępnej fazie projektu, w celu sprawdzenia poprawności koncepcji oraz obliczeń teoretycznych wystarczy założenie o idealności wszystkich elementów (pojemności, klucze, wzmacniacze operacyjne, układ zegarowy n-fazowy o okresie równym T). Ponieważ przy przejściu z jednej fazy do drugiej następuje zmiana w położeniu kluczy, otrzymuje się N różnych liniowych układów odpowiadających N fazom układu zegarowego z warunkami początkowymi wynikającymi z ładunków zgromadzonych na pojemnościach w poprzedniej fazie. Po zastosowaniu zmodyfikowanej metody potencjałów węzłowych [2] dla dowolnej k-tej fazy, filtr SC opisany zostaje następującym równaniem rekurencyjnym:

$$\begin{bmatrix} \mathbf{Y}_k & \mathbf{B}_k \\ \mathbf{C}_k & \mathbf{D}_k \end{bmatrix} \begin{bmatrix} \mathbf{v}_{k+lN} \\ \mathbf{q}_{k+lN} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_k \mathbf{v}_{k+lN-1} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{i}_{k+lN} \\ \mathbf{e}_{k+lN} \end{bmatrix}, \tag{1}$$

gdzie N jest liczbą faz,
$\mathbf{Y}_k$, $\mathbf{B}_k$, $\mathbf{C}_k$, $\mathbf{D}_k$ — macierze opisujące układ w fazie k-tej zgodnie z regułami metody potencjałów węzłowych.
$\mathbf{v}$, $\mathbf{q}$ — wektory potencjałów węzłowych oraz ładunków na pojemnościach,
$\mathbf{i}$, $\mathbf{e}$ — wektory prądowych oraz napięciowych źródeł niezależnych,
l — indeks oznaczający kolejny numer okresu T.

Równanie (1) zostało wykorzystane do napisania prostego programu komputerowego dla komputerów typu IBM PC do analizy wielofazowych sieci SC w dziedzinie czasu. Analiza ta pozwala znaleźć błędy w pierwszym etapie procesu projektowania, np. niewłaściwy rozkład potencjałów w węzłach prowadzący do zwarcia źródła napięciowego lub rozwarcie obwodu sprzężenia zwrotnego wzmacniacza operacyjnego w dowolnej fazie układu zegarowego.

Zastosowanie przekształcenia Z do równania (1) umożliwia budowę programu do obliczania charakterystyk częstotliwościowych [12], szczególnie ważnych w przypadku analizy filtrów. Jeżeli weryfikacja projektu z zastosowaniem powyższych programów wypada pomyślnie, można przejść do następnego etapu, w którym uwzględniony zostanie wpływ elementów nieidealnych na charakterystyki wyjściowe filtru. Dobrymi narzędziami do wykonania precyzyjnej i zaawansowanej analizy w dziedzinie czasu są uniwersalne programy analizy układów elektronicznych, np. NAP-2 [10] lub PSPICE [8, 13].

```
*CIRCUIT
:ANALOGOWY INTERPOLATOR LINIOWY
:
:
:
F1/TAB2/P 4E-6 0 2E3 1E-6 2E3 1E-6 1E8 4E-6 1E8 4E-6 2E3
F2/TAB2/P 4E-6 0 1E8 1E-6 1E8 1E-6 2E3 2E-6 2E3 2E-6 1E8 4E-6 1E8
F4/TAB2/P 4E-6 0 1E8 3E-6 1E8 3E-6 2E3 4E-6 2E3 4E-6 1E8
F13/TAB2/P 4E-6 0 2E3 1E-6 2E3 1E-6 1E8 2E-6 1E8 2E-6 2E3 3E-6 2E3>
3E-6 1E8 4E-6 1E8
F234/TAB2/P 4E-6 0 1E8 1E-6 1E8 1E-6 2E3 4E-6 2E3 4E-6 1E8
F134/TAB2/P 4E-6 0 2E3 1-6 2E3 1E-6 1E8 2E-6 1E8 2E-6 2E3 4E-6 2E3
R11 1 5 1*F1(TIME)
R12 3 0 1*F1(TIME)
R13 6 0 1*F1(TIME)
R21 4 6 1*F2(TIME)
R22 4 7 1*F2(TIME)
R23 2 8 1*F2(TIME)
R24 9 0 1*F2(TIME)
R131 2 4 1*F13(TIME)
R41 3 4 1*F4(TIME)
R42 2 9 1*F4(TIME)
R234 5 0 1*F234(TIME)
R1341 7 0 1*F134(TIME)
R1342 8 0 1*F134(TIME)
:
:
:
```

Rys. 1. Opis rezystorów nieliniowych zmiennych w czasie w formacie programu NAP-2

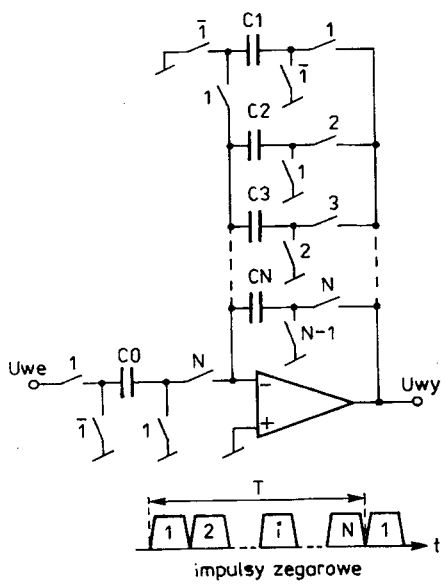
W programie NAP-2 można bardzo łatwo symulować elementy nieliniowe i dlatego prostym modelem przełącznika w tym programie jest nieliniowa rezystancja. Rysunek 1 pokazuje w jaki sposób należy opisać przełącznik jako nieliniowy rezystor z wykorzystaniem funkcji TAB2 (funkcja TAB2 wbudowana w program NAP-2, pozwala na opis tabelaryczny przez podanie współrzędnych argumentu i funkcji, przy czym jeżeli funkcja jest okresowa, wystarczy podać opis jednego okresu oraz wartość tego okresu [10]). Rezystancje modelujące przełączniki zmieniają swoje wartości w funkcji czasu od 2 kΩ przy przełączniku zamkniętym (wartość typowa dla technologii 2 μm) do 100 MΩ przy przełączniku otwartym. Po czasie T równym okresowi układu zegarowego następuje powtórzenie sekwencji sterującej tymi kluczami.

Pakiet programów SPICE pozwala na przeprowadzenie dokładniejszych symulacji i jest bardzo przydatny w przypadku wielu specyficznych układów z przełączanymi pojemnościami, jak wielofazowe wąskopasmowe filtry N-gałęziowe, filtry nierekursywne lub przetworniki analogowo-cyfrowe i cyfrowo-analogowe. Duża liczba parametrów znajdujących się pod kontrolą użytkownika, wielopoziomowego modelu wbudowanego tranzystora MOS, umożliwia symulację całego filtru na poziomie tranzystorów MOS. W przypadku rozbudowanych układów wielofazowych zbyt

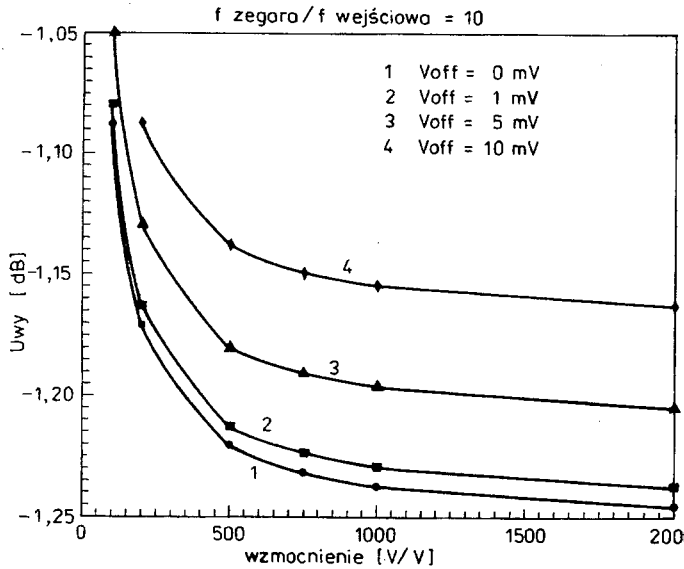
długi czas obliczeń lub brak dostatecznie dużej pamięci komputera mogą ograniczyć symulację do wybranych istotnych fragmentów całego układu, np. do analizy właściwości dynamicznych wzmacniacza operacyjnego. Istnieje również możliwość opisu części układu na poziomie pojedynczych tranzystorów (np. wzmacniacze operacyjne) a pozostałej części na poziomie funkcjonalnym (np. przełączniki jako rezystory nieliniowe). Ostateczna weryfikacja przed wykonaniem układu scalonego polega na automatycznej generacji zbioru danych bezpośrednio z wykonanego projektu układu scalonego na poziomie masek, z wykorzystaniem parametrów tranzystorów dostarczanych dla danej technologii przez fabrykę krzemu.

3. LINIA OPÓŹNIAJĄCA

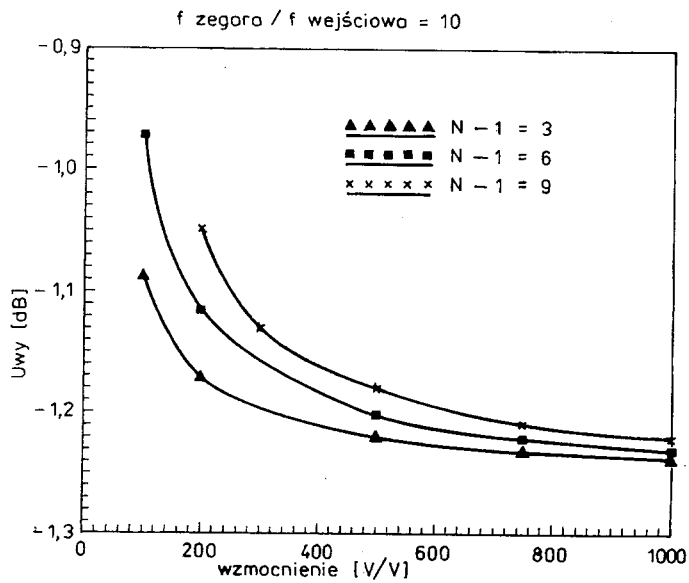
Przy realizacji filtrów o skończonej odpowiedzi impulsowej wykonywane są trzy podstawowe operacje: mnożenie, dodawanie oraz opóźnienie czasowe. W przypadku wielofazowych układów SC mnożenie realizowane jest przez właściwy dobór wartości pojemności, natomiast sumowanie polega na przesyłaniu ładunków między pojemnościami. Najtrudniejszym problemem jest realizacja linii opóźniającej, ponieważ należy minimalizować wpływ nieidealnych charakterystyk wzmacniacza operacyjnego oraz wpływ pojemności pasożytniczych. Podstawowy schemat analogowej linii opóźniającej przedstawiony jest na rysunku 2. N-fazowy układ zegarowy pozwala zrealizować (N-1) jednostkowych opóźnień. W fazie 1 pojemność C_0 jest ładowana do napięcia równego U_{we}, natomiast w fazie N-tej ładunek z pojemności C_0 jest przeniesiony do pojemności C_N włączonej w obwód sprzężenia zwrotnego wzmac-



Rys. 2. Analogowa linia opóźniająca



Rys. 3. Napięcie wyjściowe linii opóźniającej w funkcji wzmocnienia wzmacniacza dla różnych wartości napięcia niezrównoważenia

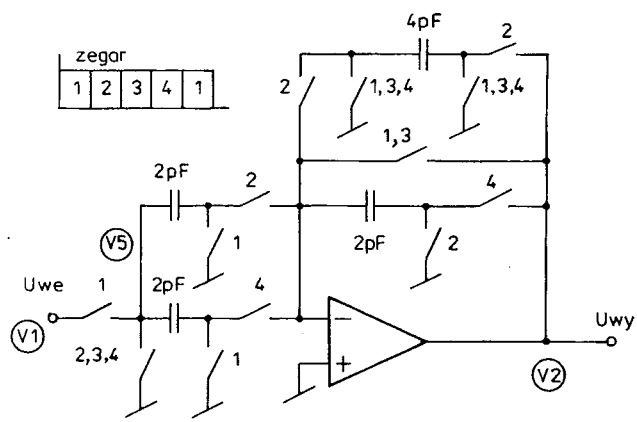


Rys. 4. Napięcie wyjściowe dla różnych długości linii opóźniającej

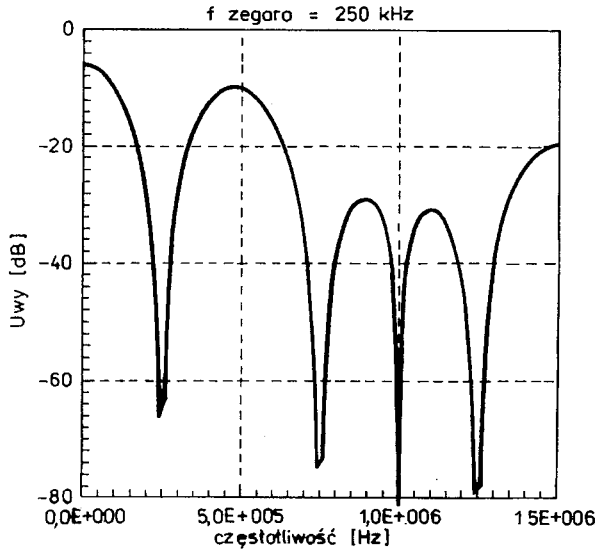
niacza. W nastepnych cyklach zegarowych ładunek ten jest przesuwany do nastepnych pojemności od C_{N-1} aż do C_1. Dokładność transferu ładunku zależy od skończonego wzmocnienia wzmacniacza operacyjnego jak również od napiecia niezrównoważenia (*off-set voltage*) tego wzmacniacza i jest funkcją liczby faz układu zegarowego. Wpływ powyższych czynników na charakterystykę amplitudową linii opóźniającej wyznaczony został przez wykonanie symulacji komputerowych programem do analizy częstotliwościowej układów SC i pokazany jest na rysunkach 3 i 4. Błędy spowodowane nieidealnymi parametrami linii opóźniającej wzrastają wraz z jej długością oraz ze wzrostem częstotliwości (maleje wzmocnienie wzmacniacza operacyjnego). Może się wówczas okazać konieczne zastosowanie układów kompensujących napięcie niezrównoważenia [12] oraz podział jednej linii na kilka mniejszych z jednoczesnym zmniejszeniem liczby cykli układu zegarowego [7].

4. INTERPOLATOR LINIOWY

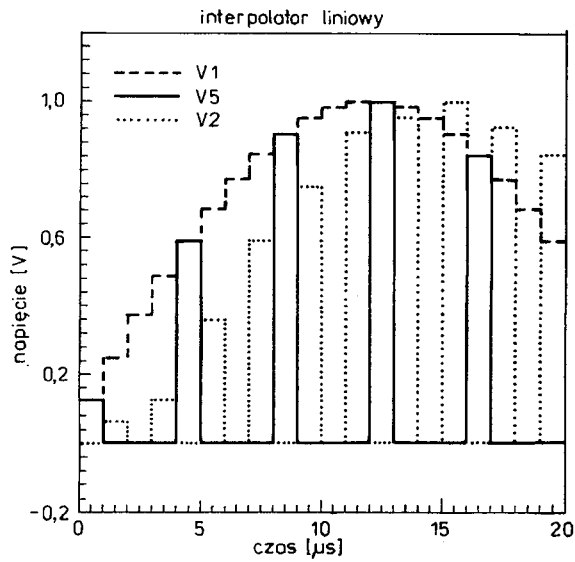
Przykładowy wielofazowy interpolator [3] przedstawiony został na rysunku 5. Pomiędzy dwa kolejne impulsy wejściowe wstawiany jest jeden dodatkowy impuls o amplitudzie równej ich średniej arytmetycznej. Impulsy wyjściowe pojawiają się zatem z częstotliwością dwukrotnie większą w porównaniu z częstotliwością wejściową. Charakterystyka amplitudowa tego interpolatora pokazana została na rysunku 6. Do sprawdzenia poprawności działania interpolatora przyjęto wszystkie elementy jako idealne. Wyniki symulacji komputerowych w dziedzinie czasu przedstawione są na rysunku 7. Dla obliczenia amplitudy interpolowanego impulsu układ musi zapamiętać dwa kolejne impulsy wejściowe i w związku z tym sygnał wyjściowy (V2) jest opóźniony w stosunku do ciągu impulsów na wejściu (V5). Wyniki symulacji tego samego interpolatora programem NAP2 z uwzględnieniem niezerowej rezystancji załączania przełączników równej 2 kΩ oraz skończonej rezystancji wyłączania (100 MΩ), pokazane są na rysunku 8.



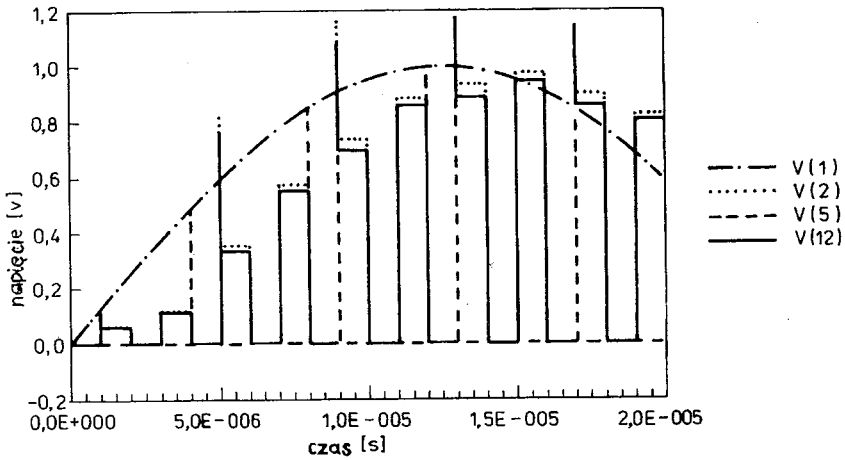
Rys. 5. Wielofazowy interpolator liniowy



Rys. 6. Charakterystyka amplitudowa interpolatora

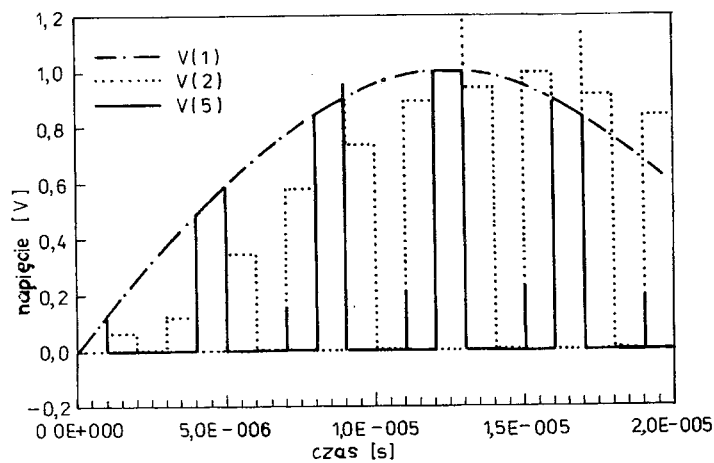


Rys. 7. Charakterystyka czasowa interpolatora — elementy idealne



Rys. 8. Analiza interpolatora w dziedzinie czasu programem NAP-2

Wzmacniacz modelowany był jako źródło napięciowe sterowane napięciem o współczynniku wzmocnienia 5000 V/V (krzywa V2) oraz 50 V/V (krzywa V12). Przesadnie mały współczynnik wzmocnienia równy 50 V/V przyjęty został w drugim przypadku dla określenia tendencji zmian napięcia wyjściowego. Dla dokładniejszego wyznaczenia przebiegów czasowych wykonane zostały symulacje komputerowe programem PSPICE, przy czym wzmacniacz modelowany był na poziomie pojedynczych tranzystorów, tzn. wprowadzone zostały modele wszystkich tranzystorów MOS wzmacniacza, natomiast w modelach przełączników uwzględniono rezystancję załączenia równą 2 kΩ oraz przy wyłączeniu przyjęto wartość 100 MΩ. Wyniki symulacji przedstawione są na rysunku 9.



Rys. 9. Analiza interpolatora w dziedzinie czasu programem PSPICE

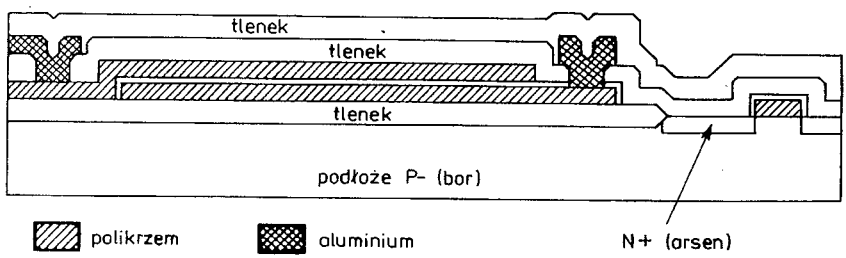
Prosty przykład interpolatora pokazuje różnice, które można zauważyć przy zastosowaniu odmiennej metodyki modelowania do symulacji tego samego układu. Modelowanie zakładające idealność wszystkich elementów (rysunek 7) może być przeprowadzone w krótkim czasie na małych komputerach, jest ono zatem polecane we wstępnej fazie projektu dla sprawdzenia poprawności przyjętej koncepcji i dokonania niezbędnych poprawek.

5. POJEMNOŚCI SCALONE

Pojemności wykonywane w technologii CMOS znalazły w ostatnich latach szerokie zastosowanie w takich układach jak filtry z przełączanymi pojemnościami, filtry ze wzmacniaczami transkondukcyjnymi, przetworniki analogowo-cyfrowe, modulatory sigma-delta, pamięci analogowe [1, 4, 11, 12]. Są one wykonywane najczęściej w technologii CMOS z podwójną warstwą polikrzemu. Dwie warstwy krzemu krystalicznego oddzielone są cienką warstwą dwutlenku krzemu, który wytwarzany jest jednocześnie z warstwami izolacyjnymi bramek tranzystorów MOS. Przekrój takiej pojemności pokazany jest na rysunku 10. Jej wartość opisuje następujące równanie [2]:

$$C = K_0 A + K_1 P_1 + K_2 P_2$$

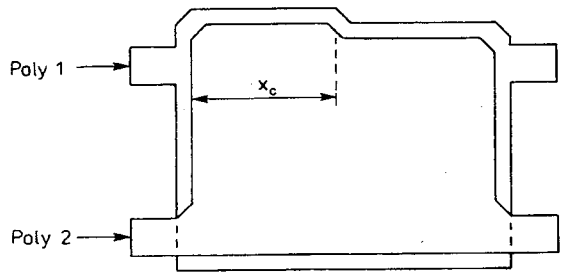
gdzie: A jest powierzchnią zajmowaną przez pojemność, P_1 jest długością tej części krawędzi dolnej warstwy polikrzemu (poly 1), która leży pod górną warstwą polikrzemu (poly 2), P_2 jest długością tej części krawędzi poly 2 która leży nad warstwą poly 1. Współczynniki K_0, K_1 i K_2 są parametrami technologicznymi, których wartości mogą się zmieniać w granicach $\pm 20\%$, jednakże na powierzchni pojedynczej płytki krzemowej każdy z tych parametrów ma wartość stałą.



Rys. 10. Przekrój pojemności scalonej w technologii CMOS z podwójną warstwą polikrzemu

Właśnie dlatego, w ramach pojedynczego układu scalonego można zachować bardzo dużą precyzję wykonania poszczególnych pojemności. Największą dokładność uzyskuje się minimalizując wpływ efektów brzegowych przez zachowanie tego samego stosunku A/P_1 oraz A/P_2. W tym celu każda pojemność składa się z szeregu

połączonych równolegle identycznych pojemności jednostkowych oraz jednej dodatkowej nie mniejszej niż pojemność jednostkowa i nie większej niż dwie takie pojemności. Staranne wykonanie tej ostatniej decyduje o precyzji całego projektu. Ponieważ dokładność rysowania krawędzi jest w praktyce ograniczona skokową rozdzielczością charakterystyczną dla danego programu, dlatego w praktyce przyjmuje się kształt zbliżony do prostokąta, jak na rysunku 11 z możliwością precyzyjnego wyznaczenia powierzchni poprzez zmianę długości x_c.



Rys. 11. Typowy kształt pojemności scalonej

Z przeprowadzonych rozważań wynika konieczność stosowania programów komputerowych, pozwalających wyznaczać w sposób optymalny wartości i kontury wszystkich pojemności tak, aby minimalizując wpływ efektów brzegowych zapewnić maksymalną dokładność ich wykonania.

WNIOSKI

Przedstawione przykłady symulacji filtrów z przełączanymi pojemnościami odzwierciedlają ogólną tendencję stosowania komputerów przy projektowaniu analogowych układów scalonych. W przypadku filtrów pracujących przy małych częstotliwościach symulacje w dziedzinie częstotliwości często wystarczają do ich poprawnego wykonania. Tym niemniej, analiza w dziedzinie czasu, na poziomie prostych funkcjonalnych modeli elementów półprzewodnikowych, pomaga często uniknąć błędów we wstępnej fazie projektowania.

Filtry analogowe pracujące przy dużych częstotliwościach rzędu megaherców wymagają zastosowania dokładniejszych symulacji, szczególnie przy nowoczesnych submikronowych technologiach CMOS, gdzie występuje już konieczność uwzględnienia wpływu efektów krótkiego kanału na charakterystyki tranzystora.

Niewłaściwy dobór pakietu programów komputerowych może zatem albo prowadzić do straty czasu przy wykonywaniu czasochłonnych symulacji uwzględniających parametry modeli nieistotne dla danego filtru, lub też brak niezbędnych symulacji komputerowych może zmniejszyć dokładność całego układu.

BIBLIOGRAFIA

1. K. Azadet: *Linear phase continuous-time video filters based on a mixed Analog/Digital structure.* XI European Conference on Circuit Theory and Design. Davos (Switzerland), September 1993, p. 73
2. Z. Ciota, A. Napieralski, J.L. Noullet: *Simulation methods and design of IC switched-capacitor networks.* Third International Workshop Power and Timing Modelling, Optimization and Simulation. La Grande Motte near Montpellier (France), October 1993, p. 237
3. Z. Ciota, A. Napieralski, J.L. Noullet: *Analog interpolated finite impulse response filters.* XI European Conference on Circuit Theory and Design. Davos (Switzerland), September 1993, p. 1367
4. Z. Ciota, A. Napieralski, J.L. Noullet: *Design process of integrated SC circuits.* Advanced Training Course: Mixed Design of VLSI Circuits, Dębe (Poland), April 1994, p. 118
5. T.S. Fiez, G. Liang, D.J. Allstot: *Switched-current circuit design issues.* IEEE J. Solid-State Circuits 1991, vol. 26, nr 3, p. 192
6. M.D. Godfrey: *CMOS device modelling for subthreshold circuits.* IEEE Trans. Circuits and Systems — II 1992, vol. 39, nr 8, p. 532
7. A. Napieralski, J.L. Noullet, Z. Ciota: *Realization of some different FIR SC filters in the CMOS technology.* Bulletin of the Polish Academy of Sciences, Technical Sciences 1994, vol. 42, nr 2, p. 269
8. *PSPICE version 5.2.* MicroSim Corporation, 1992
9. D. Radhakrishnan: *Design of CMOS circuits.* IEE Proceedings — G 1991, vol. 138, nr 1, p. 83
10. T. Rubner-Peterson: *NAP2 a nonlinear analysis program for electronic circuits.* Report 16/5-73, Technical University of Denmark, 1973
11. Y. Sun, J.K. Fidler: *Canonical realization of high-order all-pole low-pass OTA-C filters.* XI European Conference on Circuit Theory and Design. Davos (Switzerland), September 1993, p. 69
12. R. Unbehauen, A. Cichocki: *MOS switched-capacitor and continuous-time integrated circuits and systems.* Springer-Verlag, Berlin 1989
13. W.W. Wong, R.S. Witon, J.J. Liou: *Statistical and numerical method for MOSFET integrated-circuit sensitivity simulation using SPICE.* IEE Proceedings-G, 1991, vol. 138, nr 1, p. 177

Z. CIOTA

COMPUTER SIMULATION METHODS FOR ANALOG INTEGRATED FILTERS IN CMOS TECHNOLOGY

Summary

The most important methods of computer simulations in the design process of multiphase switched-capacitor filters are presented. The comparison of different simulations using different computer programs has been performed. If all components of the network are ideal, the modified nodal approach is the most suitable method for time-domain and frequency analysis of such circuit. To perform more precise time-domain analysis by using nonideal switches including non-zero switch-on resistance and finite switch-off resistance, when non-zero RC time constants appears, numerical integration methods must be applied (e.g. NAP2 and SPICE programs). The precise design method of integrated capacitors on double poly CMOS process has been also discussed.

Key words: computer simulation, analog CMOS circuits, integrated capacitors.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 3, ss. 361–378

Projektowanie systemów wieloprocesorowych z procesorami sygnałowymi

BOGUSŁAW WIĘCEK, ANDRZEJ SZKODZIK

Instytut Elektroniki, Politechnika Łódzka

Otrzymano 1995.02.15

Autoryzowano do druku 1995.06.26

W pracy przedstawiono przegląd stało i zmiennoprzecinkowych procesórów sygnałowych na przykładzie rodziny TMS320, podkreślając możliwość ich zastosowania w strukturach wieloprocesorowych. Zaproponowano ocenę niezrównoważenia obciążeń procesorów w układzie wieloprocesorowym stosując statystyczny model obrazu. Wyniki symulacji potwierdziły możliwość uzyskania optymalnego podziału zadań w strukturze dwuprocesorowej zrealizowanej do przetwarzania obrazów.

Słowa kluczowe: Procesory sygnałowe, systemy wieloprocesorowe, pamięć wspólna, analiza obciążeń procesorów.

1. WPROWADZENIE

Rozwój metod przetwarzania sygnałów wywołał rozwój układów elektronicznych, w tym głównie cyfrowych i mikroprocesorowych. Procesory sygnałowe — DSP (ang. Digital Signal Processors) są przykładem, jak za rozwojem teorii nadąża technologia wytwarzania układów scalonych. Procesory DSP są jednym z kierunków rozwoju układów scalonych i współistnieją z transputerami i procesorami typu RISC (ang. *Reduced Instruction Set Computer*). Wszystkie z nich mają zdolność wykonywania milionów operacji na sekundę i są zalecane przy realizacji arytmetycznie złożonych algorytmów przetwarzania [5].

Procesory sygnałowe posiadają jedną niewątpliwą zaletę — są procesorami jednoukładowymi i mogą tworzyć złożone systemy przy minimalnym nakładzie na elementy towarzyszące: pamięć, sterowanie, układy we/wy. Choć procesory sygnałowe mają dużą moc obliczeniową, ich słabą stroną była w pierwszych rozwiązaniach ograniczona możliwość komunikacji z otoczeniem. Współczesne DSP są przystosowane do współpracy równoległej, co znakomicie zwiększa szybkość

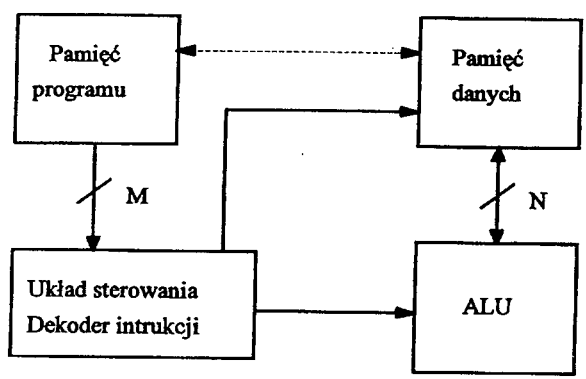
wymiany informacji w systemie. Wydaje się, że DSP są preferowane w rozwiazaniach specjalizowanych, zwartych w sensie struktury (ang. *embeded*) i opracowywanych w krótkim czasie (kilka miesięcy).

Istnieje wiele grup procesorów sygnałowych, z których najbardziej szeroką, pokrywającą wielorakie zastosowania, w tym przede wszystkim umożliwiającą pracę równoległą jest rodzina procesorów TMS 320Cxx. W pracy scharakteryzowano podstawowe procesory tej rodziny podając parametry przetwarzania danych (np.: szybkość wymiany informacji). Ponadto przedstawiono projekt wieloprocesorowego systemu graficznego wraz ze statystyczną analizą nierównomierności obciążeń procesorów pracujących równolegle.

2. PROCESORY STAŁOPRZECINKOWE

Rodzinę cyfrowych procesorów sygnałowych TMS320 zapoczątkował opracowany w 1982 r. procesor TMS32010 wytworzony w technologii NMOS 2.4 μm. Był on pierwszym mikrokomputerem jednoukładowym zdolnym do wykonywania 5 milionów rozkazów na sekundę (5 MIPS — ang. *Million Instructions Per Second*). Jego odpowiednikiem wykonanym w technologii CMOS 1.8 μm był procesor TMS320C10.

Procesory TMS32010/C10 dla uzyskania dużej szybkości pracy wykorzystują zmodyfikowaną architekturę harwardzką (Rys. 1). Architekturę harwardzką charakteryzuje podział pamięci na pamięć programu i pamięć danych, które znajdują się w dwóch oddzielnych przestrzeniach adresowych. To rozwiązanie pozwala na jednoczesne pobieranie i wykonywanie rozkazów. W praktyce długość słowa programu jest inna niż danych. Zmodyfikowana architektura harwardzka umożliwia dodatkowo wykonywanie transferów między przestrzenią programu i danych. W następnych generacjach procesorów rodziny TMS320 obie przestrzenie zostały ujednolicone tzn. procesor może pracować również jak typowa struktura von Neumanna [5]. Procesor TMS32010/C10 wyposażony jest w wewnętrzną pamięć programu ROM (1.5 K × 16 bitów) oraz pamięć danych RAM (144 × 16 bitów). Pamięć programu może być zwiększona do 4 K przez dołączenie pamięci zewnętrznej.



Rys. 1. Architektura harwardzka

W procesorze występują 4 podstawowe elementy arytmetyczne: ALU, akumulator, układ mnożący i układy przesuwające. Wszystkie działania arytmetyczne wykonywane są przy użyciu arytmetyki uzupełnień do 2. ALU jest jednostką arytmetyczno-logiczną ogólnego przeznaczenia pracującą z 32-bitowym słowem danych. Może dodawać, odejmować i wykonywać operacje logiczne. Wynik operacji ALU przechowywany jest w 32-bitowym akumulatorze. Odpowiednie rozkazy procesora pozwalają na zapamiętywanie w pamięci odpowiednio starszej i młodszej części akumulatora. Równoległy układ mnożący wykonuje w jednym cyklu operację mnożenia na danych 16-bitowych. Wynik operacji przechowywany jest w specjalnym 32-bitowym rejestrze P. Mnożenie i operację na akumulatorze można zrealizować w 2 kolejnych cyklach rozkazowych. W procesorze dostępne są 2 układy przesuwające. Cykliczny układ przesuwający może wykonywać przesuwanie w lewo o 0 – 16 bitów na słowach pamięci danych, które są następnie dodawane lub odejmowane od akumulatora. Równoległy układ przesuwający wykonuje przesunięcia o 0, 1, 4 bity danej znajdującej się w akumulatorze, w celu poprawnego generowania bitu znaku w obliczeniach w kodzie uzupełnień do 2.

16-bitowa równoległa magistrala danych może być wykorzystana do wykonywania operacji we/wy z szybkością 2.5 mln słów na sekundę. Czas cyklu rozkazowego podstawowej wersji TMS32010 wynosi 200 ns, ale dostępne są również procesory o cyklu 160 ns (TMS32010-25).

Kolejna wersja TMS320C15 oferuje rozszerzoną do 256 słów wewnętrzną pamięć RAM oraz wewnętrzną pamięć programu ROM o wielkości 4 K słów. Procesor ten posiada swój odpowiednik TMS320E15 zawierający zamiast pamięci ROM pamięć EPROM. TMS320C15 dostępny jest w wersji 200 ns lub 160 ns (TMS320C15-25). Procesory te są w pełni kompatybilne pod względem kodu wynikowego i końcówek zewnętrznych z TMS32010.

TMS320C17 jest specjalizowanym mikrokomputerem zawierajacym, oprócz wewnętrznej pamięci programu ROM (lub EPROM w wersji TMS320E17) o pojemności 4 K słów, także dwukanałowy port szeregowy, układ czasowy portu szeregowego do samodzielnej komunikacji oraz interfejs koprocesora. Jest zgodny z procesorem TMS32010 na poziomie kodu wynikowego.

Pierwszym procesorem należącym do drugiej generacji rodziny TMS320 był, wykonany w technologii NMOS 2.4 μm, układ TMS32020. Najważniejszym jednak przedstawicielem tej generacji jest jego udoskonalona i rozszerzona wersja wykonana w technologii CMOS 1.8 μm — procesor TMS320C25. Zestaw rozkazów tego procesora jest nadzbiorem rozkazów TMS32010 i TMS32020, przy czym zachowana jest zgodność kodu źródłowego. Ponadto jest on całkowicie kompatybilny z procesorem TMS32020 na poziomie kodu wynikowego, tak że programy dla TMS32020 mogą być bez modyfikacji uruchamiane na TMS320C25.

Cykl rozkazowy TMS320C25 został skrócony do 100 ns co, przy wykonywaniu większości rozkazów w jednym cyklu, daje możliwość wykonywania 10 milionów rozkazów na sekundę (10 MIPS). Przestrzeń adresowa procesora obejmuje 64 K pamięci programu oraz 64 K pamięci danych. Wewnątrz układu znajduje się pamięć programu ROM o pojemności 4 K × 16 bitów oraz pamięć danych RAM

o pojemnosci 544 słów 16-bitowych podzielona na 3 oddzielne bloki. Jeden z tych bloków o wielkosci 256 słów może być skonfigurowany albo jako pamięć danych, albo dodatkowa pamięć programu. Wielkość wewnętrznej pamięci danych pozwala na obsługę tablic zawierających 512 słów, przy pozostawieniu 32 komórek na tymczasowe przechowywanie danych.

Centralna jednostka arytmetyczno-logiczna (CALU — ang. Central Arithmetic Logic Unit) obejmuje 16-bitowy skalujący układ przesuwający, równoległy układ mnożący 16 × 16-bitów, 32-bitową jednostkę arytmetyczno-logiczną (ALU) i 32-bitowy akumulator. Skalujący układ przesuwajacy posiada 16-bitowe wejście połączone z magistralą danych i 32-bitowe wyjście połączone z ALU. Wykonuje on przesuwanie w lewo o 0—16 bitów na danej wejściowej. Dodatkowe układy przesuwające na wyjściach akumulatora i układu mnożącego mogą być wykorzystane do implementacji arytmetyki o zwiększonej precyzji, zapobiegania powstawaniu nadmiaru, numrycznego skalowania oraz wydzielania i maskowania bitów. 32-bitowy akumulator podzielony jest na dwie 16-bitowe części dla umożliwienia przechowywania jego zawartości w pamięci danych.

Układ mnożący 16 × 16 bitów może obliczyć 32-bitowy iloczyn w czasie każdego cyklu maszynowego. Z układem mnożącym związane są 2 rejestry: 16-bitowy rejestr tymczasowy TR (ang. *Temporary Register*) zawierający jeden z argumentów operacji mnożenia oraz 32-bitowy rejestr wynikowy PR (ang. *Product Register*) przechowujący wynik mnożenia. Zawartość rejestru wynikowego może być przesuwana w lewo o 1 lub 4 bity. Jest to użyteczne przy implementacji arytmetyki ułamkowej. Zawartość PR może być również przesuwana w prawo o 6 bitów w celu umożliwienia wykonywania do 128 kolejnych mnożeń i operacji na akumulatorze bez wystąpienia nadmiaru. Przestrzeń we-wy procesora TMS320C25 obejmuje 16 portów wejściowych i 16 wyjściowych. Operacje we-wy wykonywane są poprzez magistralę danych procesora i trwają 2 cykle maszynowe. Wykorzystanie specjalnego licznika powtórzeń (ang. Repeat Counter) powoduje skrócenie tego czasu do 1 cyklu. Procesor zawiera także wewnętrzny układ bezpośredniego dostępu do pamięci DMA (ang. *Direct Memory Access*) mający dostęp zarówno do pamięci programu jak i danych. Dostępne są 3 główne tryby adresowania: bezpośredni, pośredni i natychmiastowy. Adresowanie bezpośrednie i pośrednie używane jest w celu dostępu do pamięci danych. Adresowanie natychmiastowe pozwala na dostęp do pamięci programu. Przy użyciu adresowania bezpośredniego, 7 bitów słowa rozkazu łączonych jest z 9 bitami wskaźnika strony pamięci danych DP w celu utworzenia 16-bitowego adresu. Przy długości strony 128 słów rejestr DP wskazuje na jedną z 512 możliwych stron pamięci danych. Adresowanie pośrednie wykorzystuje 8 rejestrów pomocniczych (AR0—AR7). Wśród trybów adresowania pośredniego występuje tryb adresowania z odwróceniem bitów (ang. *Bit-reversed Addressing*) przyspieszający wykonywanie operacji we-wy przy ponownym ustalaniu kolejności wskaźników danych w algorytmie FFT.

Powstało także kilka kolejnych procesorów będących modyfikacjami TMS320C25. Należy do nich procesor TMS320E25, w którym wewnętrzna pamięć ROM została zastąpiona przez EPROM o tej samej pojemności 4 K słów. Procesor TMS320C25-50 może pracować przy częstotliwości zegara do 50 MHz. Kolejny

procesor wywodzący się od TMS320C25 to TMS320C26, w którym wewnętrzna pamięć ROM została zamieniona na RAM. Posiada on zatem 1.5 K słów wewnętrznej pamięci RAM oraz 256 słów pamięci ROM i może pracować przy minimalnej pamięci zewnętrznej [5].

Kontynuatorem stałoprzecinkowym procesorów pierwszej i drugiej generacji jest procesor TMS320C50. W procesorze tym czas cyklu rozkazowego skrócony został do 50 ns. Pomimo dodania nowych rozkazów ogólnego przeznaczenia i rozkazów DSP zachowana została zgodność z procesorami TMS320C1x i TMS320C2x na poziomie kodu źródłowego. Wewnętrzna pamięć RAM procesora powiększona została do 8.5 K × 16 bitów przy całkowitej przestrzeni adresowej równej 128 K słów. Procesor zawiera również 2 K × 16 bitów wewnętrznej pamięci ROM. Układy peryferyjne są analogiczne jak w przypadku procesora TMS320C25, przy czym szybkość pracy tych układów jest odpowiednio większa na skutek skrócenia cyklu rozkazowego.

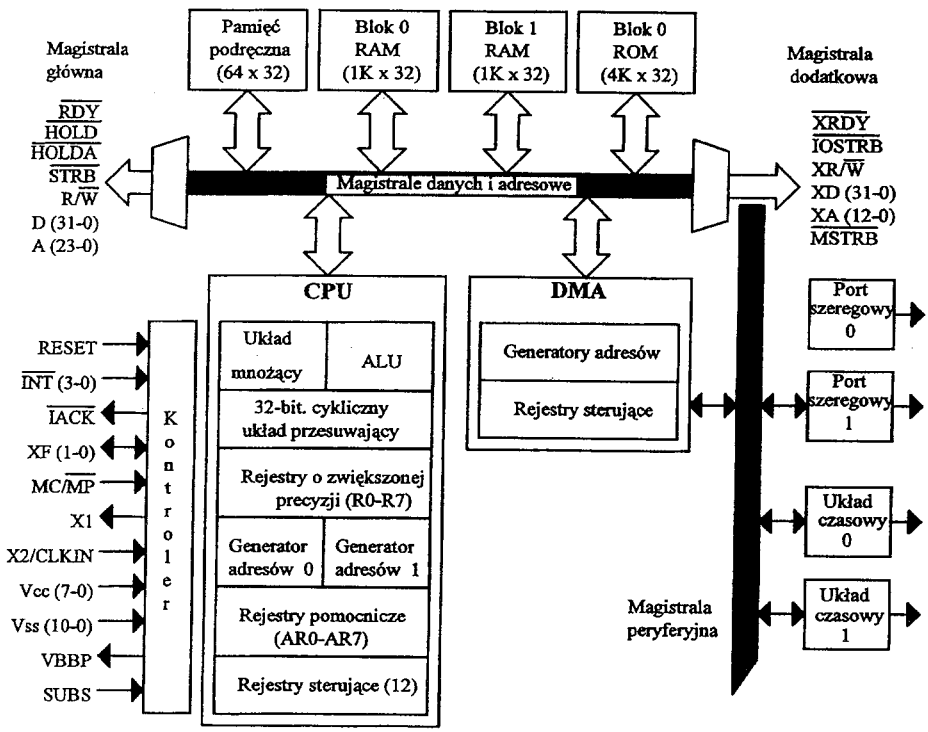
3. PROCESORY ZMIENNOPRZECINKOWE

Pierwszym procesorem rodziny TMS320 wykorzystującym arytmetykę zmiennoprzecinkową jest procesor TMS320C30. Ujednolicono przestrzeń programu i danych, zwiększono przestrzeń adresową do 16 milionów słów 32-bitowych, skrócono cykl rozkazowy do 60 ns, wykonano 24 Kbajty pamięc wewnętrznej oraz kilka urządzeń peryferyjnych. TMS320C30 daje użytkownikowi moc obliczeniową jaka w niezbyt odległym okresie była domeną wyłącznie superkomputerów. Może on wykonywać 33 miliony operacji zmiennoprzecinkowych na sekundę (33MFLOPS — ang. *Million Floating-point Operations per Second*). Aby uzyskać taką szybkość przy niskim koszcie procesor projektowany był w technologii CMOS 1 μm.

Procesor TMS320C30 posiada 2 zewnętrzne interfejsy: magistralę główną (ang. *Primary Bus*) i magistralę dodatkową (ang. *Expansion Bus*). Magistrala główna składa się z 32-bitowej magistrali danych, 24-bitowej magistrali adresowej i zestawu sygnałow sterujących. Magistrala dodatkowa obejmuje 32-bitową magistralę danych, 13-bitową magistralę adresową i odpowiednie sygnały sterujące. Za pośrednictwem magistrali głównej procesor przy pracy z maksymalną szybkością (bez oczekiwania na urządzenia zewnętrzne) może wykonywać odczyty w każdym cyklu zegara, natomiast zapisy co drugi cykl. W przypadku magistrali dodatkowej zarówno opis, jak i odczyt trwają dwa cykle zegara. Oba interfejsy dysponują zewnętrznym wejściem gotowości do generowania stanów oczekiwania w przypadku dołączenia wolnych urządzeń zewnętrznych oraz mogą wykorzystywać stany oczekiwania generowane programowo.

Jednostka centralna TMS320C30 obejmuje układ mnożący, jednostkę arytmetyczno-logiczną (ALU) do wykonywania operacji na danych zmiennoprzecinkowych i całkowitych oraz operacji logicznych, 2 jednostki arytmetyczne rejestrów pomocniczych ARAU (ang. *Auxiliary Register Arithmetic Unit*), plik rejestrów wspomagających i odpowiednie magistrale. Układ mnożący wykonuje mnożenie na 32-bitowych

liczbach zmiennoprzecinkowych, dając w wyniku liczbę 40-bitowa oraz na 24-bitowych liczbach całkowitych, dając w wyniku liczbę 32-bitową. ALU wykonuje operacje na 32-bitowych danych całkowitych i logicznych oraz 40-bitowych danych zmiennoprzecinkowych. Wyniki z układu mnożącego i ALU są zawsze utrzymywane w formacie 32-bitowym dla danych całkowitych lub 40-bitowym dla danych zmiennoprzecinkowych. TMS320C30 posiada zdolność wykonywania w jednym cyklu równoległych mnożeń i dodawań (odejmowań) na danych całkowitych lub zmiennoprzecinkowych.



Rys. 2. Schemat blokowy procesora TMS320C30

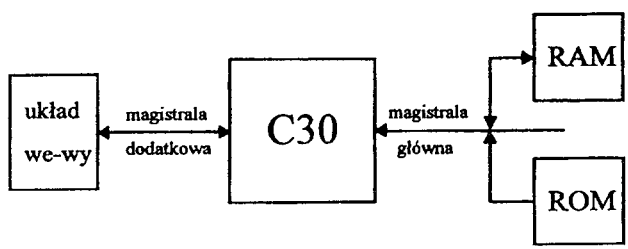
Dwie jednostki arytmetyczne rejestrów pomocniczych (ARAU0 i ARAU1) mogą w jednym cyklu generować dwa różne adresy. Pracują one równolegle z układem mnożącym i ALU. Odgrywają znaczącą rolę przy adresowaniu z przemieszczeniem, z rejestrami indeksowymi (IR0 i IR1), adresowaniu z odwróceniem bitów i adresowaniu cyklicznym (ang. *Circular Addressing*) [6]. Plik rejestrów (ang. *Register File*) składa się z 28 rejestrów, na których można wykonywać opracje za pośrednictwem układu mnożącego lub ALU. Rejestry o zwiększonej precyzji R0–R7 (ang. *Extended-Precision Registers*) umożliwiają wykonywanie operacji na 32-bitowych liczbach całkowitych i 40-bitowych liczbach zmiennoprzecinkowych. Rejestry pomocnicze AR0–AR7 (ang. *Auxiliary Registers*) służą do generowania adresów przy adreso-

waniu pośrednim, ale mogą być również wykorzystane jako 32-bitowe rejestry ogólnego przeznaczenia. Pozostałe rejestry wspomagają różne funkcje systemu, takie jak adresowanie, zarządzanie stosem i stanem procesora, powtarzanie blokowe i przerwania.

W procesorze TMS320C30 obowiązują 2 formaty liczb całkowitych ze znakiem i analogicznie 2 formaty liczb całkowitych bez znaku. Jeden z nich to format 16-bitowy używany w argumentach bezpośrednich, drugi to format 32-bitowy o pojedynczej precyzji. Przyjęte zostały również 3 formaty zmiennoprzecinkowe: format krótki 16-bitowy dla argumentów bezpośrednich (4-bitowy wykładnik, 11-bitowy ułamek, 1 bit znaku), format 32-bitowy o pojedynczej precyzji (8-bitowy wykładnik, 23-bitowy ułamek, 1 bit znaku) i format 40-bitowy o zwiększonej precyzji (8-bitowy wykładnik, 31-bitowy ułamek, 1 bit znaku).

Wewnętrzna pamięć RAM dzieli się na 2 bloki o wielkości $1K \times 32$ bity. W procesorze zawarta jest także wewnętrzna pamięć ROM o pojemności 4Kx32 bity. W jednym cyklu można uzyskać 2 dostępy do dowolnych bloków pamięci wewnętrznej. Oddzielne magistrale programu, danych i DMA pozwalają na równoległe pobieranie programu, odczytywanie i zapisywanie danych oraz wykonywanie operacji DMA. Wewnątrz procesora znajduje się także pamięć podręczna rozkazów (ang. *Instruction Cache*) o pojemności 64×32-bity. Przechowywane są w niej najczęściej powtarzane fragmenty kodu. Pozwala to na przechowywanie kodu w wolniejszej i tańszej pamięci zewnętrznej. Ponadto zwalniane są magistrale zewnętrzne, które mogą być wykorzystane przez DMA lub inne urządzenia.

Wewnętrzny sterownik DMA może wykonywać odczyty i zapisy dowolnych komórek w przestrzeni adresowej procesora bez kolidowania z działaniem CPU. Równoległe działanie DMA i CPU wpływa na zwiększenie szybkości systemu. Dla przykładowego systemu mikroprocesorowego zbudowanego w oparciu o procesor TMS320C30-33 Mz (Rys. 3) maksymalna szybkość transmisji danych z układu we-wy do pamięci RAM wynosi przy wykorzystaniu układu DMA 22.2 megabajty na sekundę (5.55 megasłów 32-bitowych na sekundę), czyli transfer jednego słowa trwa 3 cykle zegara. Przy pojedynczych transferach wymagane są 4 cykle zegara.



Rys. 3. Przykładowy systm z procesorem TMS320C30

Procesor TMS320C30 wyposażony jest także w układy peryferyjne. Należą do nich 2 układy czasowe (ang. *Timer*). Są to liczniki czasowe sterowane zegarem wewnętrznym lub zewnętrznym. Końcowka we-wy układu czasowego może być

wykorzystana jako wejście zewnętrznego zegara albo jako wyjście sygnału czasomierza. Innymi układami peryferyjnymi są 2 porty szeregowe. Każdy z nich może być skonfigurowany do przesyłania danych 8, 16, 24 i 32-bitowych. Sterowane mogą być zarówno zegarem wewnętrznym jak i zewnętrznym. Mogą być także wykorzystane do komunikacji między dwoma procesorami TMS320C30 z gwarantowaną synchronizacją [6].

Na podstawie architektury procesora TMS320C30 produkowany jest procesor TMS320C31. Wykorzystuje on tylko magistralę główną oraz jeden port szeregowy. Obszary pamięci odpowiadające magistrali dodatkowej i drugiemu portowi są zarezerwowane i nie należy ich wykorzystywać. W procesorze tym nie jest również dostępna wewnętrzna pamieć ROM użytkownika, zamiast której zrealizowana została pamięć stała zawierająca program wprowadzający (ang. *Boot Loader*). Umożliwia on załadowanie i wykonywanie programów otrzymanych z komputera głównego lub pamięci zewnętrznych. Programy, które mają być w ten sposób uruchomione muszą rezydować w odpowiednich obszarach przestrzeni adresowej procesora (Boot1, Boot2 lub Boot3), albo pobierane są za pomocą portu szeregowego. Tryb pracy programu wprowadzającego określany jest za pomocą 4 przerwań zewnętrznych procesora przypisanych odpowiednio poszczególnym obszarom *Boot* i *pottowi* szeregowemu. Nagłówek wczytywanego programu musi zawierać informację o długości słowa pamięci, rozmiarze kodu i adresie pod który program ma być wprowadzony. Procesor C31 produkowany jest również w wolniejszej wersji TMS320C31-27, w której cykl rozkazowy wynosi 74 ns, a maksymalna szybkość 27 MFLOPS.

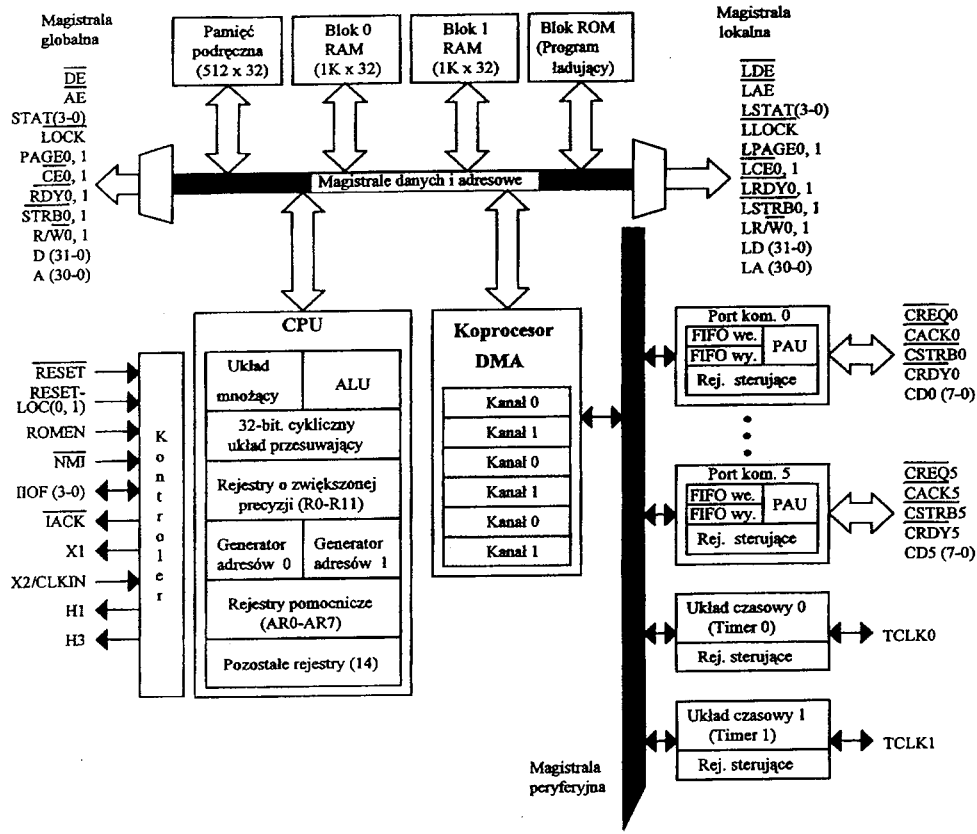
Najnowocześniejszym obecnie rozwiązaniem procesora DSP jest zmiennoprzecinkowy procesor TMS320C40 wyposażony w bogaty zestaw narzędzi do realizacji zadań przetwarzania równoległego. Procesory, które nie są specjalnie zaprojektowane do przetwarzania równoległego, nie nadają się do wykonywania zadań takich jak komunikacja wieloprocesorowa, gdyż obciąża ona bardzo układy we-wy i zmniejsza efektywność obliczeniową. Procesor C40 wyposażony został we współbieżne układy ułatwiające szybką komunikację między procesorami bez zmniejszenia szybkości działania CPU.

W skład jednostki centralnej wchodzi układ mnożący, jednostka arytmetyczno-logiczna (ALU), 32-bitowy cykliczny układ przesuwający, dwie jednostki arytmetyczne rejestrów pomocniczych (ARAU0 i ARAU1) oraz plik rejestrów CPU. Funkcje poszczególnych elementów są analogiczne jak w przypadku procesora C30. Możliwe jest wykonywanie równoległych operacji mnożenia i ALU w jednym cyklu zegara (40 ns). Formaty liczb całkowitych i zmiennoprzecinkowych są takie same jak w C30.

Przestrzeń adresowa procesora wynosi 4G słów 32-bitowych i rozdzielona jest między dwie identyczne magistrale zewnętrzne: globalną i lokalną. W przestrzeni tej odwzorowane są także pamięci wewnętrzne i układy peryferyjne. Wewnętrzna pamięć RAM dzieli się na dwa bloki, każdy o wielkości $1K \times 32$ bity. Blok pamięci ROM zawiera program wprowadzający (ang. *Boot Loader*). Umożliwia on załadowanie programu i danych z dowolnej pamięci zewnętrznej lub poprzez dowolny z portów

komunikacyjnych. Procesor C40 wyposażony jest także w pamięć podręczną programu umożliwiającą przechowywanie 128 słów kodu.

Spośród układów peryferyjnych procesora na szczególną uwagę zasługuje część 8-bitowych portów komunikacyjnych, umożliwiających wykonywanie dwukierunkowych transferów danych między procesorami z szybkością 20 megabajtów na sekundę oraz z gwarantowaną pełną synchronizacją. Każdy z portów pracuje niezależnie od pozostałych i nie wpływa na szybkość jednostki centralnej. Porty komunikacyjne wraz z interfejsami pamięci zewnętrznej umozliwiają realizację systemu przetwarzania równoległego z podziałem zadań na kilka procesorów. Możliwość tworzenia wielu wariantów połączeń pozwala na optymalne wykorzystanie procesów w konkretnym systemie.



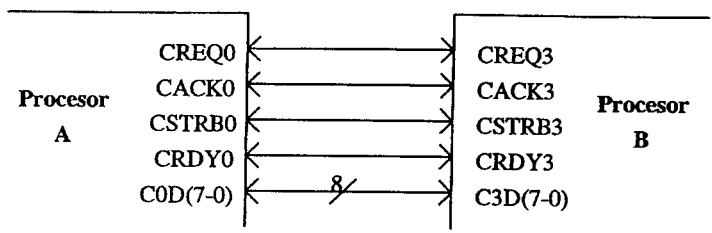
Rys. 4. Schemat blokowy procesora TMS320C40

Jednoczesne wykorzystanie wszystkich portów komunikacyjnych możliwe jest dzięki wewnętrznemu 6-kanałowemu koprocesorowi DMA. Może on dokonywać odczytu i zapisu dowolnych komórek z przestrzeni adresowej procesora bez kolidowania z działaniem CPU. Podobnie jak C30, procesor C40 posiada również 2 układy czasowe ogólnego przeznaczenia [7].

4. RÓWNOLEGŁA PRACA PROCESORÓW SYGNAŁOWYCH

W zestawie rozkazów procesora C30 i C40 znajdują się rozkazy umożliwiające komunikację wieloprocesorową. Poprzez wykorzystanie zewnętrznych sygnałów rozkazy te umożliwiają implementację mechanizmów synchronizacji dostępu do wspólnych zasobów. W procesorze C30 służą do tego celu sygnały XF0 i XF1. XF0 sygnalizuje żądanie wykonania operacji blokowanej. Operacje blokowane obejmują rozkazy ładowania i zapamiętywania. Wspomniane sygnały mogą być wykorzystane do realizacji arbitrażu przy dostępie do pamięci wspólnej kilku procesorow [6].

W podobny mechanizm wyposażony jest procesor C40. Rozkazy blokowane sterują stanem sygnału LOCK jednego z interfejsów zewnętrznych umożliwiając w ten sposób sterowanie dostępem do pamięci wspólnej za pośrednictwem tzw. semaforów. Interfejsy zewnętrzne zawierają także wyjściowe sygnały STAT3-0 określające rodzaj dostępu, jaki ma zostać rozpoczęty na danej magistrali. Na ich podstawie układ arbitrażu może zezwalać na dostęp do wspólnej pamięci i odpowiednio sterować stanami wysokiej impedancji linii sterujących, adresowych oraz danych poszczególnych procesorów.



Rys. 5. Wykorzystanie portów komunikacyjnych do połączenia dwóch procesorow TMS320C40

Chociaż współdzielenie pamięci daje korzyści w pewnych zastosowaniach, to jednak wspólna magistrala poważnie ogranicza komunikację wieloprocesorową. Przeszkodę tę omija wykorzystanie szybkich portów komunikacyjnych procesora C40 (Rys. 5). Połączenie to nie wymaga żadnych dodatkowych układów sterujących [7].

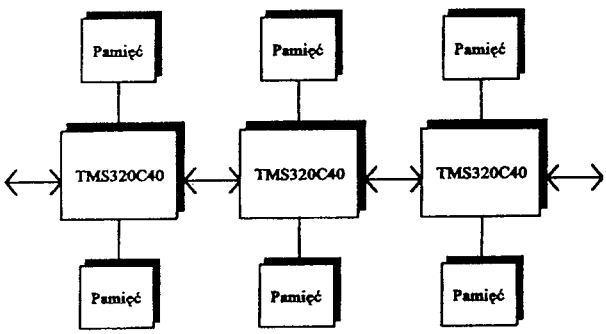
Interfejs komunikacyjny składa się z następujących sygnałów sterujących i linii danych:

- CREQx — uaktywniany przez C40, aby zasygnalizować zgłoszenie wykorzystania magistrali danych portu komunikacyjnego,
- CACKx — uaktywniany przez C40, aby zasygnalizować potwierdzenie otrzymania sygnału CREQx z innego C40,
- CSTRBx — sygnał strobujący portu komunikacyjnego. Procesor wysyłający dane uaktywnia ten sygnał, aby wskazać, że umieścił dane na magistrali danych portu komunikacyjnego,

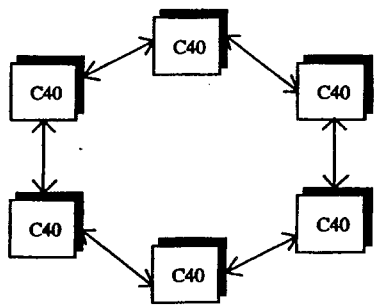
- CRDYx — sygnał gotowości portu komunikacyjnego uaktywniany przez procesor odbierający dane po zakończeniu cyklu odczytu,
- CxD(7 – 0) — magistrala danych portu komunikacyjnego. Może przesyłać dane dwukierunkowo między dwoma procesorami C40 lub między procesorem innymi urządzeniami.

Transfery danych mogą odbywać się w obu kierunkach. Jednostki arbitrażu obu portów współpracują przy generowaniu sygnałów sterujących, aby zapewnić uporządkowane transfery danych z maksymalną możliwą szybkością.

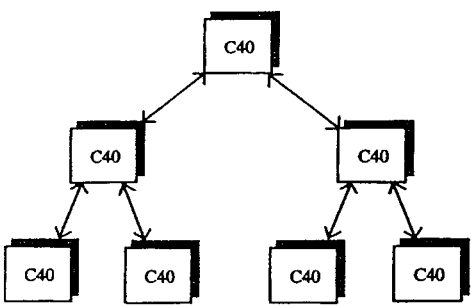
Jak już wspomniano, procesor C40 wyposażony jest w 6 identycznych portów komunikacyjnych pracujących niezależnie. Wynikiem tego jest możliwość tworzenia różnorodnych konfiguracji połączeń zawierających niemal dowolną ilość procesorów. Niektóre z nich przedstawiane są na Rys. 6 – 11. Podane są również przykładowe zastosowania poszczególnych konfiguracji.



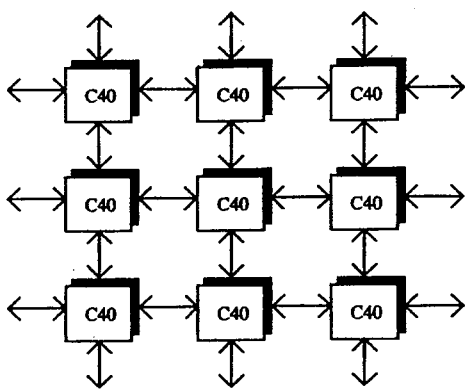
Rys. 6. Połączenie potokowe. Obliczanie splotu i korelacji, operacje potokowe w zastosowaniach graficznych



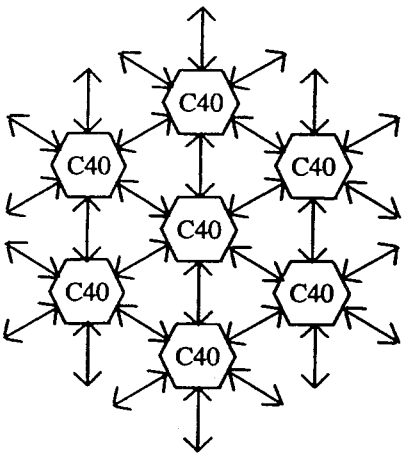
Rys. 7. Dwukierunkowy pierścień. Grupowy port dla uzyskania większej liczby urządzeń we-wy. Efektywny dla sieci neuronowych



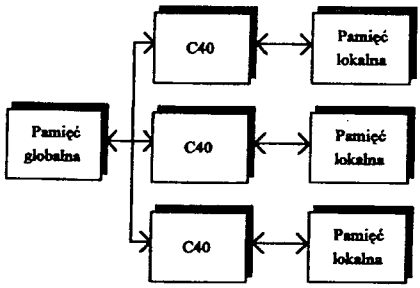
Rys. 8. Struktura drzewa. Rozpoznawanie obrazów i sygnałów mowy



Rys. 9. Dwuwymiarowa tablica. Przetwarzanie obrazów



Rys. 10. Sieć sześciokątną. Użyteczna w analizie numerycznej i przetwarzaniu obrazów



Rys. 11. Pamięć globalna i lokalna. Połączenie umożliwiające wykorzystanie wspólnej pamięci globalnej i prywatnej pamięci lokalnej

5. SPRZĘTOWE MECHANIZMY ZWIĘKSZANIA SZYBKOŚCI DZIAŁANIA PROCESORÓW SYGNAŁOWYCH

Szybkość procesorów sygnałowych wynika z zastosowania pewnych sprzętowych i programowych rozwiązań optymalizujących czas wykonywania programu. Wspomniano już o sprzętowym układzie mnożącym wykonującym operacje mnożenia w jednym cyklu zegara. Jest to szczególnie użyteczne w przypadku algorytmów DSP, w których mnożenie jest podstawową operacją.

Zwiększeniu szybkości wykonywania algorytmów DSP służą także nowe tryby adresowania pośredniego: adresowanie z odwróceniem bitów (ang. *Bit-reversed Addressing*) i adresowanie kołowe (ang. *Circular Addressing*). Adresowanie z odwróceniem bitów może być wykorzystane do przywrócenia właściwej kolejności danych przetworzonych np. za pomocą algorytmu FFT w czasie przepisywania lub transmisji danych [5–6]. Nie ma więc potrzeby wcześniejszego porządkowania danych, co jest istotne podczas pracy w czasie rzeczywistym. Adresowanie kołowe umożliwia utworzenie w pamięci kołowego bufora wymaganego przy obliczaniu splotu czy korelacji.

W opisanych wyżej procesorach zmiennoprzecinkowych rodziny TMS320 istnieje grupa rozkazów równoległych (ang. *Parallel Instructions*), skupiająca w pary pewne rozkazy, które mogą być wykonywane jednocześnie. W grupie tej znajdują się rozkazy równoległego ładowania rejestrów, równoległych operacji arytmetycznych oraz rozkazy arytmetyczno-logiczne wykonywane równolegle z rozkazami zapamiętywania.

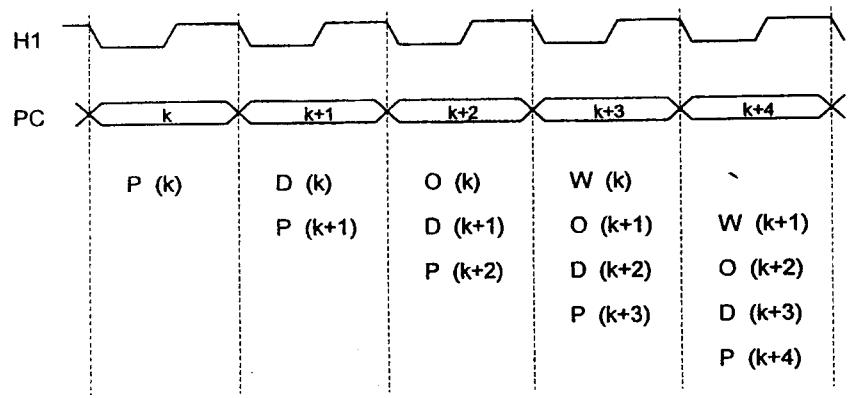
Dla przykładu rozkaz:

```
    MPYF3   srcA, srcB, dst 1
||  ADDF3   srcC, srcD, dst 2
```

powoduje wykonanie równoległych operacji mnożenia i dodawania na danych zmiennoprzecinkowych. Wszystkie rejestry z danymi wejściowymi (argumenty srcA,

srcB, srcC, srcD) odczytywane są na początku cyklu wykonawczego, natomiast rejestry zawierające wyniki operacji (argumenty dst1 i dst2), na końcu tego cyklu. Oznacza to, że jeśli jeden z rozkazów w parze modyfikuje rejestr źródłowy rozkazu drugiego, to modyfikacja ta następuje dopiero po pobraniu danej początkowej. Stwarza to szerokie możliwości optymalizacji kodu i zwiększenie szybkości wykonywania programu.

Duża moc obliczeniowa procesorów sygnałowych możliwa jest dzięki zastosowaniu przetwarzania potokowego (ang. *Pipelining*). Idea przetwarzania potokowego polega na jednoczesnym przetwarzaniu kilku kolejnych rozkazów, przy czym każdy z rozkazów znajduje się na innym etapie przetwarzania. W procesorach zmiennoprzecinkowych C30 i C40 wykorzystuje się przetwarzanie 4-poziomowe (Rys. 12). Działanie procesora sterowane jest więc przez 4 jednostki funkcjonalne.



Rys. 12. Przetwarzanie potokowe w procesorze TMS320C30, H1 — sygnał zegarowy, PC — licznik rozkazów, D — dekodowanie rozkazu, O — odczyt argumentów, W — wykonywanie rozkazu

- Jednostka pobierania — steruje aktualizacją licznika programu i pobiera słowa rozkazów z pamięci.
- Jednostka dekodowania — dekoduje słowo rozkazu i steruje generowaniem adresu.
- Jednostka odczytu — steruje odczytywaniem argumentów z pamięci.
- Jednostka wykonawcza — odczytuje argumenty z pliku rejestrów, wykonuje konieczne operacje i zapisuje wyniki z powrotem do pliku rejestrów lub pamięci.

Każdy rozkaz obsługiwany jest kolejno przez wszystkie jednostki. W celu uzyskania maksymalnej mocy obliczeniowej pracują one równolegle, zatem równocześnie obsługiwane są 4 rozkazy. W ten sposób czas wykonywania pojedynczego rozkazu ulega 4-krotnemu skróceniu.

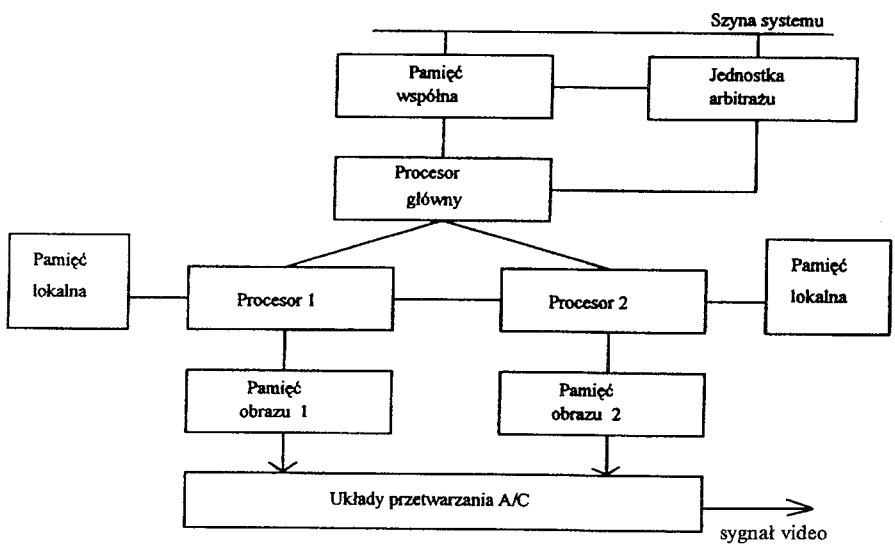
6. WIELOPROCESOROWY SYSTEM GENERACJI OBIEKTÓW GRAFICZNYCH

Przetwarzanie obrazów jest dziedziną gdzie procesory DSP mogą być efektywnie stosowane ze względu na olbrzymią ilość informacji do przetworzenia. By zwiększyć

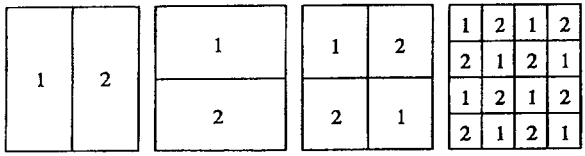
moc obliczeniową systemu i zapewnić pracę w czasie rzeczywistym stosuje się układy wieloprocesorowe [9]. Jednym z zastosowań systemu wieloprocesorowego są układy generacji realistycznej grafiki [1, 4]. Obiekty 3-wymiarowe generowane są zwykle w postaci zestawu figur płaskich (wielokątów wypukłych) aproksymujących widziane powierzchnie. Każdy wielokąt opisany jest poprzez współrzędne wierzchołków w przestrzeni trójwymiarowej i dane R, G, B określajac jego kolor [2, 3]. Aby określić, które punkty (ang. *pixel*) znajdują się wewnątrz, a które na zewnątrz danego wielokąta, procesor obrazu musi znać jego krawędzie w postaci liniowej $f(x,y)=Ax+By+C=0$. Znak funkcji f(x,y) określa położenie danego punktu względem krawędzi. Obraz na ekranie jest rzutem obiektu na wybraną płaszczyznę. Istnieją algorytmy wygaszania niewidocznych powierzchni. Jednym z częściej stosowanych jest algorytm wykorzystujący bufor głębokości (ang. *Z-buffer*). Proste porównanie współrzędnych punktu z danymi w buforze głębokości określa, czy dany punkt ma być wyświetlony czy wygaszony.

Proponowany system składa się z trzech procesorów sygnałowych TMS320C40 (Rys. 13). Procesor główny pracuje w połączeniu potokowym (ang. *pipeline*) z dwoma pozostałymi. Zadaniem jego jest obsługiwanie bazy danych obrazu tzn. generacja współrzędnych wierzchołków i atrybutów wielokątów. Odpowiada on za przekształcenie współrzędnych wierzchołków zgodnie z wybraną regułą wyświetlania (np. perspektywy). Procesor ten może także modyfikować kolory w zależności od położenia czy natężenia źródła światła.

Drugim elementem struktury potokowej jest matryca dwóch procesorów obrazowych (ang. *rendering processors*). Procesory te pracują równolegle i mają dostęp jedynie do wzajemnie wykluczających się obszarów pamięci obrazu (Rys. 14). Cyfry w kwadratach na rys. 14 oznaczają numer procesora przypisanego danemu segmentowi obrazu i pamięci. Głównym zadaniem procesorów obrazowych jest wyświetlenie wielokątów przy wykorzystaniu modelu oświetlenia, usuwanie niewidocznych powie-



Rys. 13. Przykład systemu wieloprocesorowego



Rys. 14. Podział ekranu dla procesorów obrazowych

rzchni, itp. Działania te wymagają przetwarzania olbrzymiej liczby punktów obrazu, a co za tym idzie dużej mocy obliczeniowej.

Każdy z procesorów posiada własną pamieć lokalną wykorzystywaną do realizacji wymaganych algorytmów i przechowywania tymczasowych danych. Zastosowano koncepcję pamięci wspólnej do blokowej wymiany informacji między procesorem głównym i otoczeniem systemu oraz szybkie łącza równoległe do transmisji poleceń i fragmentów obrazów między wszystkimi elementami systemu. Każdy z procesorów steruje oddzielnym bankiem pamięci i przetwarza wybraną część obrazu. Idea rozwiązania polega na podzieleniu ekranu oraz pamięci obrazu na mniejsze części, przy czym każdy z procesorów obrazowych przetwarza fragmenty obrazu zawarte w swojej własnej pamięci. Możliwe są rożne sposoby podziału ekranu i przyporządkowania poszczególnych części procesorom obrazowym (rys. 14). Oba procesory są równouprawnione przy dostępie do wspólnych zasobów. Oznacza to, że dostęp do pamięci wspólnej jest wykonywany aż do jego zakończenia lub upłynięcia czasu przeznaczonego na ten proces. Wyższy priorytet ma jedynie proces odświeżania pamięci dynamicznej.

Istotą właściwej pracy systemu wieloprocesorowego jest optymalny podział zadań, tak aby zrównoważyć obciążenia procesorów działających równolegle. By przeprowadzić analizę obciążeń procesorów obrazowych założono, że czas przetwarzania (wyświetlania obrazu na ekranie) jest liniową funkcją liczby punktów obrazu, podlegających przetwarzaniu (wyświetlaniu). Po wyznaczeniu parametrów pierwszego wielokąta przez procesor główny wyniki przesyłane są do procesorów obrazowych. Kolejne wielokąty procesor główny wyznacza w czasie gdy procesory obrazowe kreślą obraz na ekranie. Średni całkowity czas wyświetlania obiektu dla jednego procesora obrazowego wynosi:

$$T = T_w + T_o * N_p, \tag{1}$$

gdzie: T_w, T_p, N_p oznaczają odpowiednio: czas wyznaczania wierzchołków i atrybutów wielokąta przez procesor główny, czas przetwarzania punktów przez procesor obrazowy i liczbę punktów. Dla matrycy dwóch procesorów obrazowych czas niezbędny do wyświetlenia obiektu jest czasem przetwarzania procesora, któremu przypisana jest większa liczba punktów. Równanie (1) przyjmuje postać:

$$T = T_w + T_p \max\langle k, N_p - k\rangle \tag{2}$$

gdzie: k oznacza liczbę punktów obrazu przetwarzaną przez jeden z procesorów obrazowych. Oczywiście optymalne przetwarzanie ma miejsce, gdy $k = N_p - k$, tj. gdy $k = N_p/2$. Zachodzi wówczas przypadek całkowitego zrównoważenia obciążeń procesorów.

Do dalszej analizy wprowadzono współczynnik niezrównoważenia obciążeń procesorów UBF jako stosunek różnicy punktów przypisanych każdemu z nich do całkowitej ich liczby N_p.

$$UBF = \frac{k-(N_p-k)}{N_p} = \frac{2k}{N_p} - 1 \tag{3}$$

Wartość współczynnika UBF zmienia się w zakresie $\langle -1,1 \rangle$, przy czym zero oznacza całkowite zrównoważenie. Wartości dodatnie oznaczają, że jeden z procesorów jest bardziej obciążony, ujemne, że drugi. Wartości ± 1 są wtedy, gdy tylko jeden procesor przetwarza obraz. Aby uzyskać bardziej ogólne wyniki zaproponowano statystyczny model obrazu. Model obrazu jest zbiorem punktów o współrzędnych (x,y), których prawdopodobieństwo pojawienia się na ekranie jest określone gęstością prawdopodobieństwa rozkładu dwuwymiarowego. Punkty są generowane na ekranie w wyniku niezależnych zdarzeń losowych. Wybrano rozkład normalny określony gęstością prawdopodobieństwa postaci:

$$P(u,v) = \frac{1}{2\pi\sqrt{1-\rho}} e^{-\frac{1}{2(1-\rho^2)}(u^2-2uv+v^2)}, \tag{4}$$

gdzie: $u = \frac{x-\mu_x}{\sigma_x}$, $v = \frac{y-\mu_y}{\sigma_y}$ oznaczają standaryzowane wartości współrzędnych punktów obrazu, ρ — współczynnik korelacji, μ_x, μ_y, σ_x, σ_y — wartości średnie i odchylenia standardowe zmiennych losowych x, y.

Przeprowadzono statystyczne wnioskowanie dotyczące zaliczania typowych obiektów graficznych do klasy opisanej rozkładem normalnym. W praktyce obiekty, których rzuty na płaszczyznę ekranu są figurami wypukłymi mogą być opisane rozkładem normalnym. Zastosowano typowy test χ^2 do badania zgodności rozkładu empirycznego z rozkładem normalnym [10]. Uzyskane wyniki dla typowych obrazów optycznych i termalnych potwierdziły z prawdopodobieństwem $\beta = 1-\alpha = 0.95$ (α — poziom istotności) słuszność założenia, że rozkład punktów na ekranie o współrzędnych (x,y) jest rozkładem normalnym [10]. Można wykazać, że parametry rozkładu mają związek z kształtem i położeniem obrazu na ekranie. Wartości średnie wpływają na położenie obiektu na ekranie. Zerowe wartości średnie oznaczają, że środek rozkładu gęstości prawdopodobieństwa — środek układu współrzędnych Ouv pokrywa się w przeprowadzonych symulacjach ze środkiem ekranu. Niezerowe wartości średnie powodują przesunięcie środka rozkładu gęstości prawdopodobieństwa, a więc i obrazu o wektor $\vec{\mu}_x + \vec{\mu}_y$.

Zbliżone odchylenia standardowe oznaczają większą symetrię obrazu względem osi prostopadłej do płaszczyzny rzutowania i przechodzącej przez maksimum rozkładu gęstości prawdopodobieństwa. Oznacza to, że punkty obrazu rozkładają się równomiernie w obu kierunkach układu Ouv. Wartości odchyleń standardowych bezpośrednio korespondują z wielkością obrazu: większe wartości odchyleń standardowych — większy obraz.

Współczynnik korelacji określający statystyczną zależność między współrzędnymi punktów jest związany z liniową transformacją współrzędnych i określa obrót obrazu

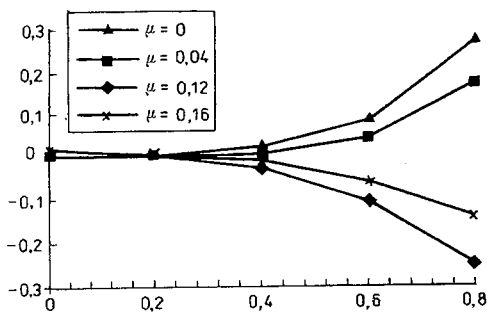
względem osi prostopadłej do płaszczyzny rzutowania i przechodzącej przez maksimum rozkładu.

Aby wyznaczyć współczynnik UBF dla obrazu segmentowanego wyznaczono k i N_p na podstawie rozkładu prawdopodobieństwa pojawienia się punktów obrazu w danym segmencie:

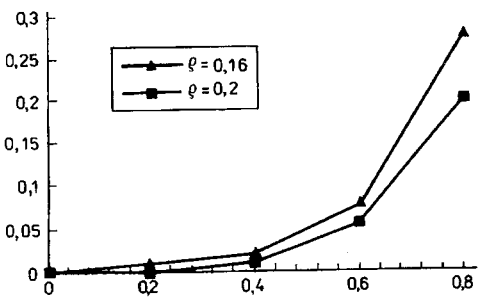
$$k \sim \sum_{i,j} \int_{u_i}^{u_{i+1}} \int_{v_j}^{v_{j+1}} P(u,v)\,du\,dv \qquad N_p \sim \int_{u_0}^{u_{\max}} \int_{v_0}^{v_{\max}} P(u,v)\,du\,dv \tag{5}$$

przy czym u_i, v_j, u_0, v_0, $u_{\max}$, $v_{\max}$ oznaczają granice segmentów przypisane jednemu z procesorów obrazowych oraz granice ekranu. Symulacje zostały przeprowadzone dla modeli obrazów przy segmentacji 4×4. Zmieniano wszystkie parametry rozkładu symulując wielkość, położenie i orientację generowanych obrazów. Wyniki potwierdziły przewidywania. Współczynnik korelacji wpływa silnie na niezrównoważenie obciążeń. Większa statystyczna zależność między współrzędnymi dla obrazów obróconych względem osi prostopadłej do płaszczyzny rzutowania i przechodzącej przez maksimum rozkładu prawdopodobieństwa zwiększa niezrównoważenie. Większe obrazy zapewniają lepsze parametry systemu. Ponadto nie bez znaczenia jest położenie generowanych obrazów. Największe zrównoważenie obciążeń można osiągnąć umieszczając obiekt tak, by współrzędne maksimum rozkładu prawdopodobieństwa pokryły się z węzłami siatki segmentacji obrazu.

Reasumując, z przeprowadzonych badań wynika, że podział ekranu na większą liczbę części sprzyja lepszemu zrównoważeniu systemu. Ponadto lepsze zrównoważenie osiągnięto dla dużych obiektów, co jest szczególnie ważne, gdyż to właśnie duże obiekty pochłaniają dużo czasu procesora.



Rys. 15. Niezrównoważenie obciążeń procesorów w funkcji ρ (orientacji) dla różnych położeń obiektu: $\mu_x = -\mu_v = 0$, 0.04, 0.12, 0.16, $\sigma_x = \sigma_v = 0.2$



Rys. 16. Niezrównoważenie obciążeń procesorów w funkcji ρ (orientacji) dla różnych wielkości obiektu: $\mu_x = \mu_v = 0$, $\sigma_x = 0.2$, $\sigma_v = 0.16$, 0.2

PODSUMOWANIE

W pracy przedstawiono przegląd procesorów sygnałowych rodziny TMS320Cxx, ze szczególnym uwzględnieniem możliwości ich aplikacji w systemach wieloproceso-

rowych. Zaproponowano strukturę systemu z trzema procesorami i przedstawiono analizę statystyczną obciążeń procesorów pracujących równolegle. Badania wykazały, że można osiągnąć zrównoważenie obciążeń i optymalne wykorzystanie procesorów w systemach grafiki realistycznej. Statystyczne podejście do modelowania obiektów zapewniło uogólnienie wyników symulacji na klasę obiektów opisanych danym rozkładem prawdopodobieństwa.

BIBLIOGRAFIA

1. J. Poulton, H. Fuchs, J. Austin, J. Eyles, J. Heinecke, Ch. Hsieh, J. Glodfeather, J. Hultquist, S. Spach: *Pixel-Planes: Building a VLSI-Based Graphic System.* Tutorial: Computer Graphics Hardware. Computer Society Press. Los Angeles, 1988
2. A. Fujimoto, Ch. Perrott, K. Iwata: *A 3-D Graphics Display System with Depth Buffer and Pipeline Processor.* Tutorial: Computer Graphics Hardware. Computer Society Press, Los Angeles, 1988
3. H. Fuchs, B. Jonson: *An Explanable Multiprocessor Architecture for Video Graphics.* Selected Reprints on VLSI, Technologies and Computer Graphics. IEEE Computer Society Press. Los Angeles, 1983
4. R. Earnshaw, T. Heywood, P. Dew: *Parallel Processing for Computer Vision and Display,* Addison Wesley Publishing Company, New York, 1989
5. *Digital Signal Processing, Applications with the TMS320 Family. Theory, Algorithms and Implementation.* vol. 3, Texas Instruments, 1990
6. *Third-Generation TMS320 User's Guide.* Texas Instruments, 1989
7. *TMS320C4x — User's Guide.* Texas Instruments 1992
8. B. Więcek: *Image Processing for Shape Defect Detection.* Rozprawy Elektrotechniczne, 1989, nr 1
9. B. Więcek: *Advanced Architecture of Multi-DSP System for Graphics and Video Applications.* Proc. 13 Conf. IASTED Modeling, Identification and Control, Feb. 21—23.1994, Grindelwald, Szwajcaria
10. S. Firkowicz: *Statystyczne badanie wyrobów.* WNT, Warszawa 1970

B. WIĘCEK, A. SZKODZIK

MULTI-DSP SYSTME DESIGN

Summary

In the work some chosen aspects of Digital Signal Processor systems design are presented. Applying DSP in multiprocessor structures is emphasised. Some hardware mechanisms to speed up its operation and perform some operations in parallel are discussed. As an example the system with three DSPs working for graphics application is presented. The statistical image modelling has been applied to evaluate the system performance — processors load balancing.

Key words: signal processors, multiprocessor systems, processors load analyse

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 3, ss. 379–410

Metodyka charakteryzowania wad rur na podstawie sygnałów przetworników wiroprądowych

ANNA LEWIŃSKA-ROMICKA

Instytut Metrologii i Systemów Pomiarowych, Politechnika Warszawska

Otrzymano 1995.06.14

Autoryzowano do druku 1995.07.28

Opisano nowe kierunki jakie pojawiły się w badaniach nieniszczących przy wykorzystaniu metody prądów wirowych. Trendy te polegają na nowym podjściu do sygnałów przetworników przedstawianych w postaci trajektorii zmian na płaszczyźnie zmiennej zspolonej. Przedstawiono metodykę rozróżniania wad rur na podstawie rozkładu sygnałów przelotowych przetworników generacyjnych w szereg Legendre'a.

Słowa kluczowe: prądy wirowe, przetworniki wiroprądowe.

1. WSTĘP

1.1. WPROWADZENIE DO OPISU METODYKI KLASYFIKACJI WAD OBIEKTÓW

Obecne tendencje rozwojowe w badaniach niniszczących obiektów metalowych przy wykorzystaniu defektoskopii wiroprądowej polegają na dążeniu do uzyskania możliwie najbogatszej informacji nie tylko o występowaniu wad w obiektach, lecz także o charakterze i położeniu wad względem powierzchni obiektów. Ideę tę można zrezlizować poprzez analizę „całkowitego" napięcia (impedancji) przetworników o dowolnej konfiguracji (przelotowe, stykowe, bezwzględne, różnicowe). Przedmiotem analizy są przedstawione na płaszczyźnie zmiennej zespolonej, dla poszczególnych wad obiektów i określonej częstotliwości pracy przetwornika, sygnały przetworników w postaci trajektorii zmian ich napięć lub impedancji [48, 49].

Rozważania i wyniki badań modelowych dotyczą opisu modelu klasyfikacji sygnałów przetworników wieloprądowych, obejmującego automatyczne określanie

wartości wybranych cech odpowiednio przetworzonych sygnałów wiroprądowych generacyjnych przetworników przelotowych różnicowych.

Do charakteryzowania sygnałów przetworników wiroprądowych wybrano współczynniki rozwinięcia zbiorów próbek sygnałów opisujących wady obiektów — w wielomiany Legendre'a.

1.2. PRZEGLĄD PRAC Z ZAKRESU KLASYFIKACJI WAD OBIEKTÓW NA PODSTAWIE SYGNAŁÓW PRZETWORNIKÓW WIROPRĄDOWYCH

Od połowy lat osiemdziesiątych w defektoskopii wiroprądowej obserwuje się tendencje do prezentacji sygnałów, wywołanych przez wady obiektów, na płaszczyźnie zmiennej zespolonej. Sygnały przetworników prezentowane są w postaci trajektorii. W trajektoriach zmian sygnałów przetworników zawarta jest informacja o wielkości, rodzaju i położeniu wad w badanym obiekcie. Mimo, że w trajektoriach zawarta jest tak istotna informacja, dotychczas nie zostały w pełni opracowane syntetyczne metody opisu trajektorii, umożliwiające wykorzystanie tej informacji w systemach kontroli.

Znana metoda tego opisu wykorzystuje aproksymację trajektorii za pomocą szeregu Fouriera, wyniki zastosowania której autorzy zagraniczni badali przykładowo dla rur wykonywanych z materiałów nieferromagnetycznych zawierających wady sztuczne [75, 91].

Przy opracowaniu tej metody założono, że:

1) trajektoria jest opisana zespolonym sygnałem napięciowym przetwornika $u(l)$, gdzie l jest parametrem określającym położenie punktu u na trajektori,
2) $u(l)=u(l+L)$, gdzie L jest długością trajektorii,
3) trajektoria napięcia jest krzywą zamkniętą.

Przy przyjęciu powyższych założeń, S.S. Udpa i W. Lord [91], aproksymowali napięcie $u(l)$ bezpośrednio za pomocą szeregu Fouriera:

$$u(l)=\sum_{n=-\infty}^{\infty} c_n \exp(\mathrm{j}2\pi nl/L)$$

$$c_n=\frac{1}{L}\int_0^L u(l)\exp(-\mathrm{j}2\pi nl/L)\,\mathrm{d}l. \tag{1}$$

W celu uzyskania informacji zawartej w trajektorii o rodzaju wady badanego obiektu zdefiniowano funkcje współczynników szeregu Fouriera c_n, niezależne od skalowania, przesunięcia, rotacji i wyboru punktu startowego. Tworzą one wektor cech, według którego za pomocą metod rozpoznawania obrazów (poprzez analizę skupień, ang. *cluster analysis*) podjęto próbę określenia rodzaju wady [91]. Metoda ta była już przedtem stosowana m.in. do automatycznego odczytywania rękopisów [31, 61, 63, 96] oraz analizy kształtu [44].

Zaletą takiego sposobu podejścia do sygnałów przetworników wiroprądowych jest prostota oraz atrakcyjność, związana z informacją zawartą we współczynnikach szeregu Fouriera. Wadą tej metody jest to, że krzywa rekonstruowana na podstawie współczynników rozkładu trajektorii — w szereg Fouriera nie zawsze jest zamknięta.

Powyższy sposób podejścia do sygnałów przetworników wiroprądowych zakwalifikować można jako parametryczny. Wcześniej, przez badaczy amerykańskich, podejmowane były próby wykorzystania metod nieparametrycznych, związanych z określaniem cech sygnałów w dziedzinie np. czasu i częstotliwości. Wadą metod nieparametrycznych jest to, że wybrane cechy sygnałów są określane na bazie wnioskowania statystycznego i brak jest intuicyjnie satysfakcjonującej bazy dla odnoszenia wybranych cech - do sygnałów wywołanych przez wady.

Dotychczas pełny algorytm określający związek pomiędzy współczynnikami szeregu Fouriera, a poszczególnymi rodzajami wad obiektów nie został wyznaczony, m.in. z dwóch powodów:

1) niejednoznaczność w opisie trajektorii przez współczynniki Fouriera sygnałów wywołanych przez wady naturalne obiektów,
2) braku atlasu wad rur i i odpowiadających im trajektorii.

Podejmowane były takze próby charakteryzowania defektów obiektów przy wykorzystaniu sieci neuronowych. Sygnały dla sztucznych wad rur wykonywanych z materiału nieferromagnetycznego, przetworzone wstępnie w układzie defektoskopu wiroprądowego, przez przetwornik analogowo-cyfrowy, podawano do komputera. Pierwszy etap przetwarzania sygnałów w komputerze polegał na kompresji danych tj. na otrzymaniu zbioru ośmiu współczynników szeregu Fouriera, zgodnie z wyżej opisana metodyką. Współczynniki szeregu Fouriera stanowiły informację wejściowa dla dwuwarstwowej sieci neuronowej. Pozwoliło to na przeprowadzenie przez L. Udpa i S.S. Udpa klasyfikacji zredukowanych współczynników Fouriera opisujących sygnały przetworników do czterech klas. I w tym przypadku została m.in. zastosowana analiza skupień. W przetwarzaniu sygnałów na bazie sieci neuronowych wprowadzane są algorytmy uczenia się. Metoda oparta na wykorzystaniu sieci neuronowych ma tę zaletę, że może być zastosowana w przypadku, gdy klasy rozpatrywanych obiektów (sygnałów przetworników) nie są liniowo separowalne.

Metodę charakteryzowania sygnałów wiroprądowych przetworników przelotowych w oparciu o cechy trajektorii zmian ich sygnałów, to jest kąt położenia fazowego na płaszczyźnie zmiennej zespolonej i tzw. rzędne charakterystyczne podano w [48, 49, 77].

1.3. ZINTEGROWANE SYSTEMY MONITOROWANIA

Niżej zostaną opisane kierunki jakie można zaobserwować przy zastosowaniu metody prądów wirowych do oceny jakości produktów i procesów wytwarzania. Dotychczas przy aplikacjach metody prądów wirowych sygnały zawierające informację o wadach obiektu badanego, podlegały przetwarzaniu w układach defektoskopów, następnie sterowały układami znakowania i sortowania obiektów, a także podlegały rejestracji graficznej. Obecnie takie podejście do zagadnień badań nieniszczących, również w odniesieniu do aplikacji metody prądów wirowych jest niewystarczające. Niezbędnym staje się nadanie układom defektoskopów sztucznej inteligencji opartej na przetwarzaniu sygnałów i stosowaniu metod rozpoznawania obrazów [11, 18, 22, 23, 45, 92]. W odniesieniu do wstępnie przetworzonych sygnałów przetworników

wiroprądowych, w układach defektoskopów, wykorzystana jest analiza statystyczna. Statystyczne metody rozpoznawania próbek są dobrym sposobem podejścia do wyników badań nieniszczących [11]. Dają one możliwosci wyboru cech, reguł decyzyjnych i implementowania analizy opartej na grupowaniu próbek, metod minimalnoodległościowych, itd.

Problemom tym poświęcono wiele prac [27 ÷ 39, 55, 81, 88, 90]. Do wieloparametrowej analizy stanu obiektów wykorzystywane są metody regresyjne.

Procesory sygnałów określają w czasie rzeczywistym wektory cech, charakterystyczne dla danego obiektu otrzymywanego w określonym procesie wytwarzania, na podstawie rozkładu amplitudy przetworzonych sygnałów przetworników wiroprądowych. Wyniki badań eksperymentalnych potwierdzają, że taki wektor cech w sposób klarowny reprezentuje wszystkie powierzchniowe wady obiektów, takich jak np. druty, pręty, rury. Metoda prądów wirowych jest szczególnie przydatna do wykrywania powierzchniowych wad obiektów.

W szczególnych przypadkach wektor cech może zawierać szereg składowych opisujących jakość obiektu (produktu). Jest to istotne w przypadku rozpoznawania (rozróżniania) wad obiektów — na podstawie sygnałów przetworników. W odniesieniu do rur najbardziej istotny w praktyce jest problem rozróżniania wad rur zainstalowanych w wymiennikach ciepła, szczególnie w siłowniach jądrowych.

Przy wdrażaniu metod rozpoznawania w badaniach nieniszczących przy wykorzystaniu metody prądów wiroowych istotne jest podejmowanie zabiegów dla poprawy stosunku sygnałów do zakłuceń. W technice wiroprądowej zakłócenia mają dwa źródła. Jednym z nich są zmiany wymiarów kontrolowanych części, co związane jest z mankamentami procesu wytwarzania a drugim wibracje obiektu badanego względem przetwornika spowodowane nieidealnym prowadzeniem obiektu względem przetwornika. Możliwe jest wykorzystanie filtrów „korelacyjnych"; przy czym problemem jest tu wybór sygnału szumu odniesienia. W pracy [67] opisano metodykę filtracji przy wzięciu pod uwagę zarówno zagadnienia przewidywania i interpolacji zakłóceń w celu zmniejszenia ich wpływu na dalej przetwarzane sygnały.

Dla poprawy stosunku sygnałów do zakłóceń stosowana jest analiza korelacyjna z wykorzystywaniem funkcji autokorelacji [27, 76, 77]. Przedmiotem analizy mogą być także skorelowane sygnału dwu jednakowych przetworników wiroprądowych umieszczonych w torze przesuwu obiektu. Aparat analizy korelacyjnej wyposażony jest w odpowiednią bazę wiedzy.

W systemach przeprowadzana jest krótko- i długoterminowa ocena statystyczna produktu. Wyniki tych analiz, w zaawansowanych systemach, służyć mogą poprzez sprzężenie zwrotne, do sterowania procesami produkcji.

Ogólne tendencje w odniesieniu do budowy inteligentnych systemów wiroprądowych są następujące. Zasadnicza ich cecha polega na zdolności do uczenia się klasyfikacji wad obiektów. Praca systemów składa się z trzech etapów: inicjalizacji, uczenia się i pracy w czasie rzeczywistym. Inicjalizacja polega na wprowadzeniu przez operatora odpowiednich danych dotyczących obiektu, takich jak materiał, wymiary itp. oraz - przetwornika. System wybiera na tej podstawie parametry pracy, m. in. prędkość obrotową sondy i częstotliwości próbkowania syganałów — stosownie do

założeń przewidywanej analizy widmowej. W fazie uczenia się systemu wprowadzane są dane, otrzymane na podstawie badania odpowiedniego zbioru dobrych i wadliwych obiektów. Poprzez zastosowanie analizy fourierowskiej analizowane jest widmo wprowadzonych sygnałów dla określenia zakresu częstotliwości, w którym należy (przy użyciu filtru cyfrowego) poszukiwać informacji odnośnie wad. W fazie pracy w czasie rzeczywistym przeprowadzana jest filtracja cyfrowa sygnałów przetwornika i porównanie tych sygnałów z określonym sygnałem progowym, którego wartości są modyfikowane na bieżąco, stosownie np. do zmian w sprzężeniu obiektu z przetwornikiem. Podejmowane są próby zastosowania różnych metod rozpoznawania próbek, przede wszystkim statystycznych i syntaktycznych.

W zależności od bazy wiedzy (odnośnie sygnałów, wad obiektów, przetwarzania sygnałów, metod wyszukiwania informacji o wadach i diagnozy końcowej) wprowadzanej *a priori* możliwe jest uzyskanie różnych stopni inteligencji systemów.

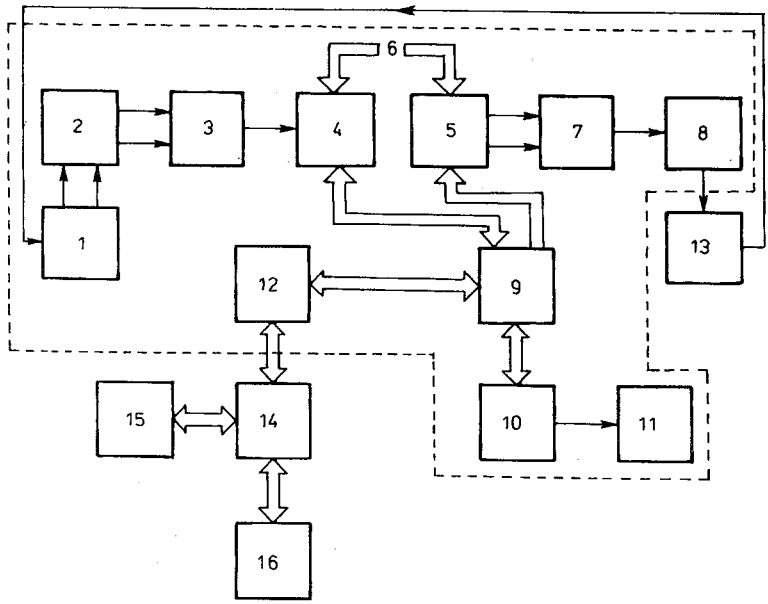
Dla celów rozróżniania wad rur, do rozkładu zbiorów próbek sygnałów przetworników, na podstawie analizy metod rozpoznawania obrazów, została wybrana metoda oparta na wykorzystaniu wielomianów Legendre'a. Podkreślić należy, że autorzy zagraniczni wykorzystywali dotychczas szeregi Fouriera do rozkładu lub rekonstrukcji zmian sygnałów przetworników wiroprądowych, a jako cechy sygnałów wybierali współczynniki rozwinięcia w szereg, ewentualnie dalej poszukiwali cech sygnałów przy zastosowaniu sieci neuronowych. Jednym z istotnych mankamentów zastosowania szeregów Fouriera do analizy sygnałów uzyskiwanych z pomiarów jest wymaganie okresowości funkcji.

1.4. METODYKA OTRZYMYWANIA DANYCH DO MODELOWANIA TRAJEKTORII ZMIAN NAPIĘCIA PRZETWORNIKÓW WIROPRĄDOWYCH

Trajektorie zmian napięcia przetworników wiroprądowych przedstawiane są na płaszczyźnie zmiennej zespolonej, przy czym wzdłuż osi odciętych odkładane są wartości składowej rzeczywistej, a wzdłuż osi rzędnych wartości składowej urojonej napięcia, bądź odpowiednio wartości składowej napięcia przetwornika będącej w fazie i przesuniętej w fazie o kąt $\pi/2$ względem napięcia generatora zasilającego przetwornik. Wymienione składowe mogą być wyodrębnione w układzie z analizatorem napięć (woltomierzem wektorowym) lub w układzie dwukanałowego defektoskopu wiroprądowego.

Schemat blokowy systemu umożliwiającego otrzymanie i prezentację trajektorii przetworników wiroprądowych przedstawiono na rys. 1. System zawiera analizator napięć, przetwornik wiroprądowy oraz mikrokomputer personalny współpracujący z urządzeniami peryferyjnymi. Bloki funkcjonalne analizatora oznaczono na rys. 1 liczbami od 1 do 12.

Analizator zawiera generator napięcia sinusoidalnego (5). Częstotliwość tego napięcia nastawiona jest bezpośrednio w układzie generatora. Amplituda i składowa napięcia wyjściowego generatora (5) nastawiana jest w układzie (7). Napięcie z wyjścia układu (7) doprowadzane jest do wzmacniacza mocy (8), napięcie wyjściowe którego podawane jest do uzwojeń wzbudzających przetwornika wiroprądowego (13). Napięcie wyjściowe przetwornika (13) podawane jest do połączonych kaskadowo układów



Rys. 1. Schemat blokowy systemu z analizatorem napięć 1÷2 — bloki funkcjonalne analizatora napięć; 1 — wzmacniacz wstępny, 2 — układ przełączania kanałów, 3 — układ zmiany zakresów, 4 — układy pomiarowe, 5 — generator, 6 — szyna synchronizacji, 7 — układ nastawiania amplitudy i składowej stałej, 8 — wzmacniacz mocy, 9 — mikrokomputer wewnętrzny, 10 — układy obsługi wyświetlacza i klawiatury, 11 — wyświetlacz i klawiatura, 12 — łącze szeregowe, 13 — przetwornik wiroprądowy, 14 — mikrokomputer zewnętrzny, 15 — monitor, 16 — drukarka

wzmacniacza wstępnego (1), układu połączania kanałów (2), układu zmiany zakresów (3) oraz układów wzmacniacza wstępnego (1), układu przełączania kanałów pomiarowych (4). W układach pomiarowych (4) wyodrębniane są wstępnie pierwsze wyrazy (a_1, b_1) rozwinięcia mierzonego napięcia w szereg Fouriera

$$u(t) = 0.5a_0 + a_1 cos\,\omega t + a_2 cos\,2\omega t + \ldots + a_n cos\,n\omega t + \ldots + \\ + b_1 sin\,\omega t + b_2 sin 2\omega t + \ldots + b_n sin\,n\omega t + \ldots + ,$$

przy czym:

$$a_1 = (2/T)\int_0^T u(t)\,cos\,\omega t\,dt$$

$$b_1 = (2/T)\int_0^T u(t)\,sin\,\omega t\,dt\,. \qquad (2)$$

W przypadku wykorzystania tego sposobu podejścia do analizy sygnałów przetworników wiroprądowych, w układach pracujących w czasie rzeczywistym, można zastosować szybką transformację Fouriera.

Szyna synchronizacji (6) umożliwia przekazywanie informacji odnośnie fazy okresu napięcia generatora, względem którego aktualnie przeprowadzane są obliczenia.

Wewnętrzny mikrokomputer (9) przeprowadza przeliczenia danych z pomiarów, umożliwia komunikację z układem pomiarowym (4) i generatorem (5) — w celu wyboru opcji pomiarów, tj. rodzaju wykonywanych pomiarów (np. charakterystyka częstotliwościowa) oraz cyklu pomiarowego: powtarzalnego i czasu całkowania przy obliczaniu współczynników Fouriera lub pomiarów pojedynczych i wyboru kanału. Mikrokomputer ten zadaje też parametry napięcia generatora (amplituda, składowa stała) oraz steruje pracą układów obsługi wyświetlacza (10) i klawiatury (11).

Otrzymywane wyniki pomiarów (składowe napięcia przetworników) prezentowane są w układzie współrzędnych kartezjańskich (a, jb) — co jest bezpośrednio przydatne dla wyżej sformułowanych celów defektoskopii wiroprądowej.

Analizator, poprzez wbudowane doń łącze szeregowe RS 232 (12), może komunikować się z mikrokomputerem zewnętrznym (14) współpracującym z urządzeniami peryferyjnymi, takimi jak monitor (15) i drukarka (16).

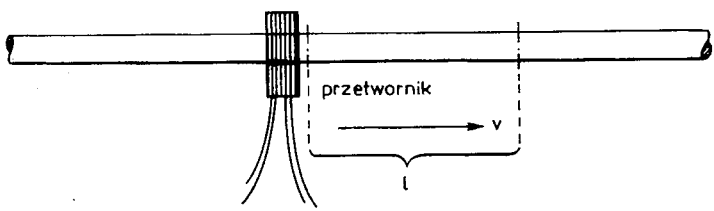
Dane dotyczące składowych napięć przetwornika są wczytywane do zbioru w mikrokomputerze (14). Odpowiedni program umożliwia wizualizację i wydruk trajektorii.

2. METODY SZEREGÓW ORTOGONALNYCH W ROZPOZNAWANIU RODZAJU WAD OBIEKTÓW

2.1. WPROWADZENIE

Niżej opisano metodę rozróżniania wad obiektów dla przypadku, gdy obiekt badany porusza się ze stałą prędkością. Ma to więc bezpośrednie odniesienie do warunków wytwarzania obiektów walcowych, na przykład rur, które zostały wybrane jako przykładowe obiekty badane.

Metodyka polega na znajdowaniu rozkładu w szereg ortogonalny wyników pomiarów sygnałów przetworników, które to wyniki otrzymywane są dla określonego odcinka l w określonym odstępie czasu T (rys. 2). Przy czym zachodzi związek $l = vT$, gdzie l — długość odcinka odniesienia, v — prędkość liniowego przesuwu obiektu, T — czas.



Rys. 2. Ilustracja do opisu metodyki rozróżniania wad obiektów

Istotnym problemem jest dobór długości l odcinka odniesienia lub odstępu czasu T, na którym powinny być wykonywane obliczenia. Wymienione wielkości, jak to wynika z doświadczeń, są ściśle związane z wymiarami przetwornika, a wobec tego

z rozkładem pola magnetycznego, wytwarzanego przez przetwornik (składowa wzdłużna).

Prowadzone na bieżąco obliczenia współczynników szeregu ortogonalnego na zadanym odcinku pozwalają na przeprowadzenie klasyfikacji sygnałów przetworników w zależności od rozróżnianych wad stosując metody rozpoznawania obrazów.

2.2 SZEREGI ORTOGONALNE

Szeregi ortonormalne określane są w przestrzni funkcyjnej (Hilberta) w taki sposób, że iloczyn skalarny

$$<\varphi_i(x), \varphi_j(x)> = \delta_{ij}, \tag{3}$$

gdzie $\varphi_i(x)$ jest funkcją będącą elementem szeregu ortogonalnego,

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \quad \text{jest symbolem Kroneckera,}$$

$< >$ oznacza iloczyn skalarny.

Przy czym, jeśli iloczyn skalarny $<\varphi_i, \varphi_j>$ pozostaje równy zeru dla $i \neq j$, a nie jest równy jedności dla $<\varphi_i, \varphi_i>$, to szereg jest ortogonalny, a nie ortonormalny. Iloczyn skalarny funkcji w przestrzeni L^2 (funkcje całkowalne z kwadratem) określamy jako

$$<f(x), g(x)> = \int_0^T f(x)\cdot\overline{g(x)}\,dx, \tag{4}$$

gdzie $f(x)$, $g(x)$ — funkcje w przestrzeni L^2.

Współczynniki rozkładu funkcji w szereg ortogonalny otrzymuje się ze wzoru Fouriera

$$c_i = \frac{1}{<\varphi_i, \varphi_i>} \int_0^T f(x)\,\varphi_i(x)\,dx. \tag{5}$$

Dla funkcji całkowalnych z kwadratem $f(x) \in L^2$ ciągi c_i należą do klasy l^2, tzn. $\sum_0^\infty c_i^2$ jest skończona.

W przypadku pomiarów, wykonywanych przez nowoczesne przyrządy cyfrowe całkę można zastapić sumą

$$c_i = \frac{1}{<\varphi_i, \varphi_i>} \sum_{k=0}^{\infty} \varphi_i(x_k) f(x_k). \tag{6}$$

Dla przypadku pomiarów, przeprowadzanych w celu wyszukiwania wady obiektu badanego, wzór powyższy przyjmuje postać następującą:

$$c_i = \frac{1}{<\varphi_i, \varphi_i>} \int_0^T f(t-T+\tau)\,\varphi_i(\tau)\,d\tau = \sum_{k=1}^{N} f(t-NT_p+kT_p)\,\varphi_i(kT_p) \tag{7}$$

gdzie N jest liczbą pomiarów $N = T/T_p$, a T_p — okresem powtarzania pomiarów.

Funkcje ortonormalne tworzą ciąg zupełny, jeśli z równości $<f(x), \varphi_i(x)> = 0$ dla $i = 0,1,2...\infty$ wynika, że funkcja $f(x)$ jest elementem zerowym $f(x) = 0$ prawie wszędzie.

Dla zupełnego ciągu funkcji ortonormalnej zachodzi związek

$$f(x) = \sum_{i=0}^{\infty} c_i \varphi_i, \tag{8}$$

gdzie c_i — współczynniki szeregu Fouriera określone wzorem (5).

Równość powyższa jest rozumiana w sensie identyczności w klasie L^2. Wartość całki kwadratu funkcji można obliczyć ze wzoru

$$\int_0^T f^2(t)\,dt = \sum_{i=0}^{\infty} c_i^2 \tag{9}$$

Ponadto dla dwóch funkcji $f(x)$, $g(x) \in L^2$ zachodzi twierdzenie Parsevala, czyli

$$\int_0^T f(x)\,g(x)\,dt = \sum_{i=0}^{\infty} c_i^2\,d_i, \tag{10}$$

lub w zapisie uproszczonym

$$<f(x),g(x)> = <c_i, d_i>\,. \tag{11}$$

gdzie:

c_i — współczynniki rozkładu szeregu Fouriera funkcji $f(x)$,

d_i — współczynniki rozkładu szeregu Fouriera funkcji $g(x)$.

Znaczy to, że iloczyn skalarny w przestrzeni Hilberta jest równy iloczynowi skalarnemu w przestrzeni Hilberta ciągów.

2.3. ZASTOSOWANIE TRYGONOMETRYCZNYCH SZEREGÓW FOURIERA DO OCENY INFORMACJI ZAWARTEJ W TRAJEKTORIACH ZMIAN SYGNAŁÓW PRZETWORNIKÓW WIROPRĄDOWYCH

Trygonometryczne ciągi funkcji ortogonalnych mogą być zapisane następująco:

$$1, \cos\omega t, \sin\omega t, \cos 2\omega t, \sin 2\omega t, ...$$

i tworzą one szeregi zupełne w przedziale $0 - T$, gdzie $T = 2\Pi/\omega$.

Współczynniki trygonometryczne szeregu Fouriera są określone następującymi wzorami

$$a_i = \frac{2}{T}\int_0^T f(x)\cos i\omega t\,dt, \qquad b_i = \frac{2}{T}\int_0^T f(x)\sin i\omega t\,dt, \tag{12}$$

Jeżeli dla pewnego zbioru wartości x suma $s_i(x)$ przy $i \to \infty$ dąży do określonej granicy $s(x)$, to dla tych wartości x mamy do czynienia ze zbliżonym szeregiem Fouriera danej funkcji $f(x)$:

$$s(x)=\frac{1}{2}a_0+a_1 \cos \omega x+a_2 \cos 2\omega x+\ldots+a_i \cos i\omega x+\ldots \tag{13}$$

$$+b_1 \sin \omega x+b_2 \sin 2\omega x+\ldots+b_i \sin i\omega x+\ldots$$

Zespolony szereg trygonometryczny tworzy ciąg funkcji $e^{j\omega it}$, gdzie $i=0, \pm 1, \pm 2,\ldots$, $j=\sqrt{-1}$, $\omega=2\Pi/T$, a T jest okresem. Współczynniki tego szeregu mają następującą postać:

$$c_i=\frac{1}{T}\int_0^T e^{-j\omega it} f(t)\,dt \qquad \text{dla } i=0, \pm 1, \pm 2,\ldots \tag{14}$$

W postaci zespolonej szereg Fouriera można zapisać w sposób następujący:

$$s(x)=\sum_{i=-\infty}^{+\infty} c_i e^{ji\omega x}, \tag{15}$$

gdzie:

$$c_i=\frac{1}{T}\int_0^T f(x)\, e^{-ji\omega x}\,dx=\begin{cases}\frac{1}{2}(a_i-jb_i), & \text{gdy } i>0\\ \frac{1}{2}(a_{-i}+jb_{-i}), & \text{gdy } i<0.\end{cases}$$

Zastosowaniu szeregów trygonometrycznych do analizy, przedstawionych w postaci trajektorii zmian, sygnałów przelotowych wewnętrznych przetworników wiroprądowch przeznaczonych do kontroli rur, poświęcona była praca S.S. Udpy i W. Lorda [91].

Trygonometryczne szeregi Fouriera można stosować do analizy próbek sygnałów poprzez obliczenie w następujący sposób współczynników ich rozwinięcia

$$c_i=\frac{1}{T}\int_0^T e^{-ji\omega t}[a(t)+j\,b(t)]\,dt \tag{16}$$

lub

$$c_i=\frac{T_p}{T}\sum_{k=0}^{N}[a(kT_p)+j\,b(kT_p)]\,e^{-ji\omega kT_p} \tag{17}$$

przy czym $T=N\,T_p$, gdzie N— liczba pomiarów, T_p —okres pomiarów.

S.S. Udpa i W. Lord, na podstawie wartości ośmiu współczynników szeregu Fouriera, rekonstruowali trajektorie zmian sygnałów przetwornika różnicowego dla wady w postaci sztucznego otworu średnicy około 4,75 mm — w zależności od

odległości miejsca rury (wykonanej z materiału Inconel) z wadą od strony ściany sitowej wymiennika ciepła. Jednym z istotnych mankamentów zastosowania szeregów Fouriera do analizy przebiegów uzyskiwanych z pomiarów jest wymaganie okresowości funkcji, co w przypadku rozważania funkcji na odcinku sprowadza się do konieczności spełnienia warunku jednakowych wartości funkcji na obu krańcach. W przypadku, gdy powyższe wymaganie dla funkcji ciągłych nie jest spełnione szeregi też są zbieżne (po uzupełnieniu funkcji przez punkt na krańcach przedziału o wartości równej średniej arytmetycznej granic na obu krańcach). Jednak wprowadzenie tego wymagania dla obliczeń prowadzonych na skończonej liczbie uwzględnianych wartości próbek sygnałów jest kłopotliwe i prowadzi do wolnej zbieżności szeregu.

2.4. WIELOMIANY LEGENDRE'A

Wielomiany Legendre'a są funkcjami z następującego szeregu:

$$\begin{aligned}
\varphi_0(x) &= 1, \\
\varphi_1(x) &= x, \\
\varphi_2(x) &= \frac{1}{2}(3x^2-1), \\
\varphi_3(x) &= \frac{1}{2}(5x^3-3x), \\
\varphi_4(x) &= \frac{1}{8}(35x^4-30x^2+3), \\
\varphi_5(x) &= \frac{1}{8}(63x^5-70x^3+15x), \\
\varphi_6(x) &= \frac{1}{16}(231x^6-315x^4+105x^2-5), \\
\varphi_7(x) &= \frac{1}{16}(429x^7-693x^5+315x^3-35x).
\end{aligned} \tag{18}$$

Współczynniki wielomianów Legedre'a dobierane są metodą ortogonalizacji Schmidta tak, aby

$$<\varphi_i, \varphi_j> = 0 \qquad \text{dla } i \neq j,$$

przy czym zachodzi związek

$$<\varphi_i, \varphi_j> = \frac{2}{2i+1} \qquad \text{dla } i=j, \tag{19}$$

gdzie:

$$<\varphi_i(x), \varphi_j(x)> = \int_{-1}^{1} \varphi_i(x)\, \varphi_j(x)\, dx. \tag{20}$$

Inaczej mówiąc wielomiany Legendre'a są ortogonalne na odcinku $x\in(-1,1)$ w przestrzeni Hilberta z iloczynem skalarnym określonym powyższym wzorem. Wielomiany Legedre'a na krańcach przedziału przyjmują wartości:

$$\begin{aligned} &\varphi_i(1)=1 && \text{dla } i=0,1,2,... \\ &\varphi_i(-1)=1 && \text{dla } i=0,2,4,... \\ &\varphi_i(-1)=-1 && \text{dla } i=1,3,5,... \end{aligned} \tag{21}$$

Oznacza to, że wielomiany Legendre'a nie są funkcjami o jednakowych wartościach na obu krańcach przedziału. W przypadku, gdy rozkładana funkcja ma równe wartości na obu krańcach przedziału, szereg, w który jest rozkładana jest szybciej zbieżny, niż trygonometryczny szereg Fouriera. Ta własność wielomianów Legendre'a jest przyczyną, dla której coraz częściej w rozważaniach przyjmuje się je do rozkładu funkcji na szeregi ortogonalne i dla której wykorzystano je do rozkładu analizowanych sygnałów przetworników wiroprądowych.

Do analizy sygnałów przetworników wiroprądowych wykorzystano następujące zależności:

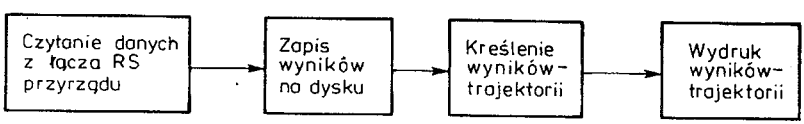
$$\begin{aligned} c_{ai}(t) &= \int_{t-2T}^{t} a(\tau)\, \varphi_i\left(\frac{\tau - t + T}{T}\right) d\tau \\ c_{bi}(t) &= \int_{t-2T}^{t} b(\tau)\, \varphi_i\left(\frac{\tau - t + T}{T}\right) d\tau, \end{aligned} \tag{22}$$

lub w postaci sum:

$$\begin{aligned} c_{ai} &= \frac{1}{N} \sum_{k=0}^{N} a\,(t-(N-k)\,T_p)\, \varphi_i\left(\frac{2k}{N}-1\right) \\ c_{bi} &= \frac{1}{N} \sum_{k=0}^{N} b\,(t-(N-k)\,T_p)\, \varphi_i\left(\frac{2k}{N}-1\right) \end{aligned} \tag{23}$$

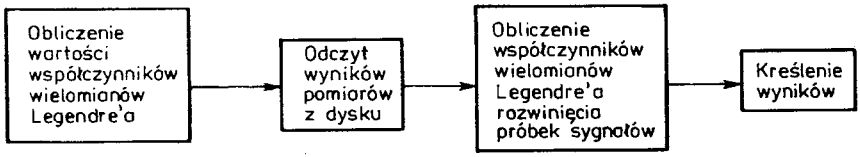
przy czym argument wielomianów Legendre'a zmienia się w przedziale od -1 do $+1$.

Na rys. 3 przedstawiono ogólny schemat funkcjonalny obrazujący współpracę układu pomiarowego (modelu defektoskopu) z komputerem.



Rys. 3. Schemat funkcjonalny ilustrujący transmisję i przetwarzanie sygnałów przetworników

Na rys. 4 pokazano ogólną sieć działań programu obliczania współczynników rozwinięcia w wielomiany Legendre'a próbek sygnałów przetworników wiroprądowych.



Rys. 4. Schemat funkcjonalny ilustrujący etapy otrzymywania współczynników rozwinięcia w wielomiany Legendre'a wyników pomiarów — próbek sygnałów przetworników

3. METODYKA ROZPOZNAWANIA WAD OBIEKTÓW

Do rozróżniania wad obiektów zostały zastosowane metody rozpoznawania obrazów — w oparciu o współczynniki rozwinięcia w szereg wielomianów ortogonalnych, uzyskanych na podstawie pomiarów, próbek sygnałów przetworników wiroprądowych. Przykładowymi obiektami badanymi były rury wykonane z materiałów nieferromagnetycznych (miedziane i mosiężne), zawierające wady o różnych konfiguracjach. Użyto rur o średnicach z zakresu 6 mm — 16 mm. Do badania rur wykorzystano generacyjne zewnętrzne przetworniki wiroprądowe różnicowe.

Metody rozpoznawania obrazów charakteryzują się następującą procedurą postępowania:

1). Zbieranie danych
2). Wybór cech
3). Klasyfikacja obiektu.

ad 1). Zbieranie danych polega w rozpatrywanym przypadku na prowadzeniu pomiarów i rejestracji wyników pomiarów składowej rzeczywistej i składowej urojonej sygnałów uzyskiwanych z przetworników wiroprądowych. Dane te, podczas prowadzonych eksperymentów, podlegały akwizycji w mikrokomputerze kompatybilnym z komputerem IBM PC, poprzez łącze RS 232.

ad 2). Wybór cech w czasie trwania eksperymentu sprowadza się do znajdowania współczynników rozwinięcia zarejestrowanych próbek sygnałów przetworników w szeregi Legendre'a wzdłuż długości obiektu badanego. Z otrzymanych eksperymentalnie wyników i z prostych rozważań wynika, że dla przypadku wystąpienia w obiekcie wady krótkiej, środkowi tej wady odpowiada środek odcinka odniesienia, na którym przeprowadzony jest rozkład zarejestrowanych sygnałów w wielomiany Legendre'a jeśli pierwszy współczynnik rozwinięcia w szereg przyjmuje wartość ekstremalną (powyżej pewnej wartości), a drugi współczynnik jest bliski zeru. Dla tak dobranej chwili wyboru współczynników otrzymuje się ciąg współczynników rozwinięcia w wielomiany Legendre'a — próbek sygnałów. Suma kwadratów

współczynników rozwinięcia w wielomiany zawiera informację o wielkości zarejestrowanej wady obiektu, a ich wartości — informację o rodzaju wady.

Przeprowadzając normalizację współczynników — dla danej wady

$$C_i = \frac{c_i}{\sqrt{\sum_{k=0}^{M} c_k^2}} \tag{24}$$

gdzie:

c_i — współczynnik rozwinięcia,

C_i — znormalizowany współczynnik rozwinięcia,

M — liczba uwzględnionych wielomianów

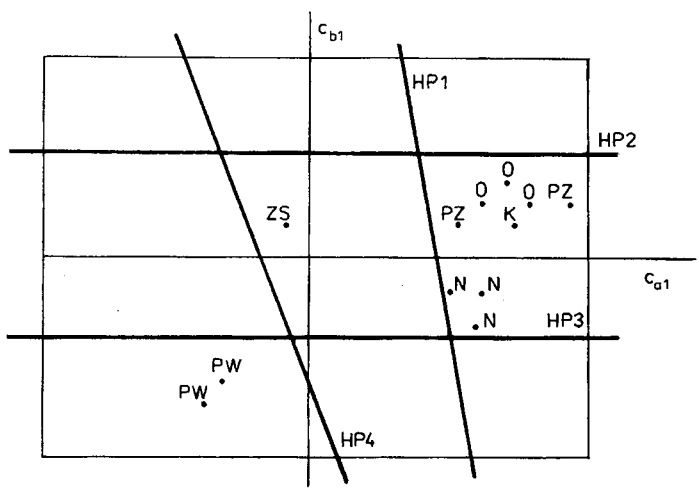
uniezależniamy wyniki od wielkości wady.

Wybierając punkty odpowiadające środkowi wady obiektu, w środku odcinka na którym prowadzony jest rozkład w wielomiany i jako cechy przyjmujące znormalizowane współczynniki wielomianów Legendre'a, dalsze rozważania prowadzone są w przestrzeni cech.

ad 3). Klasyfikacja obiektu prowadzona jest przy zastosowaniu zabiegu wykreślania hiperpłaszczyzny w M-wymiarowej przestrzeni cech, gdzie M jest liczbą współczynników branych pod uwagę.

Hiperpłaszczyzną w M-wymiarowej przestrzeni jest zbiór liniowy o wymiarach M-1; na przykład w przestrzni trójwymiarowej hiperpłaszczyzną jest płaszczyzna. Hiperpłaszczyzną na płaszczyźnie jest prosta.

Na rys. 5 przedstawiono ideę reprezentacji na płaszczyźnie cech wartości cech (znormalizowanych współczynników rozwiniecia w wielomiany Legendre'a) — dla



Rys. 5. Reprezentacja na płaszczyźnie cech wartości cech (znormalizowanych współczynników rozwinięcia w wielomiany Legendre'a) sygnałów przetwornika wiroprądowego różnicowego zawierającego rurę z wadami (rysunek poglądowy) HP1 ÷ HP4 — hiperpłaszczyzny

różnych wad rur. Na osi odciętych odłożono wartości współczynnika C_{a1}, a na osi rzędnych — współczynnika C_{b1}, obliczone na podstawie zależności (23) i (24) na odcinku, którego środek znajduje się w środku wady.

Symbole literowe (ZS, O, PZ, K, N, PW), odpowiadają poszczególnym rodzajom wad sztucznych (imitujących wady naturalne) rur, jakie mogą być brane pod uwagę przy badaniu sygnałów przetworników wiroprądowych. Jako ZS oznaczono zmniejszenie zewnętrznej średnicy rur (zmniejszenie grubości ścianki), O — otwór przelotowy przez jedną ściankę rury, K — kanałek (rowek) wzdłużny od strony zewnętrznej ścianki rury, N — nacięcie poprzeczne od strony zewnętrznej ścianki rury, PW — pierścień od strony wewnętrznej ścianki rury.

Zależność opisująca płaszczyznę w trójwymiarowej przestrzni x, y, z dana jest poniższym równaniem

$$ax+by+cz+d=0. \tag{25}$$

W rozpatrywanym przypadku klasyfikacja poszczególnych rodzajów wad — na podstawie współczynników rozwinięcia w wielomiany Legendre'a próbek sygnałów (składowej rzeczywistej i składowej urojonej) przetwornika odbywa się poprzez obliczanie wartości i określanie znaku następującego wyrażenia:

$$f_k C_{a1}+g_k C_{b1}+h_k \tag{26a}$$

w przypadku uwzględniania dwóch cech C_{a1}, C_{b1} oraz

$$f_k C_{a1}+g_k C_{b1}+h_k C_{a3}+l_k C_{b3}+m_k \tag{26b}$$

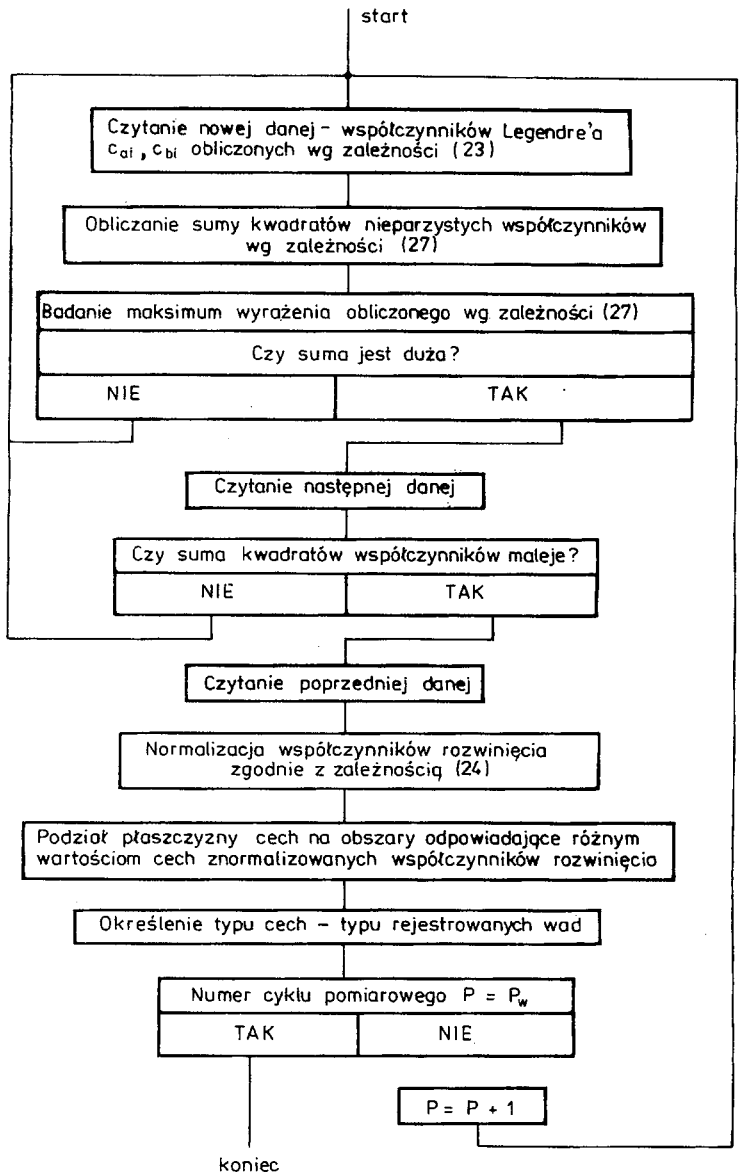
w przypadku uwzględniania czterech cech C_{a1}, C_{b1}, C_{a2}, C_{b2}, przy czym f_k, g_k, h_k, l_k, m_k są stałymi współczynnikami określającymi k-tą hiperpłaszczyznę. Jest to więc sprawdzenie, czy wartość danej cechy należy do zbioru, który mieści się po odpowiedniej stronie właściwej hiperpłaszczyzny (wyrażenie (26) przyjmuje wartości odpowiednio mniejsze lub większe od zera).

Na rys. 6 przedstawiono sieć działań programu do badania cech, przetworzonych zgodnie z wyżej opisaną metodyką, sygnałów przetworników. Program realizuje zadanie, które polega na określaniu na zbiorze wielomianów Legendre'a C_{ai} i C_{bi}, kiedy poniższe wyrażenie przybiera wartość większą od przyjętej wartości progowej

$$\sum_{i=1}^{M} C^2_{a(2i+1)}+C^2_{b(2i+1)}, \tag{27}$$

gdzie: $i=1, 3, 5,...,M$, M — liczba branych pod uwagę wielomianów.

W wyniku prowadzonych obliczeń wartości cech (znormalizowanych współczynników wielomianów Legendre'a) sprawdzane jest położenie punktów odpowiadających wartościom cech względem płaszczyzn separujących. Otrzymuje się przy tym tablicę obrazującą położenia punktów odpowiadających wartościom cech. Przykładem niech będzie tu tablica 1. Poszczególnym położeniom wartości cech przyporządkowane zostają liczby, na przykład w kodzie dwójkowym.



Rys. 6. Sieć działań programu do badania cech przetworzonych sygnałów przetworników wiroprądowych

Tablica 1.

Przykładowa tablica obrazująca położenie wartości cech

Nr płaszczyzny	HP1	HP2	HP3	HP4	HP5
Położenia wartości cech	–	+	+	–	–
Opis liczbowy	0	1	1	0	0

Tablica 2.

Ilustracja sposobu klasyfikacji wad — na podstawie położeń wartości cech

Kod wady	Rodzaj wady
0 0 0 0 0	nie znana
0 0 0 0 1	nacięcie
0 0 0 1 0	rowek
0 0 0 1 1	otwór
0 0 1 0 0	pierścień zewnętrzny
0 0 1 0 1	pierścień wewnętrzny
0 0 1 1 1	zmiana średnicy
.....	.....

Dalszy krok polega na przeprowadzeniu klasyfikacji wad obiektu badanego na podstawie otrzymanych wartości cech przy wykorzystaniu tablicy stanów, której koncepcję podano w tablicy 2. W zależności od położeń wartości cech obliczonych dla sygnałów, wywołanych przez poszczególne wady obiektu, poszczególnym rodzajom wad obiektu badanego przypisane zostają liczby, na przykład w kodzie dwójkowym.

Wyniki przeprowadzonych eksperymentów i prac nad przetwarzaniem sygnałów przetworników wiroprądowych, zgodnie z opracowaną metodyką, wskazują, że metodyka ta może zostać wykorzystana do odróżniania sygnałów uzyskiwanych przy występowaniu następujących wad obiektów: zewnętrznych, wewnętrznych i zmian średnicy.

Prezentowana metodyka daje wyniki, które mogą być wprawdzie uzyskane poprzez wizualną ocenę obrazu na ekranie monitora trajektorii zmian sygnałów przetworników wiroprądowych ale jest metodyką umożliwiającą automatyczną klasyfikację sygnałów (próbek). Przy pojedynczych „ręcznych", pomiarach do przyjęcia jest wizualna klasyfikacja wad obiektów, natomiast przy badaniach podczas i poza procesami wytwarzania obiektów (rur, prętów, drutów) taka metoda klasyfikacji jest nie do przjęcia. Ponadto metodyka rozróżniania wad obiektów pozwala na zmniejszenie wpływu zakłóceń na sygnał przetworników wiroprądowych (na przykład czynników powodujących zmiany sprzężenia obiektów z przetwornikami jak nierównomierności prowadzenia obiektów), ponieważ każdy ze współczynników rozkładu, wstępnie przetworzonych w układzie defektoskopu, sygnałów przetworników obliczny jest na podstawie wielu punktów pomiarowych. W eksperymencie otrzymano

dobre wyniki przy wzięciu pod uwagę 20 punktów pomiarowych. Koszt prowadzenia obliczeń według opracowanego algorytmu w przyrządzie inteligentnym (to jest zawierającym mikroprocesor) jest pomijalny.

4. METODA PRZEBIEGÓW WZORCOWYCH

W oparciu o metodę rozkładu próbek sygnałów, otrzymywanych z defektoskopu, możliwe jest dalsze przetwarzanie informacji w szereg ortonormalny przy wykorzystaniu przebiegów wzorcowych. Jeśli dany rodzaj wady jest opisany przez funkcję $w(x)$, a inny rodzaj wady przez funkcję $u(x)$, które to funkcje mogą być rozłożone w szeregi ortogonalne i chcemy odróżniać sygnały wywołane tylko przez te dwa rodzaje wad, to proponuje się następującą metodykę odróżniania sygnałów. Metodyka ta jest oparta na wykorzystaniu prebiegów wzorcowych.

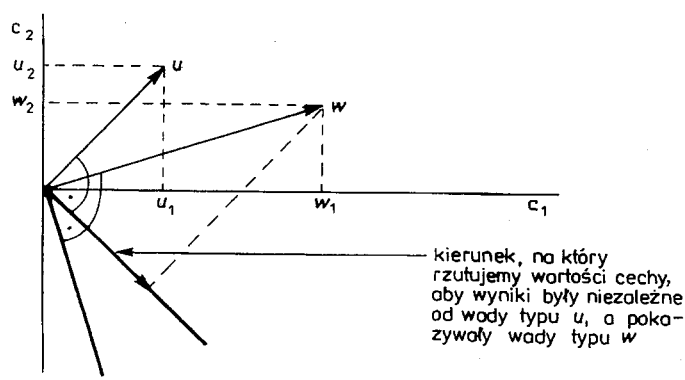
Rozpatrywane funkcje zapiszemy następująco:

$$w(x) \cong \sum_{i=0}^{N} w_i \varphi_i(x)$$
$$u(x) \cong \sum_{i=0}^{N} u_i \varphi_i(x) . \tag{27}$$

W przestrzeni cech opisujących sygnały (w przestrzeni współczynników rozwinięcia w szeregi ortogonalne) można znaleźć hiperpłaszczyzny otrogonalne do płaszczyzn, w których występują sygnały odpowiadające odróżnianym rodzajom wad obiektów (na rys. 7 płaszczyzny ortogonalne oznaczono grubą kreską).

Zależność opisująca płaszczyznę w trówymiarowej przestrzeni cech dana jest równaniem;

$$w_1 c_1 + w_2 c_2 + w_3 c_3 + \ldots = 0 , \tag{28}$$



Rys. 7. Ilustracja do opisu metodyki rozróżniania sygnałów przetworników wiroprądowych

gdzie c_i są wartościami nowych współrzędnych w przestrzni cech. Zazwyczaj informacja wnoszona przez współczynnik o indeksie zero jest mało przydatna, więc w_0 nie występuje w przestrzeni cech.

Rzutując wektory, na przykład **w** na hiperpłaszczyznę prostopadłą do **u**, otrzymujemy wektor składowych w przestrzeni cech dobrany tak, aby był ortogonalny do cechy, którą chcemy odróżnić. Przy dwóch obiektach (rodzajach wad) odróżnianych w przestrzeni cech możliwe jest następujące postępowanie.

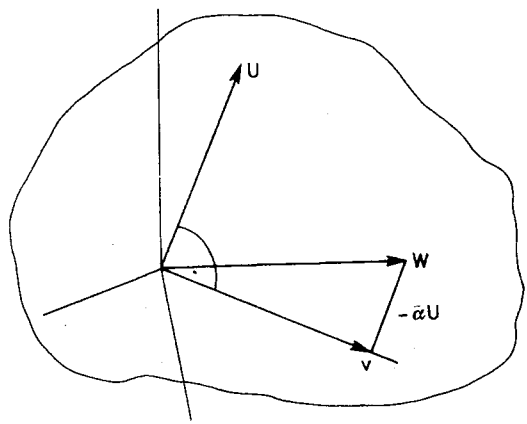
Opisujemy hiperpłaszczyznę normalną do funkcji u:

$$u_1v_1+u_2v_2+\ldots+u_nv_n=0\,. \tag{29}$$

gdzie v_1 — współrzędne płaszczyzny prostopadłej do u.
Znajdujemy rzut wektora **w** na płaszczyznę v prostopadłą do **u**. Rzut wektora **w** na płaszczyznę v otrzymuje się przez odjęcie od wektora **w** wektora $\alpha\cdot\mathbf{u}$. Wektor **v** rzutu wektora **w** na płaszczyznę prostopadłą do **u** (rys. 8) można zapisać jako:

$$\mathbf{v}=\mathbf{w}-\alpha\mathbf{u}\,, \tag{30}$$

gdzie α — skalar.



Rys. 8. Rzutowanie wektorów cech

Ponieważ zachodzi też związek

$$<\mathbf{u},\,\mathbf{v}>=0 \tag{31}$$

(warunek ortogonalności), więc po znalezieniu iloczynu skalarnego obu stron równania (30) przez u otrzymamy

$$<v,\,u>=<w,\,u>-\alpha<u,\,u>=0 \tag{32}$$

lub

$$\alpha=\frac{<w,\,u>}{<u,\,u>}\,. \tag{33}$$

Korzystając z zależności (33) można obliczyć współrzędne rzutu wektora **w** na płaszczyznę prostopadłą do **u**

$$v = w - \frac{<w,\ u>}{<u,\ u>} \cdot u, \tag{34}$$

lub zapisując w postaci sumy:

$$v_i = w_i - \alpha u_i \tag{35}$$

$$\alpha = \frac{\sum_{i=1}^{N} w_i u_i}{\sum_{i=1}^{N} u_i^2} \tag{36}$$

gdzie N jest liczbą rozważanych współczynników funkcji ortogonalnych.

Funkcja wzorcowa $w(x)$, odpowiadająca określonej wadzie W, może być zastąpiona funkcją

$$v(x) = \sum_{i=1}^{N} v_i(x)\, \varphi_i(x), \tag{37}$$

która na mocy twierdzenia Parsevala jest ortogonalna do funkcji $u(x)$ odpowiadającej wadzie lub innemu zdarzeniu, jak np. zmianie średnicy U, tzn.

$$<v(x),\ u(x)> \overset{df}{=} \int_0^T v(x)u(x)\,dx \quad \overset{\text{z twierdz. Parsevala}}{=} \quad <v_i,\ u_i> \overset{df}{=} \sum_{i=1}^{\infty} v_i c_i \tag{38}$$

ale ponieważ zachodzi zależność (31), więc funkcje $v(x)$ i $u(x)$ są ortogonalne.

Inaczej mówiąc, jeśli cel obliczeń polega na wyszukiwaniu wady W przy jednoczesnej eliminacji wpływu zdarzenia U (np. zmiany średnicy obiektu), to należy obliczyć funkcję $v(x)$ na podstawie wcześniejszych pomiarów, a czasie trwania eksperymentu obliczać wartość następującego wyrażenia:

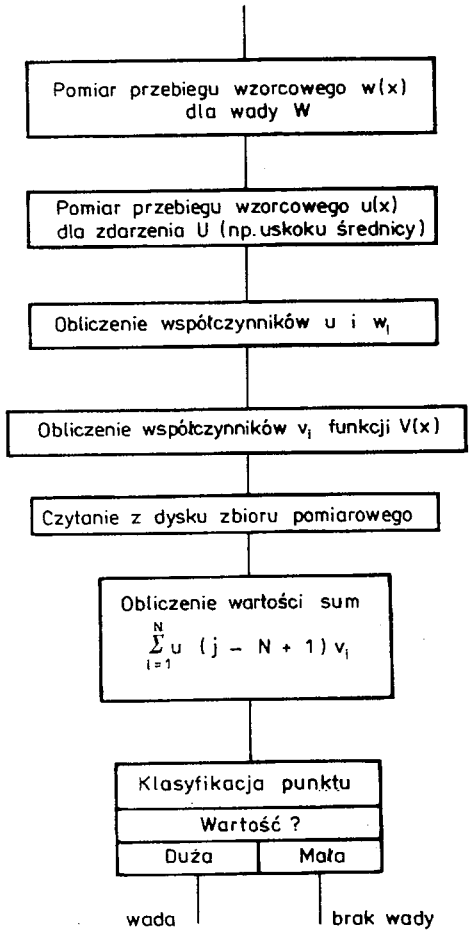
$$\int_0^T f(t - T + \tau)\, v(\tau)\, d\tau, \tag{39}$$

lub sumę

$$\sum_{k=1}^{N} f(t - NT_p - kT_p)\, v\left(\frac{kT_p}{T}\right). \tag{40}$$

Otrzymana w ten sposób suma teoretycznie przyjmuje wartość równą zeru dla zdarzenia o kształcie U, a daleka jest od zera — dla wady — zdarzenia o kształcie W.

Sieć działań programu testującego wpływ wady — przy wykorzystaniu przebiegów wzorcowych przedstawiono na rys. 9.

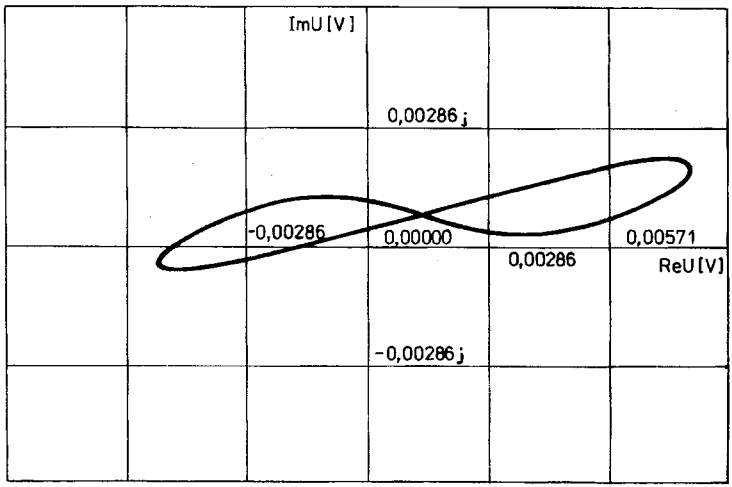


Rys. 9. Sieci działań programu realizującego algorytm postępowania z wykorzystaniem przebiegów wzorcowych

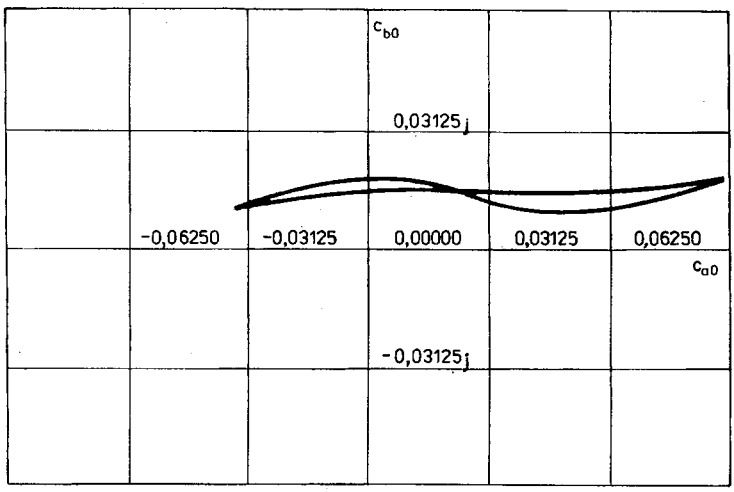
5. PRZYKŁADOWE WYNIKI ZASTOSOWANIA PROCEDURY ROZRÓŻNIANIA WAD RUR PRZY WYKORZYSTANIU WIELOMIANÓW LEGENDRE'A

Niżej przedstawiono przykładowe wyniki badań i przetwarzania sygnałów wiroprądowych przetwornika przelotowego przeznaczonego do badania rur o średnicy zewnętrznej 16 mm. Jako przykładowe obiekty badane zostały wybrane rury wykonane z materiałów nieferromagnetycznych, tj. z miedzi i mosiądzu, o średnicach w zakresie od 6 do 20 mm. W rurach wykonano szereg wad sztucznych, takich, jak otwory przelotowe przez jedną ściankę, kanałki (rowki) wzdłużne i nacięcia poprzecz-

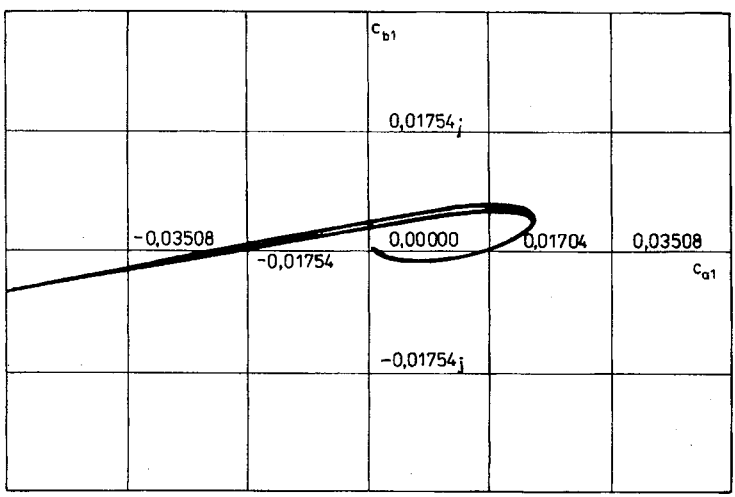
a)


ImU[V]
0,00286 j
-0,00286
0,00000
0,00286
0,00571
ReU[V]
-0,00286 j

b)


c_{b0}
0,03125 j
-0,06250
-0,03125
0,00000
0,03125
0,06250
c_{a0}
-0,03125 j

c)


c_{b1}
0,01754 j
-0,03508
-0,01754
0,00000
0,01704
0,03508
c_{a1}
-0,01754 j