

POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

KWARTALNIK
ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND
TELECOMMUNICATIONS
QUARTERLY

TOM 43 — ZESZYT 4/97



WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1998

RADA REDAKCYJNA

Przewodniczący

prof. dr hab. inż. ALFRED ŚWIT
członek rzeczywisty PAN

Członkowie

prof. dr hab. inż. DANIEL JÓZEF BEM — czł. koresp. PAN, prof. dr hab. inż. MICHAŁ BIAŁKO — czł. koresp. PAN, prof. dr hab. inż. STEFAN HAHN — czł. koresp. PAN, prof. dr inż. ANDRZEJ HAŁAS, prof. dr inż. ZDZISŁAW KACHLICKI, prof. dr hab. inż. WŁADYSŁAW MAJEWSKI, prof. dr hab. inż. JÓZEF MODELSKI, prof. dr inż. JERZY OSIOWSKI, prof. dr inż. WITOLD ROSIŃSKI — czł. rzecz. PAN, prof. dr hab. inż. STEFAN WĘGRZYN — czł. rzecz. PAN, prof. dr hab. inż. WIESŁAW WOLIŃSKI — czł. koresp. PAN, prof. dr inż. ANDRZEJ ZIELIŃSKI, prof. dr inż. MARIAN ZIENTALSKI

REDAKCJA

Redaktor Naczelny

prof. dr hab. inż. WIESŁAW WOLIŃSKI

Zastępca Redaktora Naczelnego

doc. dr inż. KRYSTYN PLEWKO

Sekretarz Odpowiedzialny

mgr KRYSTYNA LELAKOWSKA

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/fax (0-22) 660 77 37, 25 29 18+aut. sekr.

Telefony domowe: Redaktora Naczelnego: 12 17 65

Zastępcy Red. Naczelnego: 826 83 41

Sekretarza Odpowiedzialnego: 25 29 18

WYDAWNICTWO NAUKOWE PWN SA

Warszawa, ul. Miodowa 10

Ark. wyd. 10,75 Ark. druk. 9,5

Podpisano do druku w lutym 1998 r.

Papier offsetowy kl. III 70 g. B-1

Druk ukończono marcu 1998 r.

Skład: *Amel*, Warszawa

Druk i oprawa: Drukarnia Braci Grodzickich, Żabieniec, ul. Przelotowa 7

Intermodulation distortion effects on the bandwidth frequency of the SCM AM/VSB CATV CD system

MIKHAIEL SALIM MAKHOUL

Instytut Telekomunikacji, Politechnika Warszawska

Received 1997.06.09

Authorized 1997.09.17

This paper investigates technologies of bandwidth subcarrier multiplexing external amplitude modulation vestigial sideband cable television coherent detection (SCM AM/VSB CATV CD) systems, together with intermodulation distortion (IMD) phenomena that limit the performance of these systems. For this reason, we analyze deeply the intermodulation distortion of the AM/VSB external modulators. Also, we shall see description of the external modulation techniques. We also give attention to the system configuration and system noise effects. Reduction of IMD effects is our goal and interest so that the general system performance must be good.

Our analysis contains most of the novel aspects

1. We derive power spectral density of the SCM AM/VSB for CATV CD system.
2. We suggested a block diagram for AM/VSB CATV CD system used for transmission SCM signals.
3. Impact of the linearizer circuit for canceling third order intermodulation distortion.
4. Use general LO laser for many subscribers for minimized system costs.

We found as matter-of-fact, the second order intermodulation distortion (IMD_2) vanishes while the linear and third order intermodulation distortion (IMD_3) terms are independent of the bias point (Φ_T) of external AM/VSB modulator. By investigation it was clear IMD_3 that depends on δ parameter that limits possessively and compensates the nonlinearizer modulator. Our calculations show when δ equal to 0.5 the IMD_3 decreases by approximately 14 dB. We noted IMD_3 decreased more than 30 dB by proper selection of the band-pass frequency characteristics for transmission of video signals. In reality, we found that by proper selection of the signal band-pass frequency characteristics, we can improve the general system performance to an acceptable level. This development may possibly lead to an effective SCM AM/VSB CATV CD system. We shall also see that when the system works on one octave, the maximum intermodulation distortion power falls

within band-pass frequency SCM AM/VSB video signals and this selection of the system operation does not reduce IMD₃ effects and system performance is not improved.

Key words: modulation amplitude, cable television, systems coherent detection, intermodulation distortions.

1. INTRODUCTION

Amplitude modulation vestigial sideband (AM/VSB) coherent subcarrier multiplexed (SCM) cable television (CATV) transmission system currently receiving a resurgence of the interest due to recent advances in optical component technology. AM/VSB format reduces the complexity of the hardware required at remote sites in the transmitter, reducing both the initial and maintenance of costs. In general with all this, even at once general cost of the system still very expensive, because the complexity of the coherent SCM receivers. In this work, we report on a SCM AM/VSB CD system consisting of a diode-pumped Nd:YAG laser, amplitude modulator, linearizer circuit that inserts between the external AM/VSB and SCM radio frequency (RF) combiner, general (LO) consisting of a diode-pumped is using in the receiver for more one subscriber to reduce the total system costs. The predistortion linearized technique employed trying to decrease or cancel third-order intermodulation distortion products in the external modulator.

General performance of the system suffers from nonlinear external modulator and the sources noise as the shot noise (n_{sh}), thermal noise (n_{th}), relative intensity noise (RIN) and beat of the local oscillator (LO) laser with amplified spontaneous emission noise (n_{ASE}). The performance of the direct detection (DD) and coherent detection (CD) fibre optics links are distortion by received thermal noise at low optical signal power levels, and RIN at high optical signal power levels. The thermal and shot noise greatly affects the performance of the coherent SCM CATV system, performance such optic links are impaired by the receiver thermal noise at low optical signal power level and by laser relative intensity noise (RIN) at high optical signal power levels.

In the coherent SCM CATV systems, appear clearly, the laser phase noise may cause signal spectrum brooding and distort the general performance system seriously, but it is neglected in some cases by choice optimum IF bandwidth [1, 14].

To date, analog fibre optics links based on coherent detection have received relatively little attention; notable exceptions are subcarrier multiplexed coherent systems for video distribution. In a CD system, the signal field before detection is combining with the field from a LO laser. Coherent optical detection systems have several advantages over direct detection systems, among from these:

1. Coherent optical systems can approach shot-noise-limited performance with the sufficient LO power.
2. Coherent optical systems are able to separate wavelength-division multiplexed signals with high frequency resolution, and can detect the phase of the optical carrier.

3. Thus, while direct detection optical systems are used for amplitude modulation, coherent detection optical systems can use AM.

We are now through with the introductory part of this paper, and in the remaining parts we are going to see in details the following sections: externally modulators, system configuration, system noise effects, distortion of the unlinearized modulator, PSD of the modulation signals and minimized IMD₃ effects, in the end we shall also see discussions and results.

2. EXTERNALLY MODULATORS DESCRIPTION

Early work in external modulation for CATV transmission was based on integrated optical LiNbO₃ Mach-Zehnder modulators (MZM).

Externally modulators optical links have a higher dynamic range and lower noise figure at high frequencies than directly modulated links. Higher advantage of using an external modulator is that its L-V curve offers a bias point where second order distortion is nulled out the composite second order. Significant improvements in the dynamic range can be obtained by modifying the structure of the modulator such that its response is more linear in the required region of operation. Analog optical links employing external integrated sources are important, for a variety of applications such as CATV distribution. Distributed feedback-light diode (DFB-LD) currently offers a working solution to AM fibre optics transmission, though not free of problems.

Feasible alternate to overcoming the difficulties associated with direct laser modulation is the use diode-pumped laser such as Nd:YAG and Nd:YLF laser, in conjunction with LiNbO₃ external (amplitude) modulator.

Realization of external modulation linearized systems has required applying of a host of technology such as

1. Integrated optics.
2. Solid state lasers and RF electronics.
3. Developing novel approaches to broadband electronic circuitry and analog electrooptic system architectures.

Several modulator linearization techniques can reduce intermodulation distortion resulting from nonlinearities. Each approach has advantages, and disadvantages. Several Schemes have been proposed for linearizing the response of electrooptic modulator (EOM) such as

1. Electronic predistortion [2, 3]. This is the most commonly used technique in commercial externally modulation CATV optical transmitters.
2. Electronic feedforward [2].
3. Cascaded linearized modulator [4].
4. Polarization mixing [5].
5. Parallel EOM structure [6].

2.1. Predistortion Modulator

The predistortion linearization technique uses an additional nonlinear device that generates distortion products that are equal in amplitude, but opposite in phase, with the distortion products generated by the Mach-Zehnder, Fig. 1 shows block

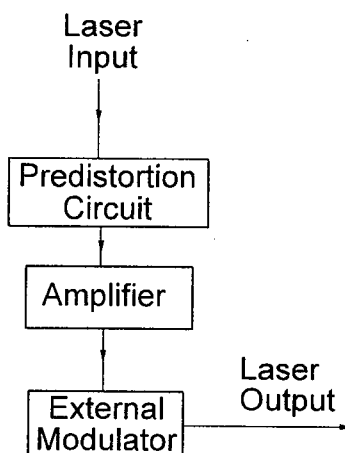


Fig. 1 schematic diagram of the laser transmitter system with the linearized electrooptic modulator use predistortion circuit

diagram of the transmitter use predistortion circuit. This technique has been previously applied to reduce the intermodulation distortion of CATV amplifiers. The video signal is applied to nonlinear predistortion circuit, which generates the necessary component. To achieve a distortion cancellation over the entire video spectrum, the linear amplifier that follows the predistortion circuit must have flat amplitude versus frequency characteristic. The diodes, which are used in the predistortion circuit, must also have negligible parasitic reactances over all frequencies of interest.

2.2. Feedforward Linearization

Idealized principle of operation of a feedforward linearization system is depicted clearly in the Fig. 2 in which all gains are assumed normalized to unity. Modulator output power is split and detected in an auxiliary receiver. Distortion and noise generated in the modulator are denoted by (e) , which defined as the difference between optical output and radio frequency (RF) input (see Fig. 2). The comparator subtracts the modulated optical output $(S + e)$ and RF input (S) of the modulator and generates an electrical error signal $(-e)$ which drives an auxiliary laser, the output of which is combined with output of the main laser to generate

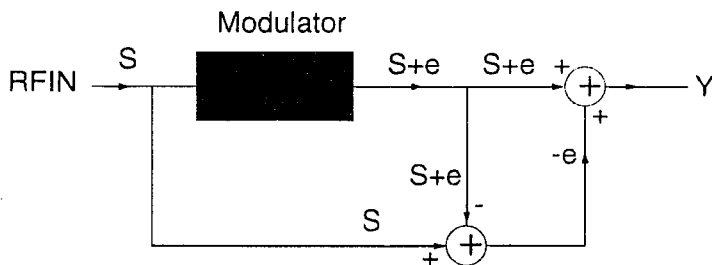


Fig. 2 general block diagrams of the feedforward linearization signal use for minimized intermodulation distrotrion

a corrected output $\{(S + e) - e = S\}$ which is ideally an undistorted replica of the input signal (S).

2.3. Cascaded Linearized Modulator

Fig. 3 shows block diagram of the cascaded linearized modulator, the linearized modulator is actually a cascade of two modulators, the second modulator compensates for the nonlinearities introduced by the first modulator. Suppression of the nonlinearities is achieved by adjusting the bias voltages of the two modulators and the relative amplitudes to each port.

Cascaded modulator has Several advantages, among of are the following

1. It is completely passive, no electronic predistortion, feedback or feedforward techniques are used.

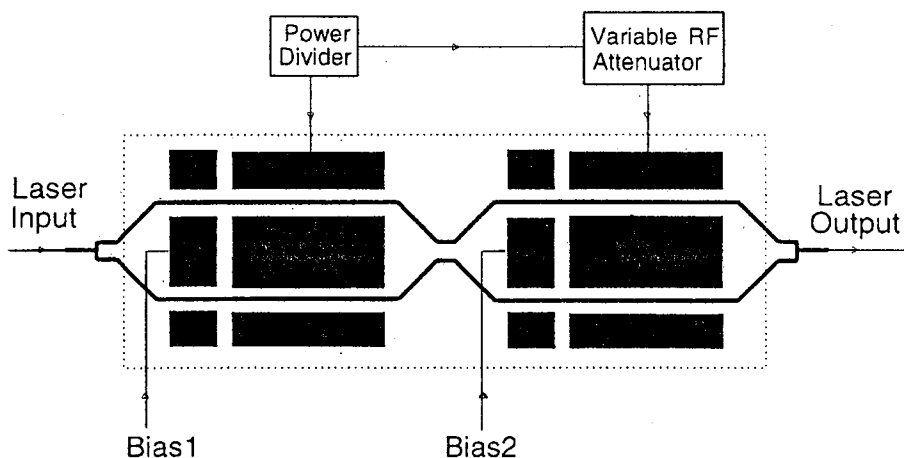


Fig. 3 block diagram of the cascaded linearized modulator, the second modulator compensates for the nonlinearities indroduced by the first modulator

2. The linearization takes place entirely on the optical chip; thus, the modulator's frequency response is not limited by the bandwidth of electronic components.
3. Cascaded modulator is less sensitive to manufacturing errors as compared to other schemes.
4. For the cascaded modulator, crosstalk problems are less severe than the parallel scheme [5, 6] and a lower bias voltage is usually required as compared to other modified directional coupler schemes.

The main disadvantages of the cascaded modulators are

1. Precise and critical adjustments of the bias voltages and the amplitudes and phases of the RF signals at the input ports are required.
2. Another problem is temperature sensitivity. When the temperature changes, the bias points of the two modulators drift in opposite directions; thus, the third order nonlinearity (IMD_3) increases much faster than in conventional modulators. Therefore, extra circuitry is needed to control the bias voltages and stabilize the temperature.

2.4. Polarization-Mixing Technique For Linearizer Modulator

Polarization-mixing technique requires no high-speed electronics and is therefore particularly suited for microwave analog optical links. The polarization-mixing linearization technique is applicable to optical modulators supporting two polarization modes with differing electrooptic sensitivities. By adjusting the modulator bias such that the transfer functions of the two polarization modes have slopes of opposite sign, and by adjusting the relative optical power in each mode, the modulator nonlinearities can be substantially reduced relative to the linear response.

Fig. 4 shows optically polarization-mixing technique applied to an interferometric modulator. A polarizer inserts between transmitter laser and the modulator. The modulator supporting TE and TM polarization modes. Under the ideal phase bias condition, both polarization modes biased at the halfpower point, which provides maximum linear response and eliminates all even-order nonlinearities, the polariza-

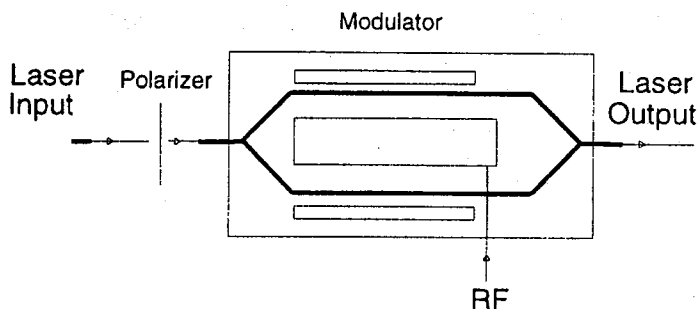


Fig. 4 schematic diagram of the polarization-mixing linearization technique is applied to an interferometric analog amplitude modulator

tion-mixing technique can also provide suppression of the cubic nonlinearity. Here, the cubic response of the TE mode exactly cancels TM mode. For nonideal modulator phase bias, the cubic nonlinearity can still be suppressed but a different polarization power ratio is necessary. Since the polarization-mixing technique requires that most of the optical power be coupled into the less-sensitive polarization mode, there is an overall reduction in modulator sensitivity. However, for a fixed modulation index relative to the total input optical power, there can be a significant reduction in distortion levels. The modulator bias voltage must be adjusted to reduce the intermodulation distortion.

2.5. Parallel Electrooptic Modulator

Fig. 5 shows dual parallel modulation based on the generalization of a linearization technique by using two polarization of the light in a single modulator. Dual parallel linearization is designed to use the distortion created by a secondary Mach-Zehnder interferometer to cancel the distortion produced by the primary Mach-Zehnder. When a multi signal is applied the IMD_2 refers to a super position of all terms within a stated bandwidth of a particular frequency that are due to second order distortion, and IMD_3 refers to a super position of all terms within the bandwidth of the frequency.

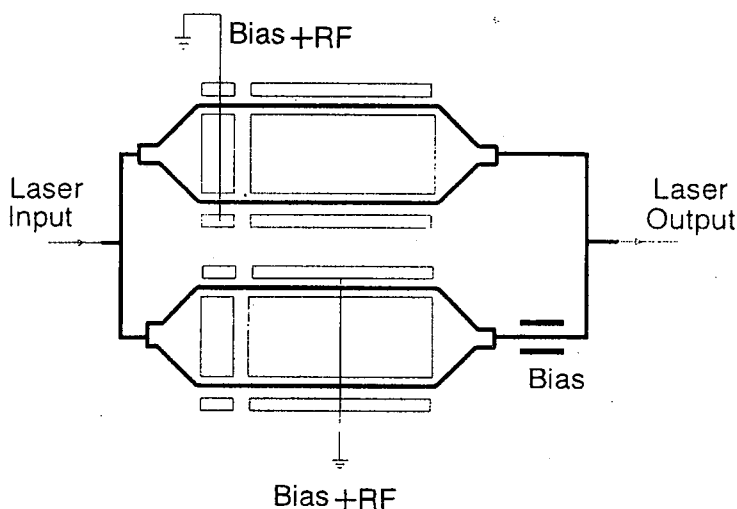


Fig. 5 configuration of the dual parallel Mach-Zehnder modulator, the second modulator uses for cancel the distortion introduced by the first modulator

The RF drive power is split between the two modulators (in a ratio of $1:\alpha^2$) so that the secondary modulator has a higher modulation depth and greater distortion ($\alpha > 1$). By providing more optical to the primary modulator than to secondary modulator, the third order distortion products created in the secondary modulator

can be made to cancel the distortion products from the primary modulator with a small cancellation of the fundamental signal.

3. SYSTEM CONFIGURATION

Fig. 6 and Fig. 7 show transmitter and receiver configuration for *subcarrier multiplexing optical amplitude modulation vestigial sideband cable television coherent detection* (SCM AM/VSB CATV CD) systems. We see clearly at the Fig. 6 transmitter (SCM AM/VSB CATV), the input signals (1 ... N) is combined by the combiner to form radio frequency SCM signal, and applied on the linearizer network.

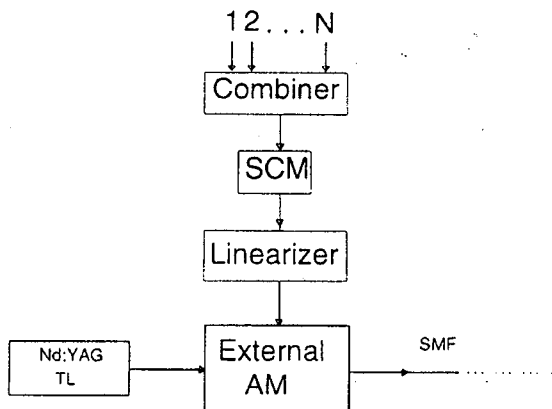


Fig. 6 block diagram of the transmitter subcarrier multiplexing for amplitude modulation vestigial sideband cable television system

The response of this net work as we assume and wont must be canceling the intermodulation distortion in a very well way. Therefore, must correct nonlinear characteristic external modulator AM/VSB that causes IMD. System uses linearizer circuit, because it is very important for improving the general performance to such systems.

Our interest is to be able to do linearizer circuit and linear network in the transmitter of the (SCM AM/VSB CATV) coherent systems. A linear net work is introduced to allow the calculation of an equivalent total system bandwidth. Linearizer circuit is tolerant to dip frequency variation of the laser's frequency response and amplitude variation. The linearizer device has well advantage for improving systems work and their performance. Linearizer circuit in the system is possible regardless of the specific frequency response of an individual laser. May be disadvantage, their implementation in the systems became somewhat uneasily.

The output of the linear network is the input to an optical external AM/VSB. Within this modulator circuit the information signals (SCM) with laser input (Nd:YAG) will be tracked. In consequence of, the output signals of the modulator is

modulated with analog amplitude modulation vestigial sideband. Clearly seen in the references [1, 2, 5, 6, 8] these kinds of the external amplitude modulation. We offered general discussion about that in the last section.

A transmitter laser (TL) is set up in the transmitter, and the modulated optical signal is mixed before transmission with the local laser light. We assumed that the local transmitter laser (TL), and the subcarrier multiplexing (SCM) signal have the same state of polarization. Therefore, the coherent (SCM AM/VSB CATV) system is unaffected by the state of the polarization in the transmission fiber. The subcarrier multiplexing amplitude modulation vestigial sideband (SCM AM/VSB) signal is transmitted, and distributed within a single mode fiber (SMF). Direct optical amplification by an optical amplifier before or after the splatter should be effective. Receiver detects the transmitted optical signal using balanced photodetector (PN), optical amplifier, baseband filter (BPF), quadratic detector and low pass filter (LPF). Fig. 7 shows that clearly.

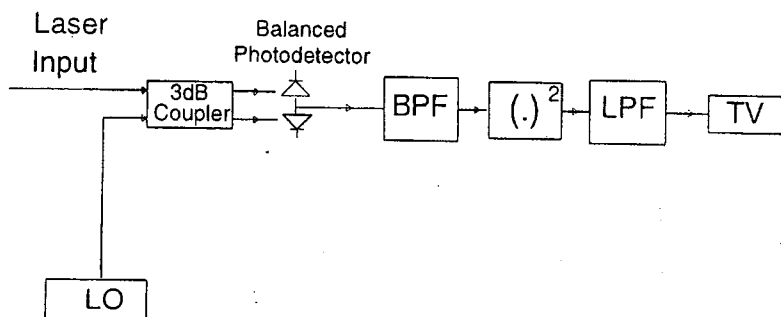


Fig. 7 block diagram of the coherent detection subcarrier multiplexing amplitude modulation vestigial sideband cable television system

The received signal is combining with the local oscillator (LO) laser signals using 3 dB couplers, which assumed coherently with the optical received signal. The LO laser is the key component of a coherent optical. For coherent system purpose, a local oscillator signal should be generated in the receiver with both the proper frequency and phase, this means that there will be phase and frequency coherence between the received signal and LO laser. Photodetector will be detecting the coherent signal, but that is never mean the detected signal is not consist the noise which one of these is the relative intensity noise (RIN). Important to say, RIN is one of many important reasons that cause distortion the general performance SCM AM/VSB CATV coherent systems.

The detected photo-current in the intermediate frequency ($\omega_{IF} = \omega_{LO} - \omega_s$) form is distorted by noises, but using linearizer circuit that is enough reason for improving total system performance. Bandpass filter (BPF) is using and matching for passing the interest signals to the quadratic demodulator. As we assume the receiver must

offer information signals that equal to the transmitted information signals, generally, which is our goal and interest for the communication systems. To reduce the total SCM AM/VSB CATV coherent system cost, we suggest common local oscillator laser (CLO) can be used in the coherent receiver system. Fig. 8 shows in detail the block diagram receiver system.

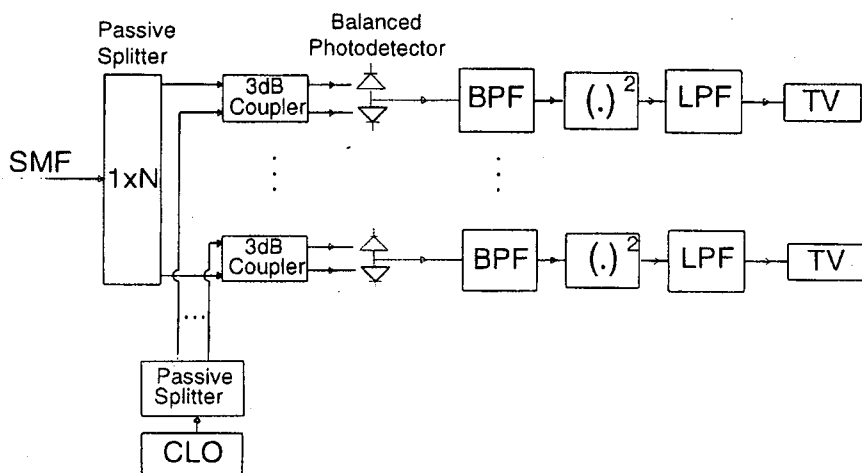


Fig. 8 schematic diagrams of the coherent detection receiver for subcarrier multiplexing amplitude modulation vestigial sideband cable television system uses common local oscillator laser

For improving general system performance, we use the linearizer circuit before the modulator at the transmitter to reduce effects of the third order intermodulation, and balanced photodetection at the receiver to reduce the effects of the relative intensity noise.

4. SYSTEM NOISE EFFECTS

In this section, we evaluate the noise components that are distortion general performance system. We consider the noise from optical modulation transmission and detection. We assume that $(\Phi_L = 0)$ in the analysis for SCM AM/VSB CATV CD system which means the laser phase noise is unimportant because the system is insensitive to linewidth laser sources.

Additive system noise $n_{\text{tot}}(t)$ which has Gaussain's distribution is white and consists of

1. Thermal noise (n_{th}) and shot noise (n_{sh}).
2. Relative intensity noise (n_{RIN}).
3. Beat of the LO laser with optical amplified.
4. Amplified spontaneous emissions (n_{ASE}) noise.

Power spectral density (PSD) will express as single side band and each noise source is given by [4, 9] and [10]. Thermal noise $n_{th}(f)$ PSD can be expressed as

$$N_{th}(f) = \frac{4K_B T}{R_L}, \quad (4.1)$$

where K_B is boltzmann's constant, T is the temperature and R_L is the input resistance of the receiver.

The shot noise power spectral density is expressed as

$$N_{sh}(f) = 2qR(P_{LO} + P_{s_{tot}}), \quad (4.2)$$

where R is the responsivity of the detector, typically is equal to ($R = 1$), P_{LO} is the optical power local oscillator, $P_{s_{tot}}$ is the optical power of the all signals incident on the photodetector receiver, which is approximately equal to the optical power of one channel that is incident on the receiver (P_s) multiplied by the number of channels N .

The local oscillator amplified spontaneous emission beat noise $N_{ASE}(f)$ is given by

$$N_{ASE}(f) = 2R^2 \cdot h\nu \cdot (F - 1) \cdot G_{tot} \cdot P_{LO}, \quad (4.3)$$

where h is Planck's constant, ν is the optical frequency, which about 200 KHz for Nd:YAG laser, F is the optical amplifier noise figure (about 3 dB), and G_{tot} is the total optical gain.

Finally, PSD of the laser relative intensity noise can be expressed as

$$N_{RIN}(f) = R^2 \left(P_{LO}^2 \cdot 10^{\frac{RIN_{LO}}{10}} + P_s^2 \cdot 10^{\frac{RIN_s}{10}} \right), \quad (4.4)$$

where RIN_s and RIN_{LO} are the relative intensity noise laser contributions ($-110 \div -165$ dB/Hz) of the transmitter and local oscillator laser, respectively.

Summing the equations from (4.1) to (4.4) the PSD of additive noise current is given by

$$N(f) = \frac{4K_B T}{R_L} + 2qR(P_{LO} + P_{s_{tot}}) + 2R^2 \cdot h\nu (F - 1) \cdot G_{tot} \cdot P_{LO} + R^2 \left(P_{LO}^2 10^{\frac{RIN_{LO}}{10}} + P_s^2 10^{\frac{RIN_s}{10}} \right). \quad (4.5)$$

When P_{LO} is sufficiently large about 20 mW for Nd:YAG (typically $\gg 0$ dB), the shot noise is large than the thermal noise. Also, local oscillator amplified spontaneous emission (LO-ASE) beat noise larger than the shot noise provided that the optical amplification is sufficiently large.

5. DISTORTION OF THE UNLINEARIZED MODULATOR

The normalized modulator output U is defined as the AC component of the modulator optical output [2]

$$U = 1 + \sin \left[\frac{\pi V(t)}{V_\pi} + \varphi_b \right], \quad (5.1)$$

where $V(t)$ is the applied RF voltage, V_π is a constant called the half-wave voltage, namely the voltage required to switch the optical output from the full extinction to maximum, and φ_b is the bias phase.

For simplify the equation (5.1) can be rewritten as

$$U = 1 + \sin (X_0 + \varphi_b) \quad (5.2)$$

with X_0 defined as the normalized modulating voltage

$$X_0 = \frac{\pi V(t)}{V_\pi} \quad (5.3)$$

expressed in radian units as the equivalent phase retardation in the electrooptic modulator interferometer arms.

A general SCM CATV AM/VSB modulating signal consisting of N unmodulated (carrier + video) signals plus direct current (dc) bias V_B

$$V(t) = V_0 \sum_{i=1}^N [\cos(\omega_i t) + S_{0i}(t)] + V_B, \quad (5.4)$$

where V_0 is the amplitude SCM Signals, $S_{0i}(t)$ is the video signal and $(\cos \omega_i t)$ is the signal carrier.

Substituting the equations (5.4) in the (5.3) yields

$$X_0 = \frac{\pi V_0}{V_\pi} \sum_{i=1}^N [\cos(\omega_i t) + S_{0i}(t)] + \frac{\pi V_B}{V_\pi}, \quad (5.5)$$

where the term $(\pi V_0/V_\pi)$ is defining as the modulation index (m), the equation (5.5) can be rewritten as

$$X_0 = m \sum_{i=1}^N [\cos(\omega_i t) + S_{0i}(t)] + \frac{\pi V_B}{V_\pi} \quad (5.6)$$

by substituted the equation (5.6) in the equation (5.1) yields

$$U = 1 + \sin \left\{ m \sum_{i=1}^N [\cos(\omega_i t) + S_{0i}(t)] + \frac{\pi V_B}{V_\pi} + \varphi_b \right\}, \quad (5.7)$$

where the total retardation bias is

$$\Phi_T = \frac{\pi V_B}{V_\pi} + \varphi_b. \quad (5.8)$$

Consider the equations (5.6) and (5.8), the equation (5.7) can be expressed as

$$U = 1 + \sin(X_0 + \Phi_T), \quad (5.9)$$

where

$$\sin(X_0 + \Phi_T) = \cos(\Phi_T) \cdot \sin(X_0) + \sin(\Phi_T) \cdot \cos(X_0) \quad (5.10)$$

the equation (5.9) can be rewritten as

$$U = 1 + \cos(\Phi_T) \cdot \sin(X_0) + \sin(\Phi_T) \cdot \cos(X_0). \quad (5.11)$$

Retaining terms up to second order in the power series expansions of the sine and cosine functions

$$\sin(X_0) = X_0 - \frac{1}{6}X_0^3 + \dots \quad (5.12)$$

$$\cos(X_0) = 1 - \frac{1}{2}X_0^2 + \dots \quad (5.13)$$

by substituting the equations (5.12) and (5.13) in the equation (5.11) yields

$$U = 1 + \cos(\Phi_T) \cdot \left(X_0 - \frac{X_0^3}{6}\right) + \sin(\Phi_T) \cdot \left(1 - \frac{X_0^2}{2}\right) \quad (5.14)$$

substituting the equations (5.6) in the last equation yields

$$U = 1 + \cos(\Phi_T) \left\{ m \sum_{i=1}^N [\cos(\omega_i t) + S_{0i}(t)] - \frac{m^3}{6} \left(\sum_{i=1}^N [\cos(\omega_i t) + S_{0i}(t)] \right)^3 \right\} \\ + \sin(\Phi_T) \cdot \left\{ 1 - \frac{m^2}{2} \left(\sum_{i=1}^N [\cos(\omega_i t) + S_{0i}(t)] \right)^2 \right\}. \quad (5.15)$$

By note the equation (5.15), we can say, the term containing the parameter m is the linear desirable term, whereas the term containing m^2 is the second order intermodulation (IMD₂) component. Apparently once Φ_T is set to zero, the (IMD₂) vanishes while the linear and (IMD₃) terms are independent of the bias point Φ_T . In this way the output of the modulator is not containing the (IMD₂) and the equation (5.15) can be rewritten as

$$U = 1 + m \sum_{i=1}^N [\cos(\omega_i t) + S_{0i}(t)] - \frac{m^3}{6} \left(\sum_{i=1}^N [\cos(\omega_i t) + S_{0i}(t)] \right)^3. \quad (5.16)$$

Temperature variation may cause the intrinsic bias to change a small fraction of V_π . Such drifts are generally show, with time constants of minute even hours. However, a tracking (dc) voltage V_B applied to the bias electrode can insure nulling of the total phase (Φ_T) if

$$V_B(t) = -\frac{\varphi_b(t) V_\pi}{\pi}. \quad (5.17)$$

This property is utilized in the realization of a parametric feedback control system for IMD_2 suppression. While biasing the modulator at the (Φ_T) points nulls the second order intermodulation (IMD_2) contribution, the bias point has no effect on the third order intermodulation (IMD_3) which must be compensated by other means.

6. POWER SPECTRAL DENSITY OF THE MODULATION SIGNALS AND MINIMIZED IMD_3 EFFECTS

Usually very useful, and is very important tracking the signals with point power spectral density. In this section we analyze power spectral density (PSD) of the subcarrier multiplexing radio frequency (RF SCM) signal and subcarrier multiplexing amplitude modulation vestigial sideband (SCM AM/VSB) signals for CATV CD transmission system.

If the carrier video signal is $(A \cos \omega_i t)$, one side carrier spectrum $x_c(\omega)$ is expressed as

$$x_c(\omega) = A \pi \delta_0(\omega - \omega_i), \quad (6.1)$$

where A is constant.

Assuming spectrum of the TV video signal to be expressed as

$$\mathcal{F}[S_{0i}(t) = S_{0i}(\omega) = \frac{a}{\sigma_i \sqrt{2\pi}} e^{-(\omega - \omega_{0i})^2 / 2\sigma_i^2} \quad (6.2)$$

where a is constant, σ_i is the half signal bandwidth ($B/2$) and ω_{0i} is the center band-pass frequency, one channel CATV AM/VSB signal can be expressed as

$$x(t) = A \cos(\omega_1 t) + S_{01}(t) \quad (6.3)$$

The first term in the (6.3) is the picture carrier, and the second term is the video signal.

Signal spectrum of the one channel AM/VSB CATV is given by Fourier transformation and can be expressed as

$$\mathcal{F}[x(t)] = A \pi \delta_0(\omega - \omega_1) + \frac{a}{\sigma_1 \sqrt{2\pi}} e^{-(\omega - \omega_{01})^2 / 2\sigma_1^2} \quad (6.4)$$

Thus, let us assume, RF SCM waveform characteristic can be expressed as

$$X_c(t) = \sum_{i=1}^N A_i \cos(\omega_i t + \varphi_i), \quad (6.5)$$

where N is the number of channels, A_i is the amplitude of the RF SCM signal, and for simplicity we assume every channel to have the same amplitude ($A_i = A$), ω_i falls within the band of the i th channel, φ_i are the phases in the various channels that can be assumed to be zero.

So, $X(t)$ is the summation of a large number of SCM video signals with random phases. $X(t)$ can be assumed to be a stationary Gaussian process. The diagram power spectral density of $X(t)$ is shown in Fig. 9, and the corresponding spectrum can be expressed as

$$X(\omega) = A\pi \sum_{i=1}^N \delta_0(\omega - \omega_i) + a \sum_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} e^{-(\omega - \omega_i)^2 / 2\sigma_i^2}. \quad (6.6)$$

Assume the external amplitude modulator has bias point (Φ_T) equal to zero (which is defined as the intrinsic and the applied DC phase biases), which means IMD_2 term vanishes while the linear and IMD_3 terms are maximized and the ratio of the linear and IMD_3 terms is independent of the bias point (Φ_T). On the another side we use linearizer circuit to reduce the negative effect of the third order intermodulation distortion. If we take these consideration the output of the AM/VSB (for one channel) will be

$$U_0(t) = 1 + m[\cos(\omega_1 t) + S_{01}(t)] + \frac{m^3 \delta}{6} [\cos(\omega_1 t) + S_{01}(t)]^3, \quad (6.7)$$

where δ is a constant, which determines compensation of the nonlinearization.

Generally, our system uses subcarrier multiplexing amplitude modulation vestigial sideband (SCM AM/VSB) signals (with linearized circuit), thus the output of the external modulator can be expressed as

$$U(t) = 1 + m \sum_{i=1}^N [A \cos(\omega_i t) + S_{0i}(t)] + \frac{m^3 \delta}{6} \sum_{i=1}^N [A \cos(\omega_i t) + S_{0i}(t)]^3. \quad (6.8)$$

To simplify, let us express the equation (6.8) as

$$U(t) = 1 + U_1(t) + U_2(t), \quad (6.9)$$

where $U_1(t)$ and $U_2(t)$ can be expressed as

$$U_1(t) = \sum_{i=1}^N [A \cos \omega_i t + S_{0i}(t)] \quad (6.10)$$

$$U_2(t) = \frac{m^3 \delta}{6} \left[\sum_{i=1}^N \cos(\omega_i t) + S_{0i}(t) \right]^3. \quad (6.11)$$

We will be using Fourier transformations for estimating the power spectral density of subcarrier multiplexing amplitude modulation vestigial sideband cable television (PSD sCM AM/VSB CATV) signals. Therefore, the Fourier transformation of the equation (6.9) becomes

$$U(\omega) = 2\pi\delta_0(\omega) + U_1(\omega) + U_2(\omega), \quad (6.12)$$

where the Fourier transformation of 1 is the $2\pi\delta_0(\omega)$, $U_1(\omega)$ is the Fourier transformation of $U_1(t)$ and $U_2(\omega)$ is the Fourier transformation of $U_2(t)$.

For frequencies analyze of SCM AM/VSB video signals expressed in the equation (6.9), let us take the second term which express as

$$U_1(t) = mA \sum_{i=1}^N \cos(\omega_i t) + ma \sum_{i=1}^N S_{0i}(t). \quad (6.13)$$

To estimate the power spectral density of the useful signal, we take Fourier transformation of the equation (6.13). Considering the equations (6.4) and (6.5), the desirable spectrum can be expressed as

$$\mathcal{F}[U_1(t)] = U_1(\omega) \quad (6.14)$$

$$\Rightarrow U_1(\omega) = mA\pi \sum_{i=1}^N \delta_0(\omega - \omega_i) + ma \sum_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} e^{-(\omega - \omega_{0i})^2 / 2\sigma_i^2}. \quad (6.15)$$

Equation (6.15) shows useful term in the spectrum that consists the transmitter information, which correspond to the useful part in the input signal to the receiver.

Now, let us estimate the power spectral density of the third order IMD signal represented by the third term in the equation (6.8). The identifies of the third power can expresses as [14, 15]

$$\left(\sum_{i=1}^N U_i \right)^3 = \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N U_i U_j U_k. \quad (6.16)$$

Thus, the third term in (6.8) can be expressed as follows

$$U_2(t) = a_0 \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N [\cos(\omega_i t) + S_{0i}(t)] [\cos(\omega_j t) + S_{0j}(t)] [\cos(\omega_k t) + S_{0k}(t)], \quad (6.17)$$

where

$$a_0 = \frac{m^3 \delta}{6}. \quad (6.18)$$

Therefore, equation (6.17) can be rewritten as

$$\begin{aligned} U_2(t) = a_0 \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \{ & \cos \omega_i t \cos \omega_j t \cos \omega_k t + S_{0i}(t) S_{0j}(t) S_{0k}(t) \\ & + S_{0k}(t) \cos \omega_i t \cos \omega_j t + S_{0j}(t) \cos \omega_i t \cos \omega_k t \\ & + S_{0i}(t) \cos \omega_j t \cos \omega_k t + S_{0j}(t) S_{0k}(t) \cos \omega_i t \\ & + S_{0j}(t) S_{0k}(t) \cos \omega_j t + S_{0i}(t) S_{0j}(t) \cos \omega_k t \}. \end{aligned} \quad (6.19)$$

Assume $U_3(t)$, $U_4(t)$, $U_5(t)$, $U_6(t)$, $U_7(t)$, $U_8(t)$, $U_9(t)$ and $U_{10}(t)$ are the terms of the equation (6.19), which can be expressed as

$$U_3(t) = \cos \omega_i t \cos \omega_j t \cos \omega_k t \quad (6.20)$$

$$U_4(t) = S_{0i}(t) S_{0j}(t) S_{0k}(t) \quad (6.21)$$

$$U_5(t) = S_{0k}(t) \cos \omega_i \cos \omega_j t \quad (6.22)$$

$$U_6(t) = S_{0j}(t) \cos \omega_i t \cos \omega_k t \quad (6.23)$$

$$U_7(t) = S_{0i}(t) \cos \omega_j t \cos \omega_k t \quad (6.24)$$

$$U_8(t) = S_{0j}(t) S_{0k}(t) \cos \omega_i t \quad (6.25)$$

$$U_9(t) = S_{0i}(t) S_{0k}(t) \cos \omega_j t \quad (6.26)$$

$$U_{10}(t) = S_{0i}(t) S_{0j}(t) \cos \omega_k t. \quad (6.27)$$

According to the above equations, we can analyze the equation (6.20) as follows

$$\begin{aligned} U_3(t) &= \cos(\omega_i t) \cos(\omega_j t) \cos(\omega_k t) \\ &= \frac{1}{4} [\cos(\omega_i + \omega_j + \omega_k) t + \cos(\omega_i - \omega_j - \omega_k) t]. \end{aligned} \quad (6.28)$$

Taking the Fourier transformation of the equation (6.28), we have

$$U_3(\omega) = \frac{\pi}{4} \delta_0 [\omega + (\omega_i + \omega_j + \omega_k)] + \frac{\pi}{4} \delta_0 [(\omega + \omega_i + \omega_j + \omega_k)]. \quad (6.29)$$

Applying the manoeuvrability convolution function [11], which is expressed as

$$\mathcal{F}[X_0(t) \cdot Y_0(t)] = \frac{1}{2\pi} [X_0(\omega) * Y_0(\omega)] \quad (6.30)$$

the Fourier transformation of the equation (6.21) becomes

$$\mathcal{F}[S_{0i}(t) S_{0j}(t) S_{0k}(t)] = S_{0i}(\omega) * S_{0j}(\omega) * S_{0k}(\omega) \quad (6.31)$$

$$U_4(\omega) = S_{0i, 0j, 0k}(\omega) = \left(\frac{1}{2\pi} \right) \frac{a^3}{\sigma \sqrt{6\pi}} e^{-\frac{[\omega - (\omega_{0i} + \omega_{0j} + \omega_{0k})]^2}{6\sigma^2}} \quad (6.32)$$

$$\frac{B}{2} = \sigma_i = \sigma_j = \sigma_k, \quad (6.33)$$

where the parameters ω_i , ω_j , ω_k , ω_{0i} , ω_{0j} and ω_{0k} are shown in Fig. 9.

Now, let us analyze the term $U_5(t)$ which can be expressed as

$$U_5(t) = \frac{1}{2} S_{0k}(t) \{ \cos[\omega + (\omega_i + \omega_j)] t + \cos[\omega + (\omega_i + \omega_j)] t \}. \quad (6.34)$$

Applying the manoeuvrability delta function [11], which is expressed as

$$\mathcal{F}[x_0(t) \cdot \cos \omega_0 t] = \frac{1}{2} [X_0(\omega - \omega_0) + X_0(\omega + \omega_0)] \quad (6.35)$$

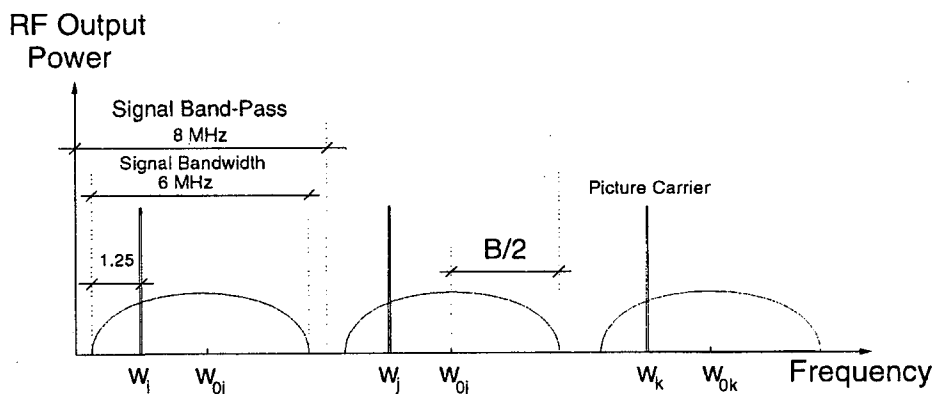


Fig. 9 diagram power spectral density of the AM/VSB video signals

and consider (6.30), the Fourier transformation of (6.34) gives

$$U_5(\omega) = \frac{1}{4} S_{0k}[\omega + (\omega_i + \omega_j)] + \frac{1}{4} S_{0k}[\omega + (\omega_i - \omega_j)]. \quad (6.36)$$

In the same manner, we analyze the terms $U_6(t)$ and $U_7(t)$. Thus, the Fourier transformation to these terms can be expressed as

$$U_6(\omega) = \frac{1}{4} S_{0j}[\omega + (\omega_i + \omega_k)] + \frac{1}{4} S_{0j}[\omega + (\omega_i - \omega_k)] \quad (6.37)$$

$$U_7(\omega) = \frac{1}{4} S_{0i}[\omega + (\omega_j + \omega_k)] + \frac{1}{4} S_{0i}[\omega + (\omega_j - \omega_k)]. \quad (6.38)$$

Considering the equations (6.31), (2.36), and taking the Fourier transformation of equations (6.25), (6.26) and (6.27) we obtain on

$$U_8(\omega) = \frac{1}{2} S_{0j,0k}(\omega + \omega_i) \quad (6.39)$$

$$U_9(\omega) = \frac{1}{2} S_{0i,0k}(\omega + \omega_j) \quad (6.40)$$

$$U_{10}(\omega) = \frac{1}{2} S_{0i,0j}(\omega + \omega_k), \quad (6.41)$$

where

$$S_{0j,0k}(\omega) = S_{0j}(\omega) * S_{0k}(\omega) = \mathcal{F} [S_{0j}(t) \cdot S_{0k}(t)] \quad (6.42)$$

$$S_{0j,0k}(\omega) = \frac{a^2}{2\sigma\sqrt{\pi}} e^{-\frac{[\omega - (\omega_{0j} + \omega_{0k})]^2}{4\sigma^2}} \quad (6.43)$$

$$S_{0i,0k}(\omega) = \frac{a^2}{2\sigma\sqrt{\pi}} e^{-\frac{[\omega - (\omega_{0i} + \omega_{0k})]^2}{4\sigma^2}} \quad (6.44)$$

$$S_{0j,0j}(\omega) = \frac{a^2}{2\sigma\sqrt{\pi}} e^{-\frac{[\omega - (\omega_{0i} + \omega_{0j})]^2}{4\sigma^2}} \quad (6.45)$$

Note: the derivation of (6.32), (6.43), (6.44) and (6.45) is not shown because it goes beyond the scope of this paper.

Finally, substituting the equations (6.15), (6.39), (6.32), (6.34), (6.37), (6.38), (6.39), (6.40), and (6.41) in the equation (6.12) yields

$$\begin{aligned} U(\omega) = & 2\pi\delta_0(\omega) + A\pi m \sum_{i=1}^N \delta_0(\omega - \omega_i) + am \sum_{i=1}^N \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\omega - \omega_{0i})^2}{2\sigma^2}} \\ & + \frac{m^3\delta}{6} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \left\{ \left(\frac{1}{2\pi} \right) \frac{a^3}{\sigma\sqrt{6\pi}} e^{-\frac{[\omega - (\omega_{0i} + \omega_{0j} + \omega_{0k})]^2}{6\sigma^2}} \right. \\ & + \frac{\pi}{4} \delta_0[\omega + (\omega_i + \omega_j + \omega_k)] + \frac{\pi}{4} \delta_0[\omega + (\omega_i - \omega_j - \omega_k)] \\ & + \left(\frac{1}{2} \right) \frac{a^2 A}{2\sigma\sqrt{\pi}} e^{-\frac{[\omega + \omega_i - (\omega_{0j} + \omega_{0k})]^2}{4\sigma^2}} \\ & + \frac{1}{2} \left[\frac{a^2 A}{2\sigma\sqrt{\pi}} e^{-\frac{[\omega + \omega_j - (\omega_{0i} + \omega_{0k})]^2}{4\sigma^2}} + \frac{a^2 A}{2\sigma\sqrt{\pi}} e^{-\frac{[\omega + \omega_k - (\omega_{0i} + \omega_{0j})]^2}{4\sigma^2}} \right] \\ & + \frac{1}{4} \left[\frac{a A^2}{\sigma\sqrt{2\pi}} e^{-\frac{[\omega - \omega_{0k} + (\omega_i + \omega_j)]^2}{2\sigma^2}} + \frac{a A^2}{\sigma\sqrt{2\pi}} e^{-\frac{[\omega - \omega_{0k} + (\omega_i - \omega_j)]^2}{2\sigma^2}} \right] \\ & + \frac{1}{4} \left[\frac{a A^2}{\sigma\sqrt{2\pi}} e^{-\frac{[\omega - \omega_{0j} + (\omega_i + \omega_k)]^2}{2\sigma^2}} + \frac{a A^2}{\sigma\sqrt{2\pi}} e^{-\frac{[\omega - \omega_{0j} + (\omega_i - \omega_k)]^2}{2\sigma^2}} \right] \\ & \left. + \frac{1}{4} \left[\frac{a A^2}{\sigma\sqrt{2\pi}} e^{-\frac{[\omega - \omega_{0i} + (\omega_j + \omega_k)]^2}{2\sigma^2}} + \frac{a A^2}{\sigma\sqrt{2\pi}} e^{-\frac{[\omega - \omega_{0i} + (\omega_j - \omega_k)]^2}{2\sigma^2}} \right] \right\}. \quad (6.46) \end{aligned}$$

Thus, spectrum of the output external modulator signals (SCM AM/VSB video signals) can be expressed as

$$R(\omega) = |U(\omega)|. \quad (6.47)$$

finally, power spectral density PSD of the SCM AM/VSB video signals becomes

$$R_0(\omega) = 10 \log [|U(\omega)|^2]. \quad (6.48)$$

Analysis the equation (6.46), we can say, the first term is the principle carrier, the terms in (m) is the linear desirable term that contains the information about video

signals, whereas the term in the term in (m^3) is the third order intermodulation (IMD_3) component. Apparently, IMD_2 vanishes while IMD_3 is independent of the bias point but dependent on δ parameter that limits compensate of the nonlinearizer modulator. Therefore, we can use linearizer circuit within AM/VSB optical modulator to minimize the IMD_3 component, which improves the general performance of the SCM AM/VSB CATV CD system.

Therefore, power spectral density consists of very large number of components with different frequencies. The amplitudes of these components become smaller and smaller with increasing frequency. For practice, it is enough to consider limited bandwidth frequency for system communication. This will economize the bandwidth, thus enables the transmission of the several channels over a limited total system bandwidth. Good knowledge to PSD helps on transmission and reception communication signals (information, news, picture) in a better way and economically matched.

7. DISCUSSIONS AND RESULTS

The required channel spacing, for coherent SCM AM/VSB CATV system depends on the IF signal, modulation form, and bandpass filter. Large channel spacing is necessary to avoid interference from the image spectrum of the adjacent channels. When we use a single optical carrier, there is no difference between the channel spacing in the optical and electrical domain, this means that values nearly $D = 2B_n$ is sufficient, where B_n is the RF signal frequency bandwidth. As a matter-of-fact, the crosstalk noise variance of channels, from (2) to (N-1) twice the variance noise crosstalk results from the neighbourhood channels. On the other hand, the crosstalk noise variance of channel 1 and N is equal to the variance crosstalk noise.

SCM systems are degraded by IMD, this is true for both direct detection and coherent systems. Intermodulation distortion results primarily from the nonlinear laser characteristic, with coherent SCM CATV systems, the IMD is a fundamental characteristics of the nonlinear modulation format and the coherent detection process.

For the design purpose we expressed the definitions of the system parameters as listed in Table.

Table
Definition and numerical values of system parameters

<i>Fiber attenuation</i>	α	<i>0,2 dB</i>
<i>Excess loss of 3 dB coupler</i>	δ_L	<i>0,05</i>
<i>Photodiode responsivity</i>	R	$q/h\nu = 1,25 \text{ A/W}$
<i>Charge electron</i>	q	$1.602 \times 10^{-19} \text{ C}$
<i>Temperature</i>	T	<i>300 K</i>

<i>Load resistance</i>	R_L	$50 [\Omega]$
<i>Bandwidth of the video TV signal</i>	B_S	$6 \times 10^6 \text{ Hz}$
<i>Power LO laser</i>	P_{LO}	$10 \times 10^{-3} \text{ W}$
<i>Local oscillator relative intensity noise</i>	RIN_{LO}	$-110 \div 160 \text{ dB/Hz}$
<i>Relative intensity noise of the transmitter laser</i>	RIN_s	$-110 \div 160 \text{ dB/Hz}$
<i>Photon energy</i>	$h\nu$	$1,28 \times 10^{-19} \text{ J}$
<i>Carrier-to-noise ratio</i>	CNR	$40 \div 46$
<i>Signal-to-noise ratio</i>	SNR	56
<i>Noise figure of the optical amplifier system</i>	F	3 dB
<i>Gain of the optical amplifier</i>	G	15
<i>Boltzmann constant</i>	K_B	$1,38 \times 10^{-23} \text{ J/C}$

We have been concerned only by the AM/VSB optical IMD that is intrinsic to coherent SCM-CATV optical system. IMD results from the electrical amplification of the composite microwave signal, this electrical IMD can make same degradation, but in some condition, the electrical IMD is negligible. We assume the system suffer from IMD AM/VSB optical modulator and IMD proportional to another means and coherent detection process is negligible.

We noted the modulation index (m) must be selected in better way and must not be increased more than (2%), because if the modulation index is large ($m > 2\%$) third order intermodulation distortion products will dominating. The expression for the power density spectrum in (6.46) is purposely separated into two terms to emphasize the two different types of spectral components.

The first term is the desirable PSD and its shape depends only on the spectral characteristic of the SCM AM/VSB video signals $U_1(\omega)$. Fig. 10 shows the power spectral of a AM/VSB modulated 10 lines, the bandwidth of the TV signal is 5 MHz for PAL system (or 6 MHz for SECAM system), and the lower edge of the bandwidth is 1.25 MHz below the picture carrier.

The second term in (6.46) is the undesirable $IMD_3 [U_2(\omega)]$ and that is clearly in the Fig. 11. Third order intermodulation products (IMP's) has been shown that products of the form $(f_i \pm f_j \pm f_k)$, where f is the frequency of the channel and has varying value for every channel i, j, k . As matter-of-fact, a third order IMP's that falls directly within the frequency band of ith channel in the N-channel coherent SCM CATV AM/VSB.

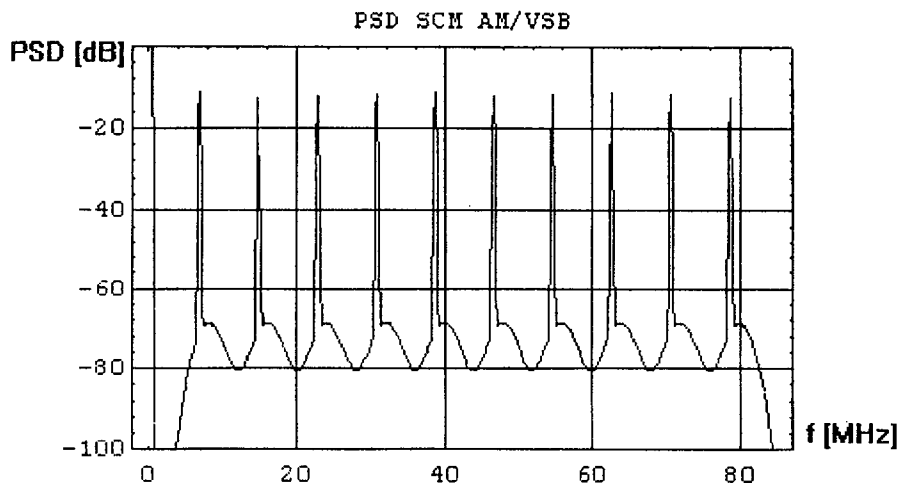


Fig. 10 power spectrum density of the 10-channels AM/VSB television signals, where $m = 0.02$, $\delta = 1$, signal bandwidth is 5 MHz (PAL systems) and channels band-pass 8 MHz

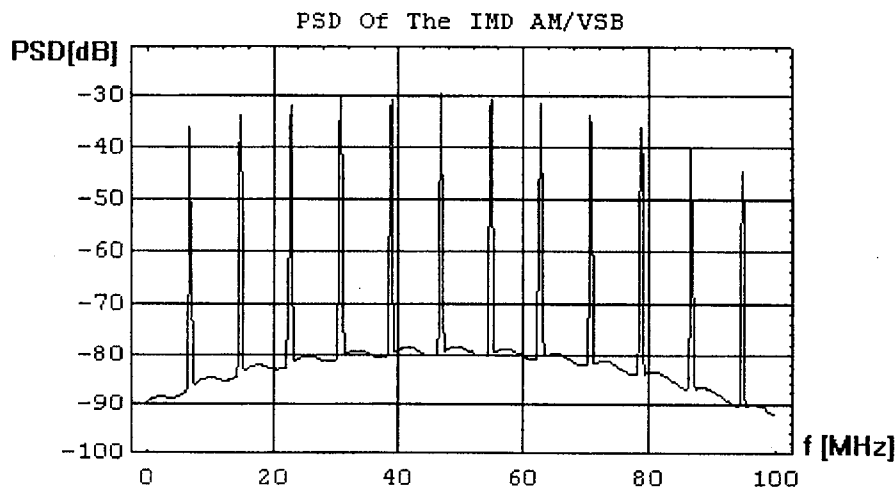


Fig. 11 power spectrum density of the third order intermodulation distortion products in external AM/VSB, where $N = 10$, $m = 0.02$, $\delta = 1$, signal bandwidth is 5 MHz channel band-pass 8 MHz

Fig. 12 shows clearly the IMD_2 power spectral components vanish when the bias point Φ_T of the optical AM/VSB equal to zero and this condition is usually desirable for the optical AM/VSB techniques for CATV CD system. Where IMD_3 takes the maximum value and set of on the center PSD of the group SCM AM/VSB video signals. Therefore, the center channel ($j = N/2$) has the most IMP's and the last channel ($j = N$) has the least IMP's. As it is seen from PSD of the pure video signals

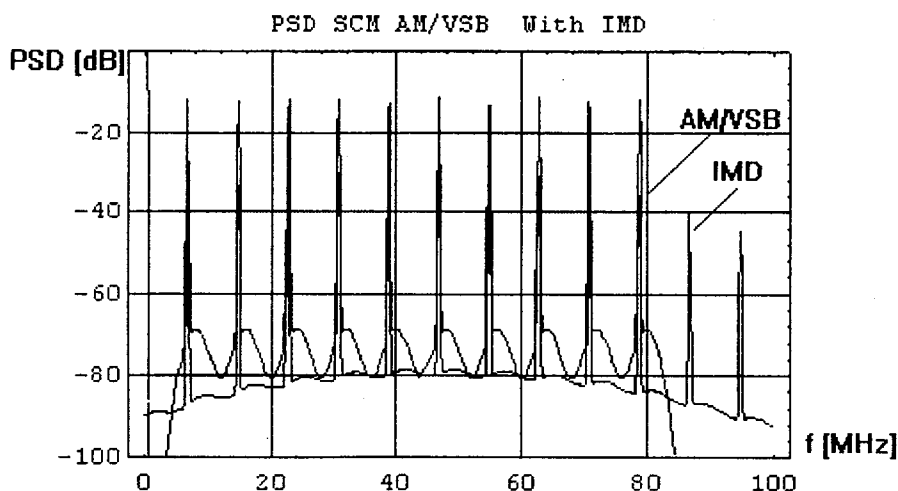


Fig. 12 power spectral density of the third order intermodulation distortion set of on the center power spectral density SCM AM/VSB video signals, where $N = 10$, $m = 0.02$, $\delta = 1$, signal bandwidth is 5 MHz and channel band-pass 8 MHz

and PSD IMD, the convoluted version of the individual signal spectra causes a wider bandwidth as shown in the Fig. 13.

Let us say after that, SCM CATV AM/VSB CD system is degraded by IMD, this is true for both direct detection and coherent detection systems. IMD results primarily from the nonlinear laser characteristics with coherent SCM CATV

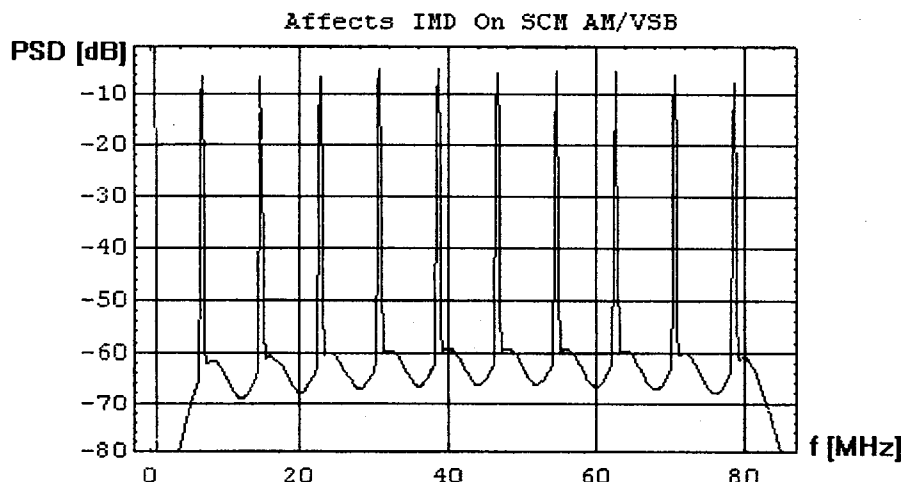


Fig. 13 shows the third order intermodulation negative effects on the power spectral density SCM AM/VSB video signal, where $N = 10$, $m = 0.02$, $\delta = 1$, signal bandwidth is 5 MHz and channel band-pass 8 MHz

AM/VSB systems the IMD is a fundamental characteristics of the nonlinear modulation format. The worst case IMD product near the center channel, is the sum of intermodulation products that involve all the channels. Since the channels near the midband contribute mostly to the worst case intermodulation third order products, it is particularly important to have uniform frequency and phase response near the midband.

As we saw, it is apparent, the (IMD_2) vanishes while the linear and (IMD_3) terms are independent of the bias point (Φ_T) and fortunately depends on another parameter (δ) which limits compensation of the unlinearizer modulator. Therefore, we can use linearizer circuit within (AM/VSB) optical modulator to minimized the IMD_3 component, which improves the general performance of the SCM AM/VSB CATV CD system.

So, comparatively with the unlinearizer modulator case $(\delta = 1)$ when δ equal to 0.25, 0.5 and 0.75, PSD of the IMD_3 reduced by approximately 25, 14, and 5 dB, respectively. Thus, by minimizing δ parameter the IMD_3 effects can be reduced. Let us then say, by proper selection of the δ the PSD of the IMD_3 is reduced and Fig. 14 shows power spectral density of the intermodulation distortions versus δ (delta). Thus, general system performance will become better.

When the video band-pass frequency SCM video signal increases from the first channel to the center group SCM signals by linear form, and decrease from the center group SCM channels in the same way to the last channels in the group SCM video signals, the IMD_3 can be reduced. That is mean, the band-pass frequency SCM video signal can be expressed as

$$B_{0i} = \begin{cases} B_0 + \sum_{i=1}^{N/2} \Delta B_0(i-1) & \text{for } 1 \leq i \leq N/2 \\ B_{N/2} - \sum_{i=\frac{N}{2}+1}^N \Delta B_0(i-1) & \text{for } \frac{N}{2} + 1 \leq i \leq N, \end{cases} \quad (7.1)$$

where B_0 is the nominal band-pass frequency of the video signal (8 MHz), ΔB_0 is the stop increase in the channel band-pass frequency, $B_{N/2}$ is the central channel band-pass, and i is the channel number.

Thus, if the PSD bandwidth of the PAL signal is 5 MHz (or 6 MHz for SECAM system) for each channel, first and last band-pass video signals have 8 MHz and center band-pass signal has 12.4 MHz where ΔB_0 is 1.1 MHz as we see in the Fig. 17. We noted here clearly PSD of the IMD_3 reduced more than 30 dB and appeared that in Fig. 18. Therefore, system performance improves as we expected.

When the system works on one octave as is clear in Fig. 19, where $N = 10$, $\delta = 1$, $m = 0.02$, and nominal channel band-pass are 8 MHz. We observed, the maximum IMD_3 falling within the channel's bandwidth. Therefore, the IMD_3 negative affects not reduced by making the system works on one octave. PSD of the IMD_3 negative effects on the PSD 10-channels AM/VSB video signals is reduced. Again, let us say after that, in this way the system performance may be not improved.

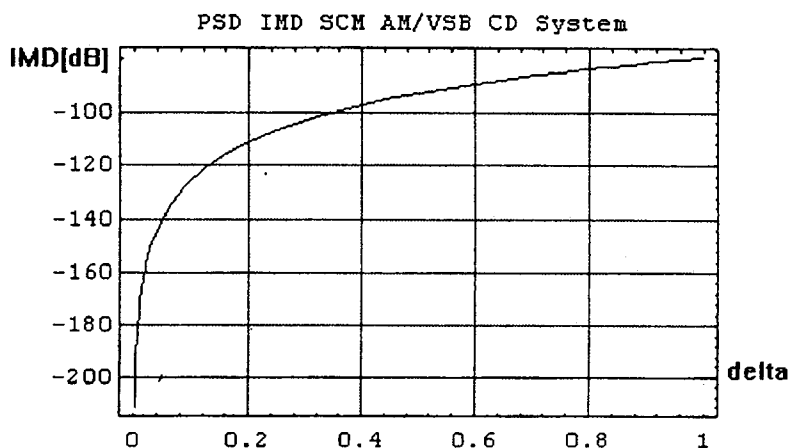


Fig. 14 intermodulation distortions power spectral density of the center channel versus δ (delta)

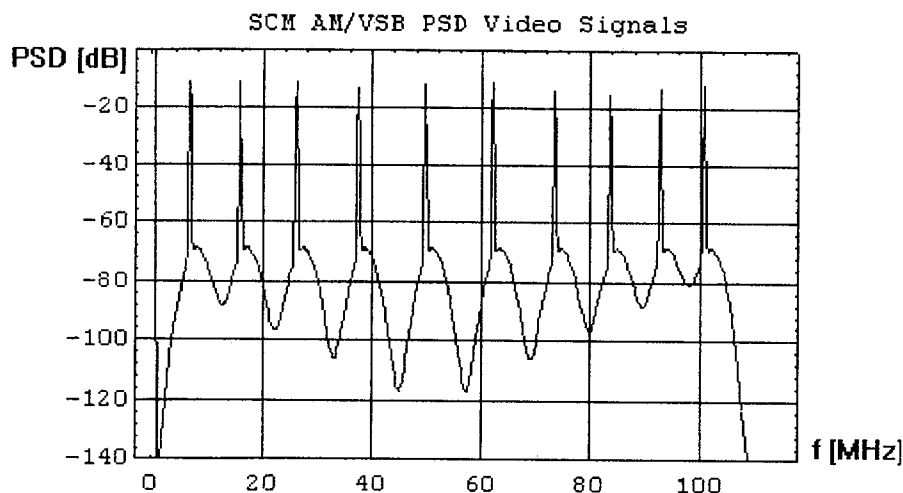


Fig. 15 Power spectral density of the 10 video channels, where $N = 10$, $m = 0.02$, $\delta = 1$, central channel band-pass 12.4 MHz, first and last band-pass signals are 8 MHz and ΔB_0 is 1.1 MHz, band-pass signals increase from the first to the center and decrease from the center to the last in the SCM AM/VSB video signals

Under all these considerations, the system designer can control the spectrum characteristics and PSD of the SCM modulated signal. Thus, by proper selection of the band-pass frequency characteristics to transmitted PSD video signals the

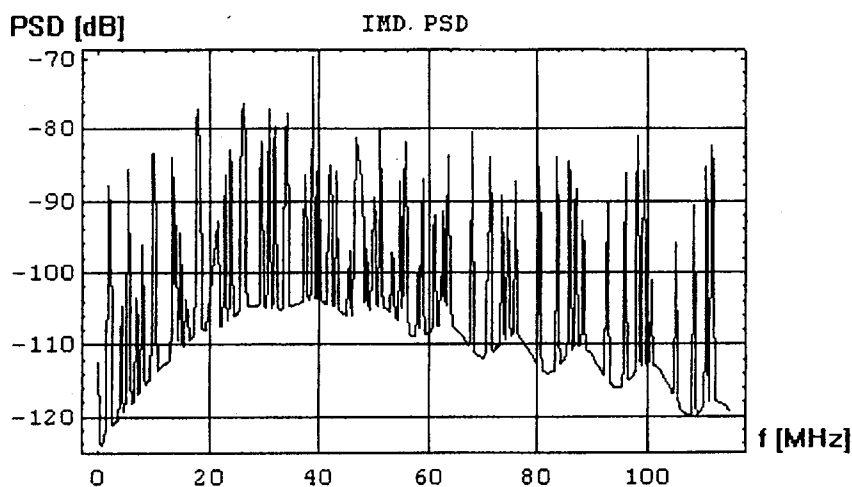


Fig. 16. Power spectral density of the third order intermodulation distortion, where $N = 10$, $m = 0.02$, $\delta = 1$, signal bandwidth is 5 MHz for PAL system, central channel band-pass 12.4 MHz, first and last band-pass signals are 8 MHz, band-pass signal increases from the first to the center and decrease from the center to the last in the SCM AM/VSB video signals

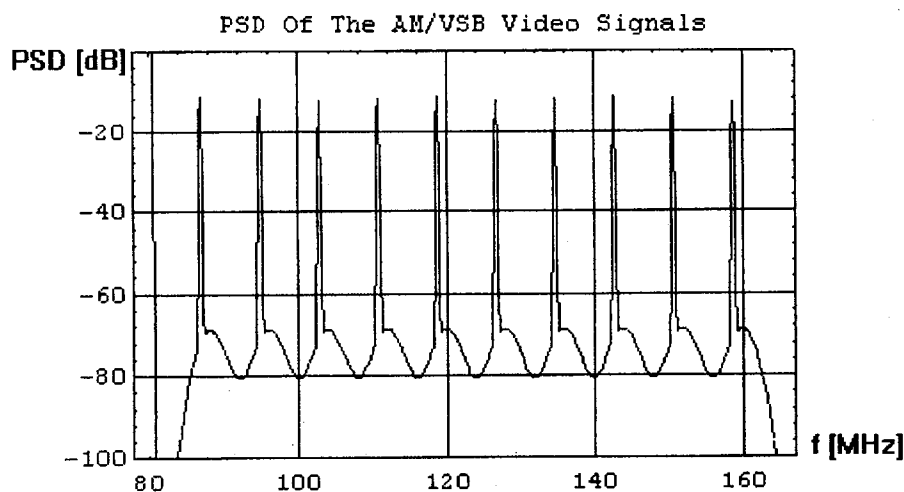


Fig. 17. Power spectrum density of the 10-channels AM/VSB video signals, when system works on one octave and where $m = 0.02$, $\delta = 1$, signal bandwidth is 5 MHz and channel band-pass 8 MHz

performance system must be increased. In addition, the SCM AM/VSB CATV CD system performance could be significantly improved by using a balanced receiver, higher power laser, linear electrooptic modulator (EOM), and high gain amplifiers.

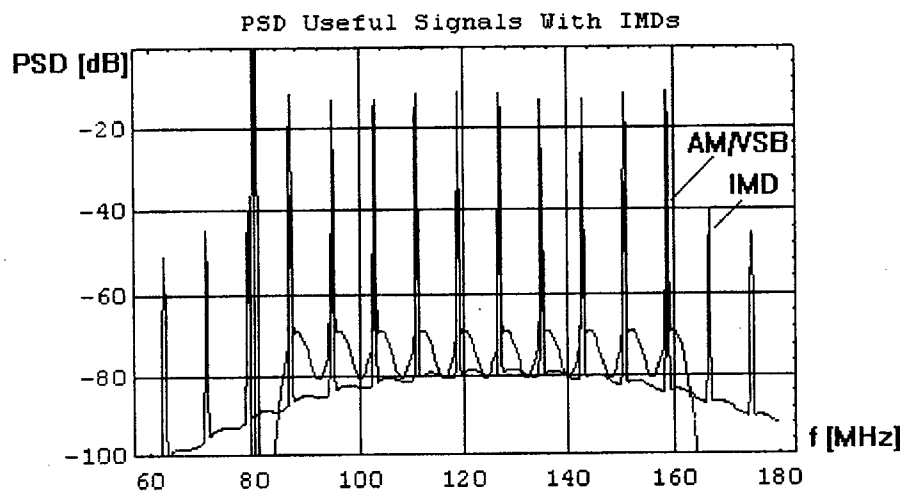


Fig. 18. PSD IMD, with 10-channels AM/VSB video signals, when system works on one octave, where $m = 0.02$, $\delta = 1$, signal bandwidth is 5 MHz (PAL system) and band-pass frequency 8 MHz

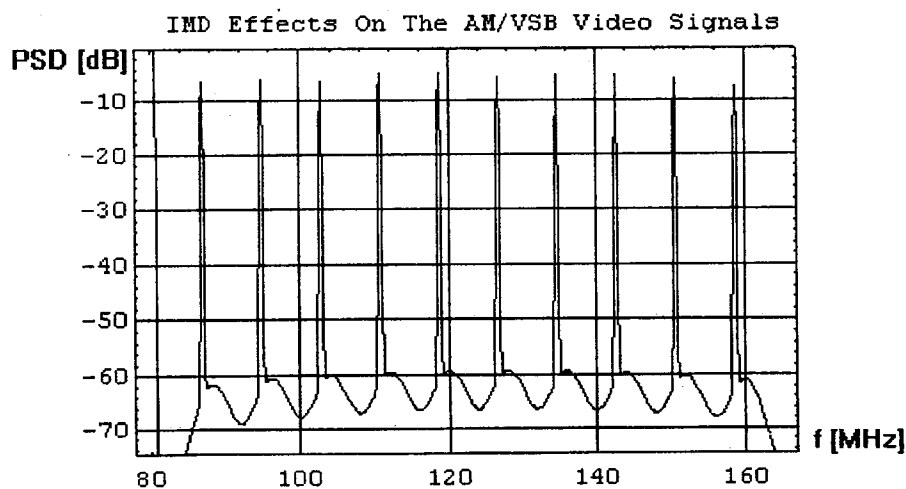


Fig. 19. Third order intermodulation distortion negatives effects on the PSD 10-channels AM/VSB video signals, when system works on one octave and where $m = 0.02$, $\delta = 1$, channel band-pass 8 MHz and signal bandwidth is 5 MHz

REFERENCES

1. T.K. Fong, D. Jay, M. Sabido IX, L.G. Kazovsky: *Linewidth-Insensitive Coherent AM Analog Optical Links Using Semiconductor Lasers*. IEEE Photonics Technology Letters, Vol. 4, No. 4, April 1993
2. Moshe Nazarathy, Josef Berger, Anthony J. Ley, Israel M. Levi, Yishai Kagan: *Progress in Externally Modulated AM CATV Transmission Systems*. Journal of Lightwave Technology, Vol. 11, No. 1, January 1993
3. R.B. Childs, V.A. O'Byrns: *Multichannel AM video Transmission Using High-Power Nd:YAG Lasers and Linearized External Modulator*. IEEE Journal On Selected Areas In Communications, Vol. 8, No. 7, September 1990
4. Ligeng Xu, R. Taylor, S.R. Forrest: *The Use of Optically Coherent Detection Techniques for True-Time Delay Phased Array and Systems*. Journal of Lightwave Technology, Vol. 13, No. 8, August 1995
5. L.M. Johnson, H.V. Roussell: *Linearization of an Interferometric Modulator at Microwave Frequency By Polarization Mixing*. IEEE Photonics Technology Letters, Vol. 2, No. 11, November 1990
6. J.L. Brooks, G.S. Maurer, R.A. Becker: *Implementation and Evaluation of a Dual Parallel Linearization System For AM-SCM Video Transmission*. Journal of Lightwave Technology, Vol. 11, No. 1, January 1993
7. M.S. Makhoul: *Optical frequency modulation for coherent subcarrier multiplexing transmission*. Kwart. Elektr. i Telekom. 1996. 42, z. 2
8. D. Jay, M. Sabido IX, Masafumi Tabara, Thomas K. Fong, Chung-Li Lu, L.G. Kazovsky (Fellow, IEEE): *Improving the Dynamic Range of a Coherent AM Analog Optical Link Using a Cascaded Linearized Modulator*. IEEE Photonics Technology Letters, Vol. 7, No. 7, July 1995
9. R.F. Kalman, C. Fan, L.G. Kazovsky: *Dynamic Range of Coherent analog Fiber-Optic Links*. Journal of Lightwave Technology, Vol. 12, No. 7, July 1994
10. M.S. Makhoul: *Using M-ary Digital Modulation For Coherent Optical Subcarrier Multiplexing System*. Kwart. Elektr. i Telekom. 1997. 43, z. 2
11. J. Szabatin: *Podstawy Teorii Sygnałów*. Warszawa, 1990
12. B.S. Glance, J. Stone, K.J. Pollock, P.J. Fitzgerald, C.A. Burrus, JR., B.L. Kasper, L.W. Stulz: *Densely Spaced FDM Coherent Star Network With Optical Signals Confined to Equally Spaced Frequencies*. Journal of Lightwave Technology, Vol. 6, No. 11, November 1988
13. W. Richard, Gross, W. Rodeout, R. Olshansky, G.R. Joyce: *Heterodyne Video Distribution Systems Sharing Transmitter and Local oscillator Lasers*. Journal of Lightwave Technology, Vol. 9, No. 4, April 1991
14. Delfin Jay M. Sabido IX, Thomas K. Fong, R.F. Kalman, L.G. Kazovsky: *Linewidth-Insensitive Coherent Optical analog Links*. 504/SPIE Vol. 1703, 1992
15. T.K. Fong, D. Jay M. Sabido IX, R.F. Kalman, Masafumi Tabara, L.G. Kazovsky: *Linewidth-Insensitive Coherent AM Optical Links: Design, Performance, and Potential Applications*. Journal Of Lightwave Technology, Vol. 12, No. 3, March 1994
16. I.S. Gradshteyn, I.M. Ryzhik: *Table of Integrals. Series, and Products* Fifth Edition, 1994 England
17. Attio J. Rainal: *Limiting Distortion of CATV Lasers*. Journal of Lightwave Technology, Vol. 14, No. 3, March 1996
18. F.W. Willems, J.C. van der Plats, F. Koelma: *Experimental Verification of the Distortion Caused By Long Optical Cavities in AM-VSB CATV Lightwave Systems*. Electronics Letters 4th January 1996, Vol. 32, No. 1
19. V.B. Gorfinkel, S. Luryi: *Fundamental Limits for Linearity of CATV Lasers*. Journal of Lightwave Technology, Vol. 13, No. 2, February 1995
20. A. Fiok: *Podstawy Ogólne*, Telewizja, Warszawa, 1996
21. Delfin Jay M. Sabido IX, Masafumi Tabara, T.K. Fong, R.F. Kalman, L.G. Kazovsky: *Experimental Linewidth-Insensitive Coherent Analog Optical*. Journal of Lightwave Technology, Vol. 12, No. 11, November 1994

M.S. MAKHOUL

EFEKTY ZNIEKSZTAŁCENIA ZŁOŻONEGO SYGNAŁU W PASMIE CZĘSTOTLIWOŚCI
SCM AM/VSB CATV CD

Streszczenie

Przeanalizowano zniekształcenia złożonego sygnału utworzonego drogą optycznej multipleksji falowej (WDM) modulowanego zewnętrznie za pomocą modulatora elektrooptycznego. Opisano koherentną transmisję sygnału o szerokim widmie optycznym w odniesieniu do przenoszenia obrazu w sieci telewizyjnej kablowej. Podano przegląd możliwych układów elektrooptycznej modulacji zewnętrznej z kompensacją zniekształceń nieliniowych. Są to kombinacje planarnych modulatorów elektrooptycznych w konfiguracji interferometru Macha-Zehndera wspomaganych elementami elektroniki. W powołaniu na dane literaturowe przedstawiono graficznie z krótką analizą możliwość kompensacji zniekształceń modulacji. Zaprezentowano konfigurację nadajnika i odbiornika koherentnej transmisji z nośnym sygnałem optycznym generowanym z lasera Nd:YAG wzbudzonym za pomocą diod. W nadajniku zastosowano układ zewnętrznej modulacji z kompensacją zniekształceń. Przedstawiono dwie konstrukcje odbiornika koherentnego z lokalnym oscylatorem w odniesieniu do jednego odbiornika i z dystrybucją sygnału i lokalnego oscylatora na N torów odbiorczych.

Istotna, oryginalna część pracy dotyczy opisu modulacji i wynikających zniekształceń nieliniowych, towarzyszących elektrooptycznej modulacji zewnętrznej i zdefiniowania warunków minimalizacji drugiego, lub trzeciego rzędu deformacji nieliniowej sygnału. Sygnał przedstawiono w notacji gęstości mocy spektralnej i przeanalizowano szczegółowo efekt minimalizacji trzeciego rzędu zniekształceń intermodulacyjnych sygnału multipleksowanego optycznie i modulowanego amplitudowo modulatorem zewnętrznym. Przedstawiono warianty kompensacji zniekształceń ilustracją wyników symulacji komputerowej.

Słowa kluczowe: zniekształcenia sygnału, optyczna multipleksja falowa, układy modulacyjne, laser Nd:YAG, telewizja kablowa.

Design of bit permutations for substitution-permutation networks

ALBERT SADOWSKI

Instytut Podstaw Elektroniki, Politechnika Warszawska

Received 1997.05.23

Authorized 1997.10.08

In this paper we consider the bit permutations used in the substitution-permutation networks (SPNs). We define the terms: *regular SP-network* and *irregular SP-network* and present the methods of designing these two kinds of networks. Presented methods allow to achieve so called *connective completeness* of the substitution-permutation network in as few number of rounds as possible.

Keywords: Substitution-permutation network, product cipher, permutation, completeness, complete SPN.

1. INTRODUCTION

In [MEYE78] author presented an analysis of building dependence of ciphertext bits on plaintext/key bits in successive rounds of the Data Encryption Standard algorithm. According to this results of this analysis, the total dependence of output bits on plaintext and key bits is achieved in the fifth round. However, author showed that, according to the approximate analysis, four is the minimal number of the round in which total dependence could potentially be achieved.

In this paper we present the methods of designing the permutations of the substitution-permutation networks which ensure achieving total dependence in minimal number of rounds. First we define a term *connective completeness*, which means, for the networks with S-boxes for which each output bit depends on all input bits, total dependence of all ciphertext bits on all plaintext bits and present a necessary condition of connective completeness of the SPN. We introduce the partition of the substitution-permutation networks into two classes:

- the *regular SP-networks* — the networks for which the number of the S-boxes in one round is a natural power of the number of S-box inputs (or outputs),

- the *irregular SP-networks* — the networks which are not regular.

For both these kinds of networks we present theorems which allow to determine the minimal number of the round in which achieving the connective completeness is possible. Moreover, we present the algorithms which may be used to design both these kinds of networks. Algorithm of designing the regular networks is similar to the one presented in [KADA79]. Second presented algorithm allows to design both regular and irregular networks and is more complex.

2. BACKGROUND AND FUNDAMENTAL CONCEPTS

In this paper we discuss an N -bit SPN consisting of R rounds of n S-boxes connected by bit permutations (Fig. 1). The S-boxes are $p \times p$ bijective mappings and the bit permutations belong to a special set of permutations for which no two

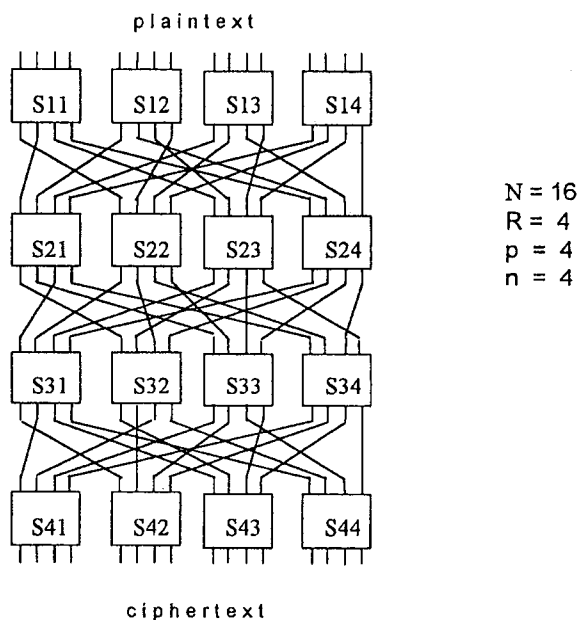


Fig. 1. A sample substitution-permutation network

outputs of an S-box are connected to one S-box in the next round. The round (or iteration) keys are in general independent on each other (or generated by a key scheduling algorithm). The keying is done by xoring the iteration keys with the inputs to the rows of the S-boxes in each round and by xoring result of the last round with the last iteration key.

Nonlinearity and diffusion

It known that the input bits of the substitution-permutation network should depend on its input bits nonlinear way. The components which provide *nonlinear dependence* of the output bits of the network on its input bits are the S-boxes. The bit permutations are linear transformations and play another role in the network.

Another requirment for the substitution-permutation network is that each its output (each output bit) depends on the greatest number of inputs (in the best case — on all inputs). If each output depends on large number of inputs then it means that there are many such inputs which influence many outputs. Above requirement means that there should be many such inputs that changing one of them may cause large number of output bits changing. To make possible changing some input bit cause large number of outputs of the network changing, the network has to be constucted this way that a change of the input bit is spread in successive rounds onto the greatest part of the network. Therefore the network should ensure a *diffusion* of the changes of its inputs. Owing to this, a small change at the inputs will be able to cause a large change at the outputs.

Both the S-boxes and the permutations are responsible for diffusion of the changes at the inputs of the network.

Participation of S-boxes in diffusion

A degree of diffusion introduced by the S-boxes is illustrated by the relationships between input and output xors of the S-box. Let $wt(v)$ represent the number of non-zero bits in vector v , or the Hamming weighth of v , dX and dY input and output xors of the S-box, respectively.

Definition 1. [HETA96] An S-box satisfies a *diffusion order* of λ , $\lambda \geq 0$, if, for $wt(dX) > 0$,

$$wt(dY) > \begin{cases} \lambda + 1 - wt(dX), & wt(dX) < \lambda + 1 \\ 0, & \text{otherwise} \end{cases}$$

Example 1. The DES S-boxes satisfy $\lambda = 1$. It means (a.o) that if one input bit changes ($wt(dX) = 1 < \lambda + 1$) then at the outputs at least two bits change.

Other properties related to the diffusiveness of an S-box are the stricte avalanche criterion (SAC) [WETA86] and the propagation criterion [PLLG91] — or the higher-order SAC. An S-box satisfies SAC if, given that a single input bit is complemented, for each output bit the probability that it changes is exactly $1/2$. Similary, an S-box satisfies k -order SAC if each output bit changes with a probability of $1/2$ when k or less input bits are complemented. The SAC and the higher-order SAC properties of an S-box imply that the expected number of output changes will not be small (i.e., on a average half the output bits will change) even if the number of input changes is small. However, unlike the diffusion order of an S-box, the SAC

and the higher-order SAC cannot be used to guarantee a lower bound on the number of output changes given a small number of input changes.

Participation of permutation in diffusion

The permutations used in the substitution-permutation networks belong to a special set Π of permutations for which no two outputs of an S-box are connected to one S-box in the next round. Owing to this, the bit changes are spread during the transition between two successive rounds (of course the number of changes does not change but their locations do). If at the output of some S-box m bits change, these changes affect m S-boxes of the next round. It amplifies the effect of diffusion provided by the S-boxes so that these two kind of effects intensify each other. If, in turn, the changes at the inputs of affected S-boxes of that next round are the only changes at the inputs of these S-boxes, at the output of each of them more than λ bits will change, where λ is a diffusion order of the used in the network S-boxes. These changes will be propagated to even one more round of substitution.

As we can see, if we want the diffusion to be the strongest then designing the permutations we have to consider a necessity of fast propagation of changes in many (not only in two) successive rounds. Therefore we have to design the permutations totally for the whole network and not only for selected neighbouring rounds. On Fig. 2a) we present a part of the network in which the permutations are designed this way that they do not enable as good diffusion of changes as in the network presented on Fig. 2b).

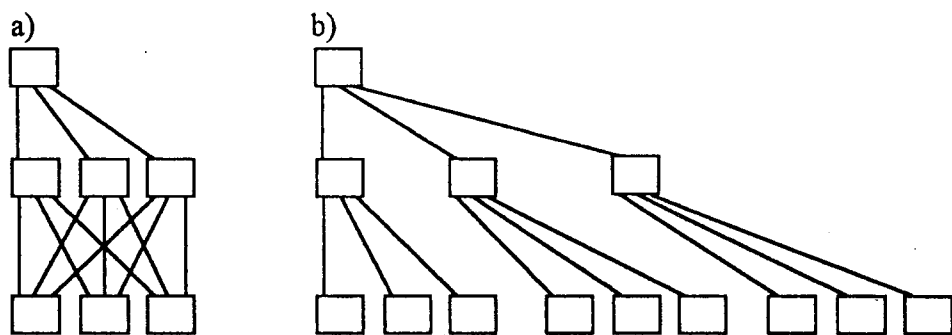


Fig. 2. A participation of permutations in diffusion

Completeness and avalanche effect

One of the requirements for the substitution-permutation network is that every output depends on the greatest number of inputs, in the best case — on all inputs. Some special kind of dependence is achieved for so called *complete networks*. *Completeness* is one of two key requirements for the designed nowadays networks.

Designing the networks which satisfy this requirement is not always easy. A necessary condition of achieving completeness of the network is ensuring so called *connective completeness*, which will be described in further part of this paper. Next we formally define described property of the network.

Definition 2. A one-one function f

$f: V_N \rightarrow V_N$, where V_N is a vector space of N tuples of elements from $GF(2)$ is said to be *complete* if

$$\forall i, j \in \{1, 2, \dots, N\} \quad \exists (v_1, v_2 \in V_N) \quad [v_1 \oplus v_2]_k = \begin{cases} 0, & k \neq i, \quad k \in \{1, 2, \dots, N\} \\ 1, & k = i \end{cases}$$

$[f(v_1) \oplus f(v_2)]_j = 1$, where $[v]_g$ represents g -th bit of the vector v and $\oplus$ represents a bitwise sum mod 2.

Above definition means that function f is complete if for any bit $i \in \{1, 2, \dots, N\}$ we can find two vectors v_1, v_2 such that v_1 and v_2 differ only in the i -th bit, for which $f(v_1)$ and $f(v_2)$ differ at least in the j -th bit.

Definition 3. An S-box is said to be *complete* if substitution it implements is a complete function. A substitution-permutation network is said to be *complete* if the function implemented by it is complete for all key values.

Another key property of the substitution-permutation network is an *avalanche property* [FEIS73], [WETA86], [YOTA96]). An SPN is considered to display good *avalanche effect* if a one bit change in the plaintext is expected to result in close to half the ciphertext bits changing. Good avalanche effect is important to ensure that network is not susceptible to statistical attacks [HETA94]. More formally, the avalanche is defined as follows:

Definition 4. [YOTA96] An SPN is said to satisfy the *avalanche effect* if, for each key, on average half the ciphertext bits change when one plaintext bit is changed. That is $E(wt(dC) | wt(dP) = 1) = N/2$, where dC and dP denote the ciphertext and plaintext change vectors, respectively (N — as above).

An extension of the above definition was proposed by Webster and Tavares [WETA86] and is referred to as the Strict Avalanche Criterion (SAC) (may be used both in respect to an S-box and to a whole network).

Definition 5. A network is said to satisfy the SAC if, for each key, each ciphertext bit changes with a probability of $1/2$ when a single plaintext bit is changed. That is $P(dC_i = 1 | wt(dP) = 1) = 1/2$ where C_i denotes the i -th ciphertext bit, $1 \leq i \leq N$.

It is clear that a network satisfying the SAC must satisfy the avalanche criterion. Satisfaction of the avalanche criterion does not necessarily imply satisfaction of the SAC. While the most SPNs, if treated as randomly selected boolean functions, are expected not to satisfy the SAC [OCON94], most of the SPNs will satisfy the SAC

after sufficiently many rounds. Also, there is no known design procedure that guarantees that the resulting SPN satisfies the SAC.

3. MEYER'S ANALYSIS AND CONNECTIVE COMPLETENESS

As we mentioned, one of the requirements for the substitution-permutation networks is that every output depends on the greatest number of inputs. The best case is that for every key value, every output bit of the network depends on *all* input bits — and not just on some subset of input bits. If the SP-network does not satisfy this property and some output bit (at least for some key values) depends only on a few input bits then it could be found the relation among considered output bit and the correspondig small subset of input bits (by analysing significant number of plaintext — ciphertext pairs). This kind of information could be then used in attacks on the network and might appear usefull.

In [MEYE78] (also in [MEMA82]) author presented an analysis of building dependence of ciphertext bits on plaintext (and key) bits for the Data Encryption Standard (DES) algorithm. In this analysis author investigated how fast the dependence of every output bit on every input (and key) bit was achieved as a function of number of rounds. An assumption in Meyer's analysis is that for each S-box each its output bit is a complex function of all its input bits.

To provide a measure of mentioned dependency, a 64×64 array $G_{a,b}$ is formed. Each element $G_{a,b}(i, j)$ specifies a dependency of j -th output bit of the round b on i -th input bit of round $a + 1$. The number of marked elements in $G_{0,r}$ indicates the degree to which total dependence was achieved by round r .

In the analysis it was shown that in DES, after five rounds each bit depends on all plaintext bits (and all key bits). However, from the aproximate analysis followed that *four* is the minimal number of rounds to achieve total dependece between ciphertext and plaintext (see [MEYE78]). Therefore total dependence has not been achieved in DES in potentially minimal number of rounds.

The dependence of each ciphertext bit on all plaintext and key bits is defined in [MEYE78] as an "intersymbol dependence". Next we define a term *connective completeness*.

Definition 6. A substitution-permutation network satisfies a property of connective completeness if each its output bit has a connection with all its input bits. (We say that it exist a connection between some input of some S-box s_i of the first round and some output of some S-box s_r of the round r if it exist such set of S-boxes $\{s_1, s_2, s_3, \dots, s_r\}$, where s_i represents an S-box of the round i , $i \in \{1, 2, 3, \dots, r\}$, that for $i \in \{1, 2, 3, \dots, r-1\}$ the S-box s_i is connected (through a wire of bit permutation) with an S-box s_{i+1} . Of course, if it exist connection between some input of the S-box s_1 and some output of the S-box s_r , then it exist a connection between every input of s_1 and every output of s_r . Then we say that S-box s_1 is connected with S-box s_r .

It is obvious that if the network satisfies the intersymbol dependence then it also satisfies a connective completeness property; connective completeness is a necessary condition of intersymbol dependence. If we assume — similarly as Meyer did — that for each S-box of the network each its output bit depends on all its input bits then if the network is connectively complete (i.e., satisfies the connective completeness property), it also satisfies the intersymbol dependence. Such assumption makes sense because just the S-boxes of this kind are used in practice. Therefore in such case the terms “intersymbol dependence” and “connective completeness” are equivalent (can be exchanged). Since a subject of paper concerns more topology of the network than its functional properties, we will use the second term, i.e., “connective completeness”.

In further part of this paper we describe designing the substitution-permutation networks for which the connective completeness is achieved in minimal number of rounds.

4. CONNECTIVE COMPLETENESS (INTERSYMBOL DEPENDENCE) VS COMPLETENESS OF THE SPN

As we mentioned, completeness is a special kind of dependence of network's output on its input (or key), defined by Definition 2. From fact that network is complete follows that it satisfies the connective completeness property. Therefore connective completeness is a necessary condition of completeness; it is obvious that network may not be complete if it is not connectively complete — in such case some outputs just do not depend on some inputs.

In general it is much harder to design a complete network than to design a connectively complete network. However there is some kind of networks which may be designed this way that satisfaction of the connective completeness property guarantees satisfaction of the completeness property if used in the network S-boxes are complete. Such networks are one of the topics of our further research.

5. DESIGNING THE PERMUTATIONS

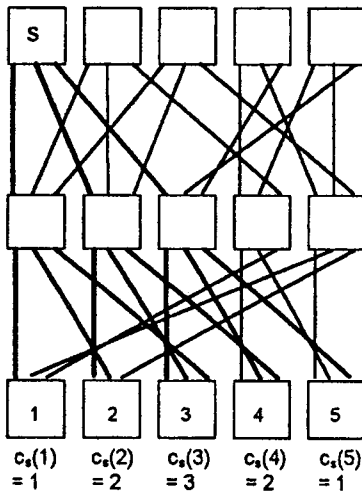
In the following part of this paper while considering the S-boxes of the particular round we will number them using the numbers 1, 2, ..., N starting from the most left S-box of the round. We will denote the S-boxes with use of the small letters (or numbers) and the sets of the S-boxes — with use of the capital letters. Moreover it will always be determined to which round given S-box or given set of S-boxes belongs. A round in which we achieve connective completeness we name a *round of connective completeness*.

Definition 7. Let S-box q belong to the round of connective completeness. An *order of connective completeness* of the S-box q for some S-box s of the first round is a number of all possible connections in the network between S-boxes q and s .

A function of order of connective completeness for some S-box s of the first round is a mapping c_s , which assigns to each S-box of the round of connective completeness a number of all conctions with S-box s (see Fig. 3).

$c_s: \{1, 2, \dots, N\} \rightarrow \{1, 2, \dots, p\}$; the set $\{1, 2, \dots, N\}$ represents the S-boxes of the round of completeness.

Minimal value of the function of order of connective completeness is equal to 1, maximal — to p .



$N = 5, p = 3$

A round of connective completeness is a round number three.

Fig. 3. A function of order of completeness for the S-box s

Definition 8. A set of S-boxes involved by S-box s of the i -th iteration, $i = 1, 2, \dots, N-1$ is a set of the S-boxes of the $i+1$ st iteration which have connections with the S-box s . Shortly we name this set a set involved by S-box s and denote it by $D_{i+1}(s)$.

A set of S-boxes involved by set A of S-boxes of the i -th iteration, $i = 1, 2, \dots, N-1$ is a sum of the sets of S-boxes involved by the elements of the set A . We shortly name this set a set involved by set A and denote it by $D_{i+1}(A)$. Next we present a necessary condition of completeness of the SP-network.

Lemma 1. (a necessary condition of connective completeness)

It is impossible to achieve connective completeness of the substitution-permutation network in the round number of which is smaller than $I_{\min}$, where

$$I_{\min} = 1 + \log_p N. \quad (0)$$

(A very simple corollary from this lemma is that the SPN must not have the number of rounds smaller then $I_{\min}$ if we want to achieve connective completeness in it).

Proof. The connective completeness we can achieve in the round, for which it will be possible to achieve connections between each S-box of this round and each S-box of the first iteration.

Consider any S-box s of the first iteration. It involves p S-boxes of the second round (because every its output must be connected to different S-box of the second round); these ones of the second round can involve at most p^2 S-boxes of the third round. Maximal number of S-boxes involved by the set A of S-boxes of particular round in the next round is equal to $p^*|A|$, where $|Z|$ represent the number of elements of the set Z . In the best case we have

$$\begin{aligned} & |D_2(s) = p, \quad |D_3(D_2(s))| = p^*p = p^2 \text{ etc., until we reach the last round} \\ \text{or} \quad & |D_i(D_{i-1})| = p^*|D_{i-1}| = p^{i-1}, \text{ where } i \text{ is the number of the round of the} \\ & \text{network, } i \in \{3, 4, \dots, R\}. \end{aligned}$$

Therefore it is impossible to achieve connective completeness in the round i if

$$p^{i-1} < N, \quad \text{or for } i < 1 + \log_p N.$$

6. REGULAR NETWORKS

Definition 9. A *regular network* is a network for which following condition is satisfied:

$$\exists u \in N \quad N = p^u.$$

Theorem 1. Consider regular substitution-permutation network. For such network the minimal number of the round in which it is possible to achieve connective completeness is equal to i_c , where

$$i_c = 1 + \log_p N. \quad (1)$$

Proof. From the Lemma 1 follows, that for the regular network, $1 + \log_p N$ is the minimal number of the round, for which the necessary condition of completeness is satisfied. Consider any S-box s of the first round. It involves p S-boxes in the second round: $|D_2(s)| = p$. In the third round maximum p^2 S-boxes can be involved: $|D_3(D_2(s))| = p^*|D_2(s)| = p^2$, etc. In the i -th round at most $p^{i-1} = N$ exactly N S-boxes will be involved. Hence we have:

$$i_c = 1 + \log_p N.$$

Assume that for considered network we have $N = p^u$, $u \in N$. From this follows that $i_c = u - 1$. This way we have shown that any S-box s of the first round can involve exactly N S-boxes in the round given by (1). Now we have to show that it is possible for all S-boxes of the first round simultaneously.

Consider any S-box s of the first round (see Fig. 4 — comments on this figure will be presented further). It involves p S-boxes of the second round (no two outputs of one S-box can be connected to one S-box of the next round). These involved S-boxes create a set $D_2(s)$. Each one of the S-boxes of the set $D_2(s)$ has more $p - 1$

free inputs, therefore more $p - 1$ S-boxes of the first round can be connected to the set $D_2(s)$; every one the same way as the S-box s .

Hence we can group the S-boxes of the second round into $N/p = p^u/p = p^{u-1}$ distinct p -element sets. We can connect to each one of them to p S-boxes of the first round, the similar as described above way. Next we consider any of these sets. Without loss of generality we consider the set $D_2(s)$. Like previously, we connect the outputs of this set (we mean: the output bits of the S-boxes of this set — there are p^2 such outputs) to p^2 S-boxes of the third round — our aim is to involve as many S-boxes of the third round as possible. We involve p^2 — element Set $D_3(D_2(s))$. Every S-box of this set has more $p - 1$ free inputs. We can connect to these inputs the outputs of $p - 1$ sets we have achieved in the second round, the same way we connected the outputs of the set $D_2(s)$. The set $D_3(D_2(s))$ has p^2 elements. We achieve $N/p = p^{ic-1}/p = p^{ic-2}$ elements and each element of this set is connected with $N/p = p^{ic-1}/p$ S-boxes of the first round.

Consider any set of these sets. We connect every one of its outputs — there are $(p^{ic-1}/p)^* p = p^{ic-1}$ of them — to one S-box of the round i_c . Since $N = p^{ic-1}$, this way we involve all S-boxes of the round i_c . Now we have the situation in which all S-boxes of the round i_c are connected with N/p S-boxes of the first round. Each S-box of the round i_c has more $p - 1$ free inputs; in this situation in similar way we can connect the outputs of each one of $p - 1$ left sets of the round $i_c - 1$. In the designed this way network all S-boxes of the round i_c have connections with all S-boxes of the first round.

Algorithm of designing the regular SP-networks

In this section we present an algorithm of designing the regular substitution-permutation networks for which the connective completeness is achieved in as few number of rounds as possible.

In the presented algorithm we use the following notation:

underline — keywords of the algorithm pseudo-language,

normal — variables,

italic — operations made in the algorithm,

italic — comments,

UPPER_LETTERS — constants.

In the algorithm we number both the S-boxes of particular round and the inputs of S-boxes starting with zero. We number the rounds starting with one.

Algorithm 1

first we group the S-boxes of every round of the network into the sets

for round = 1 to ROUND_OF_COMPLETENESS-1

begin

- Divide the round round into $N/p^{\text{round}-1}$ sets ($p^{\text{round}-1}$ — element sets) taking to the successive sets successive S-boxes starting from the S-box number 1 (the most left S-box).

- *Save the partition of the round in memory.*

end

next we create the connections

for round = 1 to ROUND_OF_COMPLETENESS-1 do

begin

for all sets of the round round:

for set = 0 to ($N/p^{\text{round}-1} - 1$) do

begin

Connect the outputs of the set set of the round round to the inputs number

(set mod p of the set [set/ p] of the round round + 1

end

end

On the Fig. 4 we present a sample regular SPN designed with use of the Algorithm 1 ($p = 3$, $N = 9$, $i_c = 3$). The groups of the S-boxes of every round achieved as a result of the first loop of the algorithm are marked with dotted boxes.

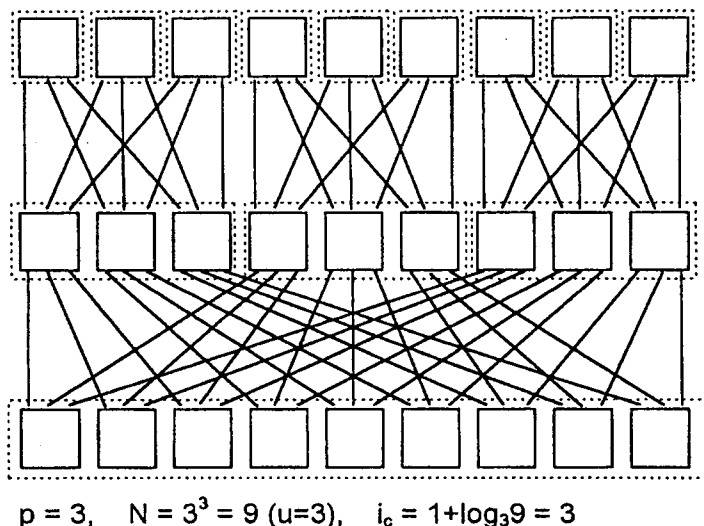


Fig. 4. A SP-network designed with use of the Algorithm 1

7. IRREGULAR NETWORKS

Definition 10. An *irregular network* is a network which is not regular, it means a network for which we have:

$$\sim (\exists u \in N : N = p^u).$$

Definition 11. An *e-set* of the S-box s of the round i , $i \in \{2, 3, \dots, r\}$ is a set of these S-boxes of the first round (represented by numbers from the set $\{1, 2, \dots, N\}$) with

which S-box s has connections. This set can contain replications of the elements; the number of occurrences of particular S-box of the first round in the e-set of the S-box s of the round i is equal to a number of connections of this S-box with the S-box s .

If the e-set of some S-box of some round is $\{1, 2, 4, 1, 3, 4\}$ then it means that this S-box has two connections with the S-box 1 of the first round, two connections with the S-box 4, one connection with the S-box 2 and one connection with the S-box 3.

Theorem 2. Consider irregular substitution-permutation network. For such network the minimal number of the round in which it is possible to achieve connective completeness is equal to i_c , where

$$i_c = [2 + \log_p N]. \quad (2)$$

Proof. (Sketch) Since

$$[2 + \log_p N] > I_{\min}, \text{ where } I_{\min} \text{ is determined by (0)}$$

then for the round number $[2 + \log_p N]$ the necessary condition of completeness is satisfied. Note that $[2 + \log_p N]$ is the smallest natural number greater than $I_{\min}$. Therefore we will show that achieving the completeness of the network is possible for the round of the smallest number for which the necessary condition of completeness is satisfied.

Consider second round. We can connect every S-box of the second round with p different S-boxes of the first round. For the second round we achieve N p -element e-sets of different elements (the elements of one e-set differ from each other). Since each S-box of the first round has connections with p different S-boxes of the second round then each element of the set $\{1, 2, \dots, N\}$ belongs to exactly p e-sets of the second round. For any S-box s of the second round we can find at least $p-1$ such S-boxes of this round that s and these S-boxes have pairwise separable e-sets (we mean the e-set of the S-box s and the e-sets of these found S-boxes are pairwise separable) if $(N/p) \geq p$. If this condition is satisfied we will be able to create for the second round p -element e-sets such that for each one of them it will be possible to find $p-1$ other e-sets which will be separable with it and pairwise separable with each other.

Assume that condition $(N/p) \geq p$ is satisfied. Consider round number three. Give any S-box a_1 of this round and connect it with any S-box b_1 of the second round. Since condition $(N/p) \geq p$ is satisfied, in the second round for the S-box b_1 we can find $p-1$ S-boxes such that b_1 and these S-boxes have pairwise separable e-sets. We can connect these $p-1$ S-boxes with the S-box a_1 of the third round. Now e-set of the S-box a_1 contains p^2 different elements. Next consider another S-box a_2 of the third round. We connect this S-box with one of these S-boxes of the second round which have the greatest number of free inputs (input bits). Let b_2 represent such S-box. For the S-box b_2 we can find at least $p-1$ S-boxes of the second round such that b_2 and these S-boxes have pairwise separable e-sets. We connect with a_2 these ones of the found S-boxes which have the greatest number of

free outputs. Now the e-set of the S-box a_2 contains p^2 different elements. Similar we can connect remaining S-boxes of the third round. If while creating new connection we have a choice of few S-boxes of the second round, we choose this S-box which has the most free outputs.

Consider any S-box a of the third round. To make finding for S-box a $p - 1$ such S-boxes that a and these S-boxes have pairwise separable e-sets possible, it must be satisfied a condition $(N/p^2) > p$.

Similar way we construct the connections for further rounds. In general, for any S-box r of the j -th round it is possible to find $p - 1$ S-boxes of this round such that S-box r and these $p - 1$ S-boxes have pairwise separable e-sets if following condition is satisfied

$$N/p^{j-1} > p \Leftrightarrow j \leq \log_p N$$

$$1 \leq j \leq i_c - 2, \quad j \in N, \quad i_c \text{ — determined by (2)}$$

If $i_c - 2$ is the greatest number of the round in which for any S-box r it is possible to find $p - 1$ S-boxes e-sets of which and e-set of r are pairwise separable then $i_c - 1$ is the greatest number of the round S-boxes of which have e-sets of different components. The e-sets of the S-boxes of the round i_c will be containing replications of the elements (because in the round number $i_c - 1$ for each S-box it is impossible to find $p - 1$ S-boxes of this round satisfying described above condition <it may be possible find less such S-boxes>).

The e-sets of the S-boxes of the round $i_c - 1$ contain p^{i_c-1} different elements. It means that every S-box of the round $i_c - 1$ has connections with p^{i_c-2} elements of the first round. Moreover, given element of the set $\{1, 2, \dots, N\}$ occurs in p^{i_c-2} e-sets of the round $i_c - 1$. We connect any S-box of the round i_c with p S-boxes of the round $i_c - 1$. The e-set of this S-box will be containing p^{i_c-1} elements, $p^{i_c-1} > N$. We can connect any S-box of the round i_c with such S-boxes of the round $i_c - 1$ that in the e-set of this S-box we will be having all elements of the set $\{1, 2, \dots, N\}$. This way we achieve completeness of the network.

Algorithm of designing the irregular SP-networks

We present an algorithm of designing the irregular substitution-permutation networks for which the connective completeness is achieved in the fewest number of rounds. This algorithm may also be used to design regular networks but it is not as simple as the Algorithm we number the S-boxes (and the rounds) starting with one.

Algorithm 2.

```

for round = 2 to ROUND_OF_COMPLETENESS do
  begin
    for sbbox = 1 to N do
      begin
        if round < ROUND_OF_COMPLETENESS
          begin
            free: = p

```

```

repeat
    Connect successive inputs of the S-box sbos of the round round with the
    outputs of the S-boxes of the round round-1
        • with which S-box sbos has no connections,
        • which have free free inputs, and
        • the e-sets of which are separable with the currently achieved e-set of
          the S-box sbos
    After each new connection update the e-set of the S-box sbos
    free: = free - 1
until all outputs of the S-box sbos are connected
end
else # round = ROUND_OF_COMPLETENESS
begin
    Connect successive inputs of the S-box sbos of the round roundn with the
    outputs of the S-boxes of the round round-1
        • with which S-box sbos has no connections
        • the e-sets of which have the fewest number of shared elements with as
          far created e-set of the S-box sbos
    If there are more then one S-boxes satisfying above conditions connect with
    this one which has the greatest number of free outputs.
    After each new connection update the e-set of the S-box sbos.
end

```

Example 2. On the Fig. 5 we present a sample SP-network designed with use of the Algorithm 2 ($p = 2$, $N = 5$). At every S-box of the rounds 2, 3 and 4 we gave the content of its e-set. The round of connective completeness of this network is the

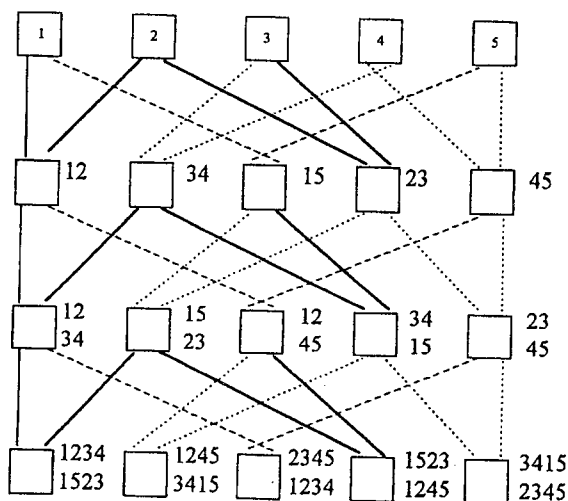


Fig. 5. A SP-network designed with use of the Algorithm 1

round number 4. According to Theorem 2, four is the minimal number of the round in which it is possible to achieve connective completeness:

$$i_c[2 + \log_p N] = [2 + \log_2 5] = 4.$$

8. FURTHER ROUNDS

At the end we proof some simple fact:

Fact 1. Each output of each round further that the round of connective completeness depends on each input bit of the network (independently on the kind of used before these further rounds bit permutations).

Proof. Consider any S-box s of the round i , $i > i_c$. This S-box has at least one connection with some of the S-boxes of the round of connective completeness, give q . Since q belongs to the round of connective completeness then it has connections with all S-boxes of the first round. Hence also s has connections with all S-boxes of the first round (of course it may have other connections than the one through q).

However we can still use Algorithms 1 and 2 in designing permutations of further rounds to achieve good diffusiveness of bit changes in the network.

SUMMARY

In the paper we introduced a term *connective completeness* of the network. This term is equivalent to a term *intersymbol dependence* (introduced by Meyer in [MEYE78]) if network contains the S-boxes for which each output bit depends on all input bits. Our term is used more in respect to a topology of the network. We noted that connective completeness is a necessary condition of completeness of the network.

Next we determined a minimal number of the round in which it is possible to achieve connective completeness, both for the regular and irregular networks. We presented the algorithms of designing regular and irregular networks which ensure achieving the connective completeness is minimal number of rounds.

In further research we are going to investigate how should the regular and irregular networks be designed to satisfy the completeness property.

REFERENCES

- [BRSE89] L. Brown, J. Seberry: *On the design of permutation P in DES type cryptosystems*. Advances in Cryptology — EUROCRYPT'89, 696—705, Springer-Verlag, 1990
- [FEIS73] H. Feistel: *Cryptography and Computer Security*. Scientific American 1973, Vol. 228, no. 5, pp. 5—23
- [HETA94] H.M. Heys, S.E. Tavares: *Key clustering in substitution-permutation network cryptosystems*. Workshop in Selected Areas of Cryptography (SAC'94), pp. 134—145, Workshop record, Ontario, 1994
- [HETA96] H.M. Heys, S.E. Tavares: *Substitution-Permutation Networks Resistant to Differential and Linear Cryptanalysis*. Journal of Cryptology, 1996, Vol. 9 No. 1, pp. 1—20

5. [KADA79] J.B. Kam, G.I. Davida: *Structured Design of Substitution-Permutation Encryption Networks*. IEEE Transactions on Computers, 1979, Vol. C-28, No. 10, pp. 747–753
6. [KIM92] K. Kim: *Construction of DES-like S-boxes Based on Boolean Functions Satisfying the SAC*. Advances in Cryptology — ASIACRYPT'91, pp. 59–72, Springer-Verlag, 1992
7. [KNUD93] L.R. Dnudsen: *Iterative Characteristics of DES and S^2 -DES*. Advances in Cryptology CRYPTO'92, pp. 497–511, Springer-Verlag, 1993
8. [MEMA82] C.H. Meyer, S.M. Matyas: *Cryptography: A New Dimension in Data Security*. John Wiley & Sons, New York 1982
9. [MEYE78] C.H. Meyer: *Ciphertext/plaintext and ciphertext/key dependence vs number of rounds for the data encryption standard* AFIPS Conference Proceedings 47, pp. 1119–1126, AFIPS Press, 1978
10. [OCON94] L.J.O'Connor: *An upper bound on the number of functions satisfying the Strict Avalanche Criterion*. Information Processing Letters 52, 1994 pp. 325–327
11. [PLLG91] B. Preneel, W. Van Leekwijck, L. Van Linden, R. Goevarts, J. VBanderwalle: *Propagation characteristics of boolean functions*. Advances in Cryptology — EUROCRYPT'90, pp. 161–173, Springer-Verlag, 1991
12. [SADO97] A. Sadowski: *Key-Controlled Bit Permutations in Substitution-Permutation Network*. Kwart. Elektr. i Telekom. 1997, Vol. 43, z. 3
13. [SEZH93] J. Seberry, X.M. Zhang: *Highly nonlinear 0–1 balanced Boolean functions satisfying strict avalanche criterion*. Advances in Cryptology — AUSCRYPT'92, pp. 145–156, Springer-Verlag, 1993
14. [YOTA96] A.M. Youssef, S.E. Tavares: *Modelling Avalanche Characteristics of a Class of Substitution-Permutation Networks*. Proceedings of PRAGOCRYPT'96, pp. 18–26, GS UCMP, 1996
15. [WETA86] A.F. Webster, S.E. Tavares: *On the design of S-boxes*. Advances in Cryptology — CRYPTO'85 pp. 523–534, Springer-Verlag, 1986

A. SADOWSKI

PROJEKTOWANIE PERMUTACJI BITÓW SIECI PODSTAWIENIOWO-PERMUTACYJNYCH

Streszczenie

W artykule opisano permutacje bitów stosowane w sieciach podstawieniowo-permutacyjnych. Zdefiniowano pojęcia: regularna i nieregularna sieć podstawieniowo-permutacyjna i zaprezentowano metody projektowania tych rodzajów sieci. Przedstawione metody pozwalają na uzyskanie tzw. połączeniowej zupełności sieci w jak najmniejszej liczbie rund.

Słowa kluczowe: permutacja, szyfry produktowe, zupełne sieci podstawieniowo-permutacyjne.

System segmentacji obrazu i śledzenia obiektów w czasie rzeczywistym

JERZY SIUZDAK

Instytut Telekomunikacji, Politechnika Warszawska

Otrzymano 1997.06.19

Autoryzowano 1997.09.21

W artykule opisano zasadę działania systemu przetwarzania obrazu umożliwiającego wydzielanie obiektów tego obrazu w czasie rzeczywistym. Po przetworzeniu na postać cyfrową, jasność obrazu podlega filtracji dwuwymiarowej w celu wydzielenia konturów. Powstały w ten sposób obraz binarny transferowany jest do pamięci RAM komputera IBM PC, gdzie podlega dalszej obróbce. Urządzenie umożliwia śledzenie w czasie rzeczywistym obiektów poruszających się na ekranie. Przedyskutowano dobór filtru dwuwymiarowego oraz wpływ zakłóceń i ruchu obiektu na pracę systemu.

Słowa kluczowe: przetwarzanie obrazów, obrazy binarne, filtry dwuwymiarowe

1. WSTĘP

W wielu zastosowaniach automatyki, elektroniki i telekomunikacji pojawia się konieczność przetwarzania obrazów. Wachlarz zastosowań jest ogromny: począwszy od najstarszych i najprostszych aplikacji, takich jak przesyłanie obrazu telewizyjnego, po najnowsze i najbardziej złożone systemy sztucznego wzroku. Jedną z ważniejszych aplikacji jest segmentacja obrazu [1, 2], która umożliwia wydzielenie z obrazu konkretnych obiektów na nim się znajdujących. Ustala się w ten sposób położenie tych obiektów, co — jeśli jest to potrzebne — pozwala na ich śledzenie. Informacja o położeniu obiektów może być też wykorzystana w inny sposób, np. przy sterowaniu automatów przemysłowych czy w kontroli jakości. W wielu zastosowaniach (automatyka przemysłowa, kierowanie pojazdami, zastosowania wojskowe) jest wymagana obróbka obrazu w czasie rzeczywistym. W dostępnej literaturze stosunkowo rzadko można znaleźć omówienie tego typu urządzeń. Jako wyjątki można tu przytoczyć prace [3, 4, 5, 6]. Istnieje kilka podstawowych podejść do segmentacji obrazu, które umożliwiają działanie w czasie rzeczywistym.

Najprostszym z nich jest porównanie jaskrawości punktów obrazu z jednym bądź kilkoma poziomami progowymi [7]. Przyjmuje się, że punkty których jaskrawość leży między wybranymi poziomami progowymi, należą do określonego obszaru (obiektu). Połączone obszary dają żadaną segmentację. Metoda ta jest najbardziej skuteczna dla jednorodnych obiektów położonych na jednakowym tle, np. liter na białej stronie. Największym problemem przy zastosowaniu tej techniki jest wybór poziomów progowych. Poziomy progowe mogą zostać wybrane przez doświadczenia ze zbiorem wybranych, typowych obrazów. Inną możliwością jest analiza histogramu jaskrawości danego obrazu. Jeżeli histogram ten ma widoczne maksima i minima, to te ostatnie wyznaczają poziomy progowe.

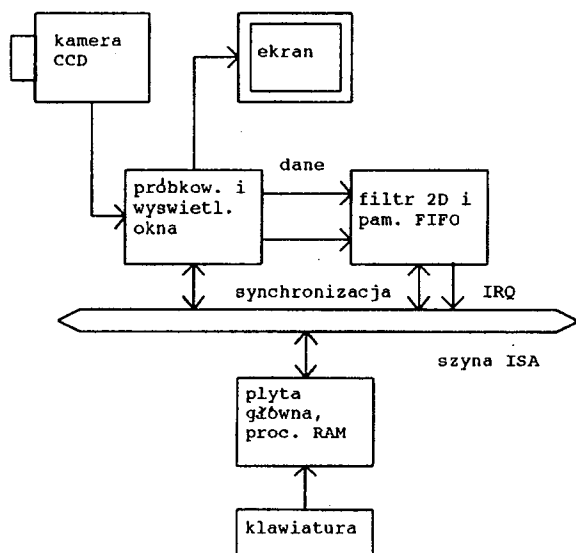
Inną ważną grupą metod segmentacji są metody bazujące na wykrywaniu krawędzi obiektów. Wykorzystują one to, że na krawędziach obiektów jaskrawość obrazu ulega szybkim zmianom. Wobec tego dokonując korelacji obrazu z odpowiednio dobraną maską lub grupą masek można te szybkie zmiany odpowiednio uwypuklić [1]. Jedną z możliwości jest tutaj zastosowanie masek pozwalających na przybliżone obliczenie gradientu jaskrawości w każdym punkcie. W dalszych etapach dokonuje się porównania modułu tak wyliczonego gradientu z uprzednio ustalonym progiem. Punkty, w których wartość progu zostaje przekroczona, są zaklasyfikowane jako punkty krawędzi obiektów, umożliwiając w ten sposób dokonanie segmentacji.

Ostatnia wreszcie grupa metod polega na dokonaniu korelacji obrazu z maską, której rozkład jaskrawości odpowiada poszukiwanemu obiektowi. Maksimum funkcji korelacji odpowiada położeniu poszukiwanego obiektu. Odpowiednie algorytmy optymalizacyjne ograniczają liczbę niezbędnych obliczeń.

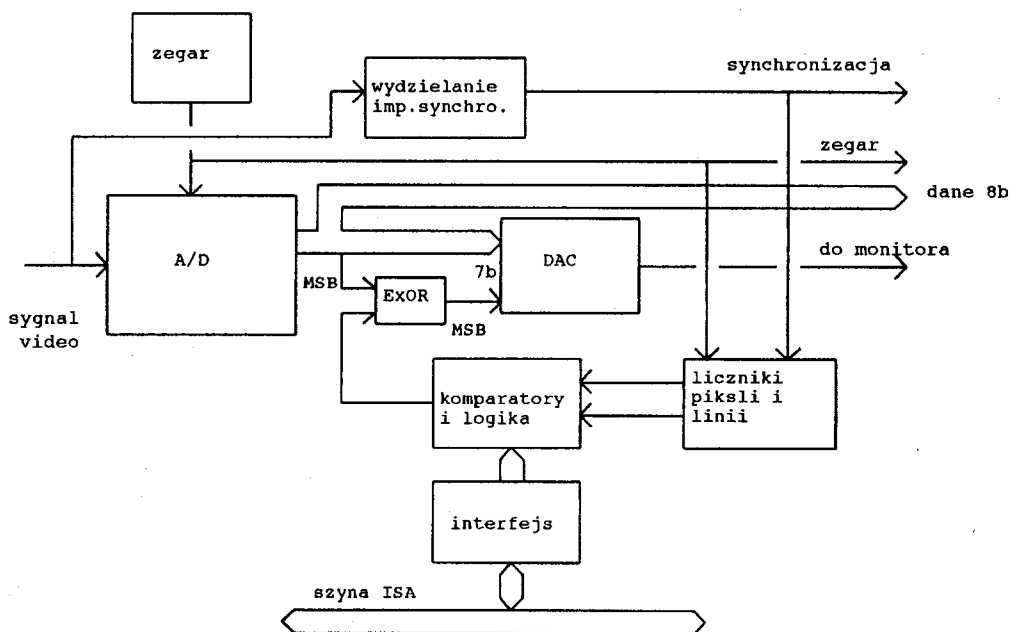
W Instytucie Telekomunikacji Politechniki Warszawskiej, korzystając z grantu rektorskiego nr 905/6, opracowano system umożliwiający dokonanie w czasie rzeczywistym segmentacji przychodzącego obrazu telewizyjnego, a następnie pozwalający na śledzenie wybranych obiektów. W systemie wykorzystano filtrację dwuwymiarową. Niniejsza publikacja jest poświęcona omówieniu budowy i działania wspomnianego urządzenia.

2. ZASADA DZIAŁANIA

System został zbudowany w oparciu o komputer IBM PC i fizycznie zajmuje dwie karty rozszerzeń współpracujące z szyną ISA komputera. Ogólny schemat blokowy całego urządzenia przedstawiono na rys. 1, a zasada jego pracy jest opisana poniżej. Kamera wideo CCD jest źródłem standardowego sygnału video ($1 V_{p-p}$). Sygnał z kamery doprowadzono do karty mieszacza wideo, której schemat blokowy pokazano na rys. 2. Tam, w układzie szybkiego przetwornika A/D typu flash, sygnał video zostaje zamieniony na 8. bitowe dane binarne, które odpowiadają jaskrawości danego punktu na obrazie. Przetwornik jest taktowany zegarem 10 MHz tak, że na 1 linię obrazu przypada 512 próbek. Zastosowano tu układ CA2218 firmy Harris. Jednocześnie w innym układzie z pełnego sygnału video wydzielane są impulsy synchronizacji poziomej H i pionowej V (linii i ramki). Dane binarne wraz

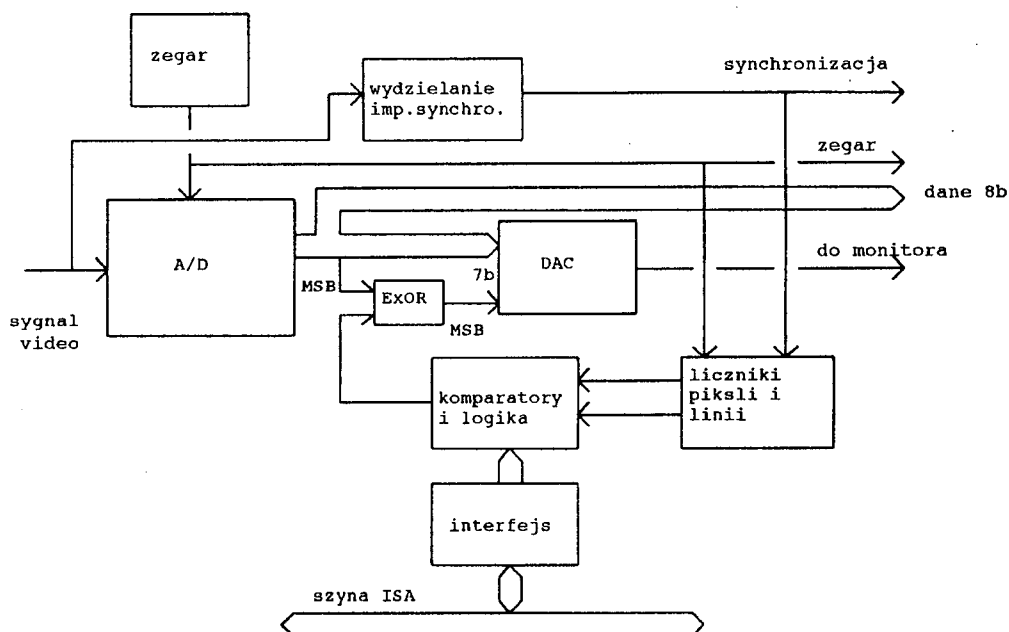


Rys. 1. Schemat blokowy urządzenia



Rys. 2. Schemat blokowy karty mieszacza wideo

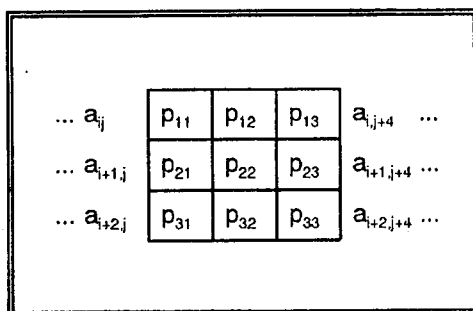
z sygnałami synchronizacji są przesyłane do drugiej karty, na której umieszczono cyfrowy filtr dwuwymiarowy i pamięć FIFO (ang. *first in first out*). Schemat blokowy układów znajdujących się na drugiej karcie pokazano na rys. 3. Dane binarne wraz z sygnałami synchronizacji dochodzą na wejścia filtru dwuwymiarowego zbudowanego w oparciu o układ PDSP 16488 firmy Plessey [8].



Rys. 3. Schemat blokowy układu filtracji dwuwymiarowej i pamięci FIFO

Dzięki wbudowanym kilku rejstrów przesuwnych, z których każdy umożliwia opóźnienie danych o jedną linię, możliwe jest przemnożenie dochodzącego w postaci szeregowej obrazu przez maskę (filtr) o wymiarach do 8×8 . Liczona jest wtedy na bieżąco korelacja obrazu z daną maską. Wynik tej korelacji otrzymujemy na bieżąco na wyjściu filtru, jedynie z niewielkim opóźnieniem spowodowanym obróbką potokową danych. Zasadę mnożenia pokazano na rys. 4 dla filtru o wymiarach 3×3 ; łatwo ją uogólnić dla większych wymiarów masek.

Przykłady możliwych do zastosowania typów filtrów pokazano na rys. 5 (filtry wykrywające krawędzie) i rys. 6 (filtry typu Laplasjanu). Wynik mnożenia przez daną maskę daje dużą (co do modułu) liczbę w przypadku jeśli jest ona dopasowana do danego fragmentu obrazu. Dla przykładu filtry z rys. 5 i 6 dają duże co do modułu wartości na wyjściu w przypadku napotkania krawędzi obiektów. Zagadnienie doboru filtrów i ich właściwości dokładniej omówimy w następnym rozdziale. Wielkość sygnału wyjściowego filtru dwuwymiarowego jest wewnętrznie porównywana z predefiniowanym progiem i daje na wyjściu wartość 1 przy przekroczeniu



Rys. 4. Ilustracja zasady filtracji dwuwymiarowej (mnożenie obrazu przez maskę)

a)

1/3	0	-1/3
1/3	0	-1/3
1/3	0	-1/3

b)

1/8	1/8	-1/8	-1/8
1/8	1/8	-1/8	-1/8
1/8	1/8	-1/8	-1/8
1/8	1/8	-1/8	-1/8

c)

1/10	1/10	0	-1/10	-1/10
1/10	1/10	0	-1/10	-1/10
1/10	1/10	0	-1/10	-1/10
1/10	1/10	0	-1/10	-1/10
1/10	1/10	0	-1/10	-1/10

d)

1/18	1/18	1/18	-1/18	-1/18	-1/18
1/18	1/18	1/18	-1/18	-1/18	-1/18
1/18	1/18	1/18	-1/18	-1/18	-1/18
1/18	1/18	1/18	-1/18	-1/18	-1/18
1/18	1/18	1/18	-1/18	-1/18	-1/18
1/18	1/18	1/18	-1/18	-1/18	-1/18

Rys. 5. Maski wykrywające krawędzie o różnych wymiarach: a) 3×3 , b) 4×4 , c) 5×5 , d) 6×6

a)

$-1/4$	$-1/4$	$-1/4$	$-1/4$
$-1/4$	$3/4$	$3/4$	$-1/4$
$-1/4$	$3/4$	$3/4$	$-1/4$
$-1/4$	$-1/4$	$-1/4$	$-1/4$

b)

$-1/5$	$-1/5$	$-1/5$	$-1/5$	$-1/5$
$-1/5$	$16/45$	$16/45$	$16/45$	$-1/5$
$-1/5$	$16/45$	$16/45$	$16/45$	$-1/5$
$-1/5$	$16/45$	$16/45$	$16/45$	$-1/5$
$-1/5$	$-1/5$	$-1/5$	$-1/5$	$-1/5$

c)

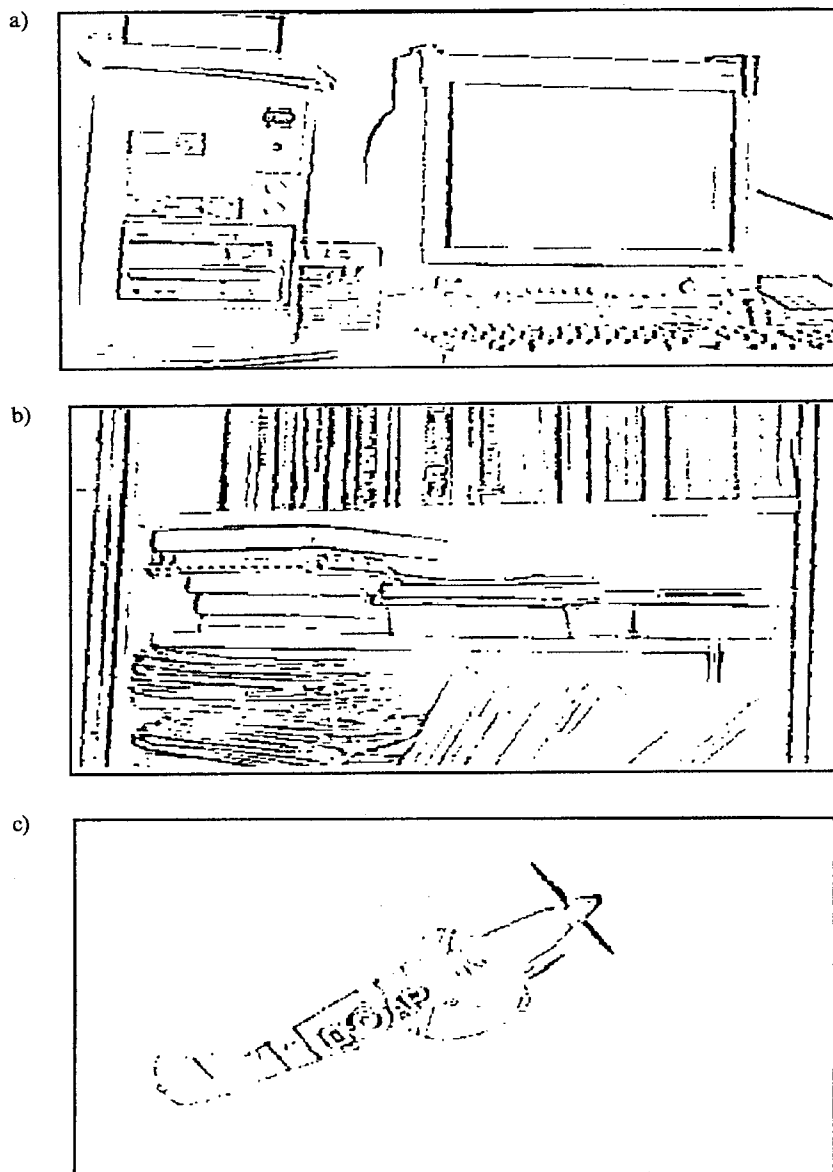
$-1/6$	$-1/6$	$-1/6$	$-1/6$	$-1/6$	$-1/6$
$-1/6$	$5/24$	$5/24$	$5/24$	$5/24$	$-1/6$
$-1/6$	$5/24$	$5/24$	$5/24$	$5/24$	$-1/6$
$-1/6$	$5/24$	$5/24$	$5/24$	$5/24$	$-1/6$
$-1/6$	$5/24$	$5/24$	$5/24$	$5/24$	$-1/6$
$-1/6$	$-1/6$	$-1/6$	$-1/6$	$-1/6$	$-1/6$

d)

$-1/12$	$-1/12$	$-1/12$	$-1/12$	$-1/12$	$-1/12$
$-1/12$	$-1/12$	$-1/12$	$-1/12$	$-1/12$	$-1/12$
$-1/12$	$-1/12$	$2/3$	$2/3$	$-1/12$	$-1/12$
$-1/12$	$-1/12$	$2/3$	$2/3$	$-1/12$	$-1/12$
$-1/12$	$-1/12$	$-1/12$	$-1/12$	$-1/12$	$-1/12$
$-1/12$	$-1/12$	$-1/12$	$-1/12$	$-1/12$	$-1/12$

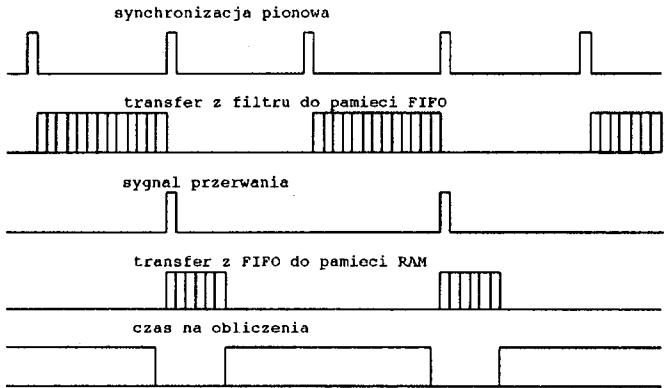
Rys. 6. Maski Laplace'a o różnych wymiarach: a) $4 \times 4(2 \times 2)$, b) $5 \times 5(3 \times 3)$, c) $6 \times 6(4 \times 4)$, d) $6 \times 6(2 \times 2)$.
W nawiasach podano wymiar centralnej części maski

progu i 0 — kiedy próg nie został przekroczony. Zatem przy doborze filtru (lub ich grupy) dopasowanych do krawędzi obiektów, na wyjściu filtru otrzymujemy obraz binarny, na którym niezerowe są jedynie krawędzie obiektów. Przykłady takich obrazów, otrzymanych z omawianego urządzenia, są pokazane na rys. 7. Ten binarny obraz poprzez 16 bitowy rejestr przesuwany podawany jest do pamięci FIFO (Matra-Harris M67205), gdzie zostaje zapamiętany. Pamięci są tak skonfigurowane, że umożliwiają zapamiętanie 256 linii obrazu, z których każda ma 512 piksli.



Rys. 7. Przykłady binarnych obrazów krawędziowych wytworzonych przez urządzenie: a) komputer, b) książki, c) model samolotu

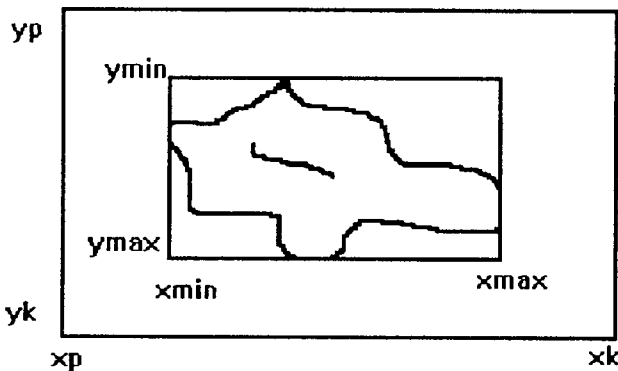
Zapamiętywanie zachodzi na bieżąco w ciągu danego półobrazu. Po zapamiętaniu całego półobrazu karta zgłasza przerwanie po szynie ISA. Program obsługi tego przerwania powoduje wczytanie danych z FIFO do pamięci RAM komputera i następnie dalszą ich obróbkę.



Rys. 8. Przebiegi czasowe w układzie

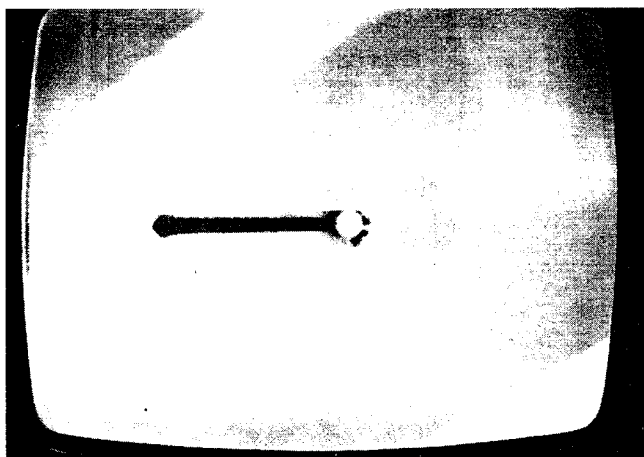
Przebiegi czasowe w układzie pokazano na rys. 8. Program, w sposób omówiony w następnym paragrafie, wyznacza położenie obiektu na obrazie i przekazuje po szynie ISA dane o wymiarach obszaru, w którym zawarty jest obiekt, do karty mieszacza wideo.

Dane te są przekazywane jako cztery współrzędne obszaru, w którym zawarty jest obiekt: $x_{\min}$, $x_{\max}$, $y_{\min}$, $y_{\max}$ (patrz rys. 9) i wyświetlane na ekranie w formie okna



Rys. 9. Współrzędne obszaru zawierającego obiekt

obserwacyjnego. Pokazano to na rys. 10. W tym celu na karcie mieszacza wideo te współrzędne są zapamiętywane, a następnie porównywane ze zmieniającą się na bieżąco zawartością liczników piksli i linii. W odpowiednich momentach układ logiczny wytwarza sygnał logiczny 0, który powoduje inwersję najstarszego bitu z przetwornika A/C i wyświetlenie okna obserwacyjnego. Dane binarne z naniesionym oknem są następnie poddawane konwersji C/A (układ DAC08) i wyświetlane jako obraz na ekranie monitora (patrz rys. 10).



Rys. 10. Fotografia okna obserwacyjnego na ekranie głównego monitora

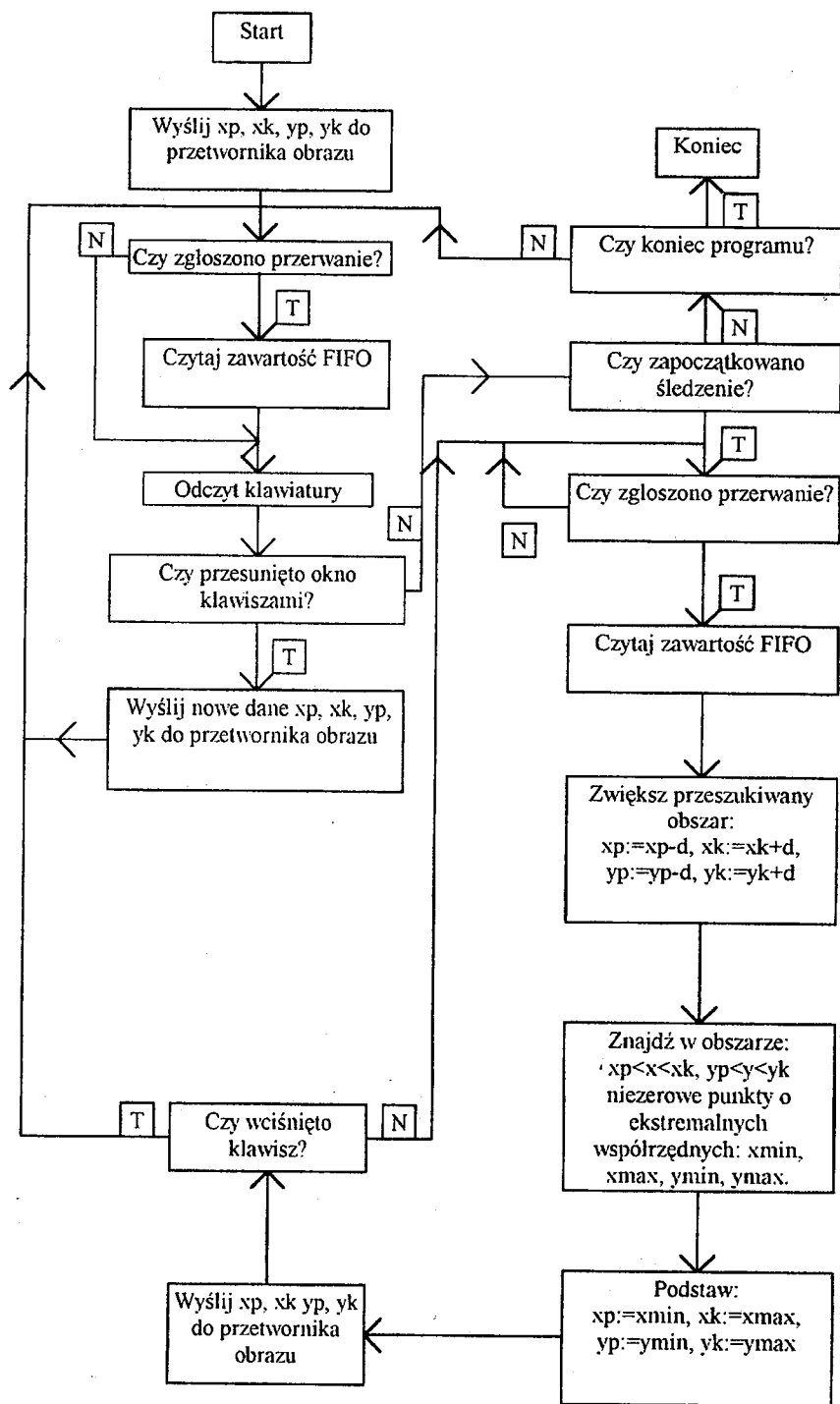
W celu kontrolnym do zapamiętywania danych z pamięci FIFO wykorzystywana jest część pamięci RAM zarezerwowana dla karty graficznej. Pozwala to wyświetlić obraz krawędzi na ekranie drugiego monitora (komputerowego). W trakcie zmian położenia śledzonego obiektu zmieniają się wartości $x_{\min}$, $x_{\max}$, $y_{\min}$, $y_{\max}$ obliczone przez komputer, a więc i położenie okna obserwacyjnego na ekranie co pozwala kontrolować prawidłowość procesu segmentacji i śledzenia.

3. OPROGRAMOWANIE

Sieć działań programu, pod którego kontrolą pracuje urządzenie, pokazano na rys. 11. Po uruchomieniu programu na ekran monitora zostają wysłane dane odpowiadające początkowemu położeniu okna obserwacyjnego (standardowe wymiary, pozycja centralna). Program śledzi na bieżąco status klawiatury i monitorując przerwania dokonuje transferu danych z pamięci FIFO do RAM. Operator wciskając odpowiednie klawisze ma możliwość przesuwania okna obserwacyjnego na ekranie. W momencie, kiedy okno obserwacyjne jest skierowane na wybrany obiekt, operator inicjuje śledzenie automatyczne poprzez naciśnięcie odpowiedniego klawisza. Program oczekuje wtedy na przerwanie z karty filtru dwuwymiarowego i pamięci FIFO. Przy wystąpieniu przerwania zostaje zainicjowane przesłanie danych z pamięci FIFO do pamięci RAM szyną ISA. Po zakończeniu transmisji program (komunikując się już tylko z pamięcią RAM) zaczyna przeszukiwanie obszaru o zwiększonych w stosunku do okna obserwacyjnego (x_p , x_k , y_p , y_k) wymiarach:

$$x_p = x_p - d, x_k = x_k + d, y_p = y_p - d, y_k = y_k + d.$$

Wyszukiwane są punkty krawędzi obiektu (niezerowe piksele) o ekstremalnych współrzędnych: $x_{\min}$, $x_{\max}$, $y_{\min}$, $y_{\max}$. Pokazano to na rys. 9. Zwiększenie wymiarów przeszukiwanego obszaru w stosunku do wymiarów okna obserwacyjnego jest



Rys. 11. Sieć działań programu

podyktowane koniecznością dopasowania wymiarów okna do wymiarów obiektu w przypadku, jeśli początkowe wymiary okna są mniejsze od wymiarów obiektu. Ponadto umożliwia to poprawne śledzenie poruszających się obiektów. Duże wartości parametru d ułatwiają śledzenie obiektów, ale utrudniają oddzielenie wybranego przedmiotu od tła, stąd wybór wartości d musi być kompromisowy. Punkty o ekstremalnych współrzędnych wyznaczają wymiary okna zawierającego obiekt. Dane te są wysyłane do miksera wideo i w konsekwencji na ekran. Ponieważ jest to czynione na bieżąco, operator może kontrolować poprawność śledzenia. W przypadku niepoprawnego działania układu operator ma możliwość wyłączenia śledzenia automatycznego i powrotu do manualnego przesuwania okna obserwacyjnego. Komputer bowiem również przy śledzeniu automatycznym monitoruje stan klawiatury.

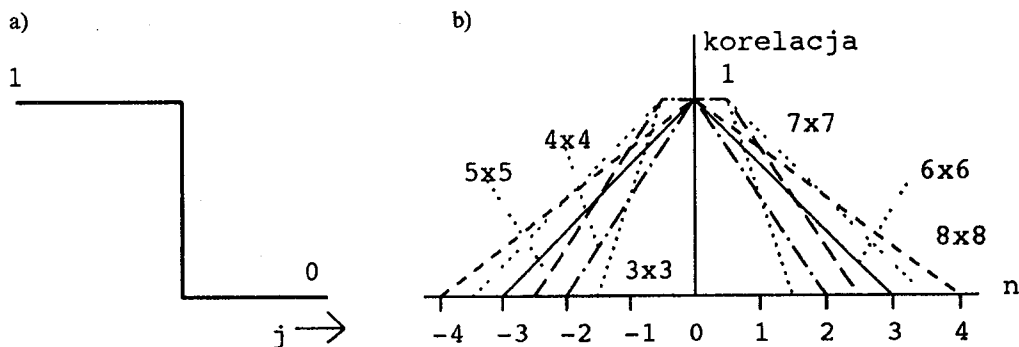
4. WYBÓR RODZAJU FILTRA DWUWYMIAROWEGO

O niewłaściwościach systemu i jego poprawnej pracy w zasadniczym stopniu decyduje wybór filtra dwuwymiarowego. Przez odpowiedni dobór współczynników tego filtra p_{ij} można uzyskać filtr dopasowany do danego przedmiotu. Taki filtr daje dużą wartość funkcji korelacji z obrazem (dużą wartość sygnału wyjściowego) w przypadku napotkania na obrazie przedmiotu, do którego jest dopasowany. Z teorii filtrów dopasowanych jednowymiarowych (elektrycznych) wiadomo [9], że dają one największy stosunek maksymalnej wartości sygnału do mocy szumu. Jeżeli jaskrawość danego obiektu jest wyznaczona przez macierz $[a_{ij}]_{N \times N}$, to filtr dopasowany jest określony macierzą $[p_{ij}]_{N \times N}$, gdzie

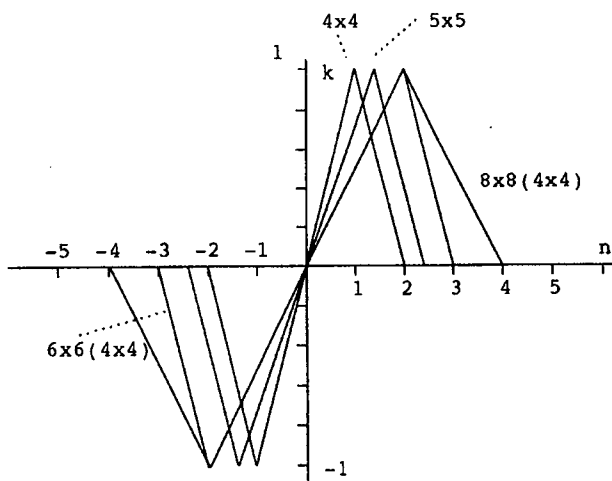
$$p_{ij} = a_{ij} - A, \quad A = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N a_{ij} \quad (1)$$

Odjęcie średniej wartości jaskrawości obiektu A zapewnia zerową korelację z tłem o stałej wartości. Ponieważ pojedynczy układ PDSP 16488 zapewnia maksymalny wymiar maski równy co najwyżej 8×8 , może być on tylko użyty do wykrywania przedmiotów o małych wymiarach. Ponadto jest konieczna dokładna znajomość funkcji jaskrawości poszukiwanego obiektu, a co więcej, filtry takie są bardzo wrażliwe na wymiary i położenie katowe szukanego obiektu. Stąd możliwość ich zastosowania jest ograniczona.

Wynika z powyższego konieczność zastosowania filtrów, które dają dobre rezultaty przy różnych kształtach przedmiotów na obrazie. Należą do nich filtry wykrywające krawędzie (rys. 5) i filtry wykrywające małe niejednorodności przypominające operator Laplace'a (rys. 6). Współczynniki filtrów na rysunkach są dobrane tak, że maksymalna wartość korelacji z prostą krawędzią przedmiotu (o maksymalnym kontraście) pokazaną na rys. 12a wynosi 1 w każdym przypadku. Wartości korelacji z prostą krawędzią z rys. 12a dla różnych filtrów krawędziowych pokazano na rys. 12b, a dla filtrów Laplace'a na rys. 13. W obydwu przypadkach



Rys. 12. a) Idealna krawędź, b) wartości korelacji tej krawędzi dla różnych filtrów krawędziowych w zależności od przesunięcia n w pikselach



Rys. 13. Wartości korelacji idealnej krawędzi dla filtrów Laplace'a w zależności od przesunięcia n w pikselach

zmienną niezależną jest przesunięcie n maski względem krawędzi wyrażone w pikselach.

Warto tu zauważyć, że w przypadku filtrów Laplace'a wartość maksimum korelacji jest nieznacznie przesunięta w stosunku do położenia krawędzi. Ten błąd rośnie wraz z wymiarami maski i pokazano go dla różnych masek w Tabeli 1. Należy też zaznaczyć, że maski te są w niniejszej pracy wykorzystywane niestandardowo, gdyż zazwyczaj w literaturze jako punkt krawędzi przyjmowane jest zero funkcji Laplace'a, a nie wartość maksymalna tej funkcji [10].

Filtry pierwszego rodzaju są bardzo wrażliwe na kierunek krawędzi i ich korelacja z krawędziami prostymi do tych, do których są dopasowane daje

Tabela 1
Błąd w określaniu położenia krawędzi dla
różnych masek Laplace'a

Wymiar maski	Błąd (piksele)
4 × 4 (2 × 2)	1
5 × 5 (3 × 3)	1.5
6 × 6 (2 × 2)	1
6 × 6 (4 × 4)	2
7 × 7 (3 × 3)	1.5
7 × 7 (5 × 5)	2.5
8 × 8 (4 × 4)	2
8 × 8 (6 × 6)	3

wartość zerową. Stąd, dla prawidłowego działania układu potrzebne są co najmniej dwa takie filtry dopasowane do krawędzi prostopadłych. Odpowiedni układ pokazano na rys. 14. Z kolei filtry Laplace'a mają podobną czułość dla krawędzi pionowych i poziomych i dają dla takich krawędzi jednakowe wartości funkcji korelacji. Wartość korelacji dla krawędzi ukośnych (45°) dla tych filtrów wynosi 60%—80% wartości korelacji dla krawędzi pionowych lub poziomych. Ilustruje to Tabela 2.

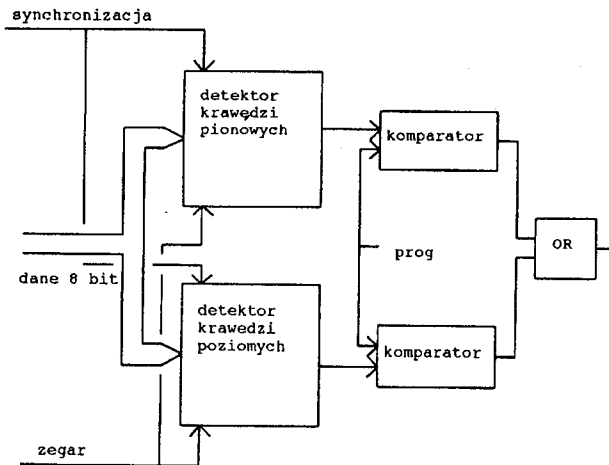
Tabela 2
Wartości korelacji krawędzi ukośnych z filtrami Laplace'a wyrażonej w procentach wartości korelacji z krawędzią pionową

Rodzaj maski	4 × 4	5 × 5	6 × 6/2 × 2	6 × 6/4 × 4	7 × 7/3 × 3	7 × 7/5 × 5	8 × 8/4 × 4	8 × 8/6 × 6
Wartość korelacji	75%	64%	83%	62,5%	71%	59%	69%	58%

Ważną cechą wszystkich filtrów jest ich odporność na zakłócenia. Dla oceny odporności różnych filtrów założymy, że do zawartości informacyjnej obrazu został dodany biały szum w taki sposób, że jasność każdego piksela jest zmienną losową o wariancji σ^2 , niezależną od wartości jasności jakiegokolwiek innego piksela. Wówczas, pamiętając o jednostkowej odpowiedzi każdego filtru na idealną krawędź, wariancja korelacji obrazu z daną maską jest miarą odporności tej maski na zakłócenia. Im ta wariancja jest mniejsza, tym odporność danego filtru (maski) na zakłócenia jest większa. Wariancja sumy niezależnych zmiennych losowych wziętych z wagami p_{ij} , odpowiadającymi współczynnikom filtru, jest równa.

$$\sigma_{\text{calc}}^2 = \sum_{i=1}^N \sum_{j=1}^N p_{ij}^2 \sigma^2 = \sigma^2 \sum_{i=1}^N \sum_{j=1}^N p_{ij}^2 = A \sigma^2, \quad (2)$$

$$A = \sum_{i=1}^N \sum_j p_{ij}^2,$$



Rys. 14. Schemat blokowy układu detekcji krawędzi w kierunkach prostopadłych

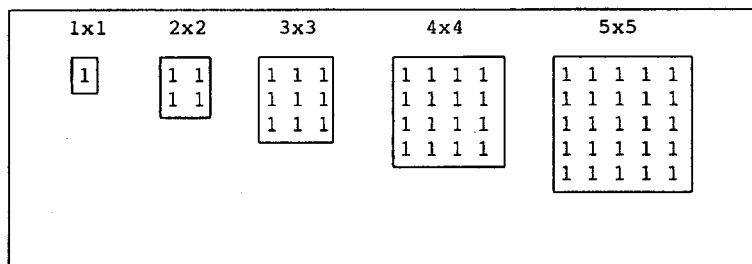
Wartość parametru A jest miarą odporności na zakłócenie. Odpowiednie wartości podano w Tabeli 3. Z zamieszczonych tam danych wynika, że odporność na zakłócenia rośnie wraz ze wzrostem wymiarów filtra (parametr A maleje ze wzrostem N). Widać też wyraźnie, że maski dopasowane do krawędzi dają o rząd wielkości mniejsze wariancje aniżeli maski Laplace'a, są więc odpowiednio bardziej odporne na zakłócenia.

Tabela 3

Miarą A odporności poszczególnych filtrów na zakłócenia

Wymiar filtra	A (filtry krawędziowe)	A (filtry Laplace'a)
3×3	0.67	
4×4	0.25	3
5×5	0.2	1.78
6×6	0.11	$6 \times 6(4 \times 4)$ 1.25 $6 \times 6(2 \times 2)$ 2.47
7×7	0.095	$7 \times 7(3 \times 3)$ 1.11 $7 \times 7(5 \times 5)$ 0.96
8×8	0.06	$8 \times 8(4 \times 4)$ 0.75 $8 \times 8(6 \times 6)$ 0.78

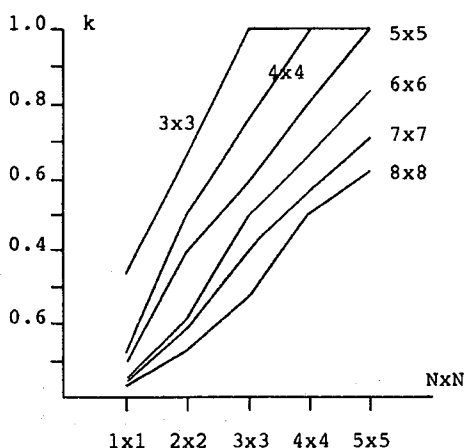
Z odpornością na szумы wiąże się również zdolność do wykrywania małych przedmiotów (niejednorodności). Dla porównania zdolności wykrywania takich obiektów przez poszczególne maski założyliśmy, że obiekty te mogą mieć zmienne

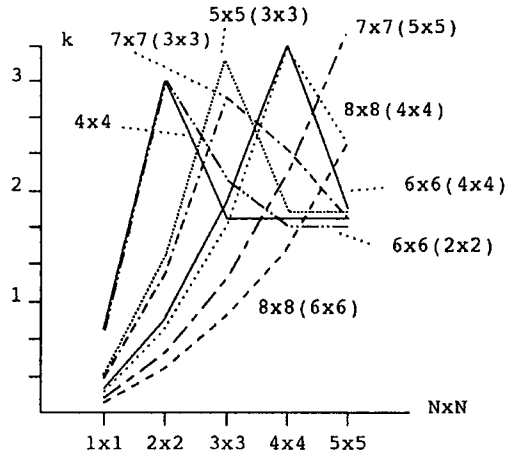


Rys. 15. Małe obiekty o różnych wymiarach

wymiary od 1×1 do 5×5 pikseli przy 100% kontraście w stosunku do tła. Pokazano to na rys. 15. Z kolei maksymalne wartości korelacji takich przedmiotów z maskami do wykrywania krawędzi są narysowane na rys. 16 a z maskami typu Laplace'a na rys. 17. Widać wyraźnie, że filtry Laplace'a dobrze nadają się do wykrywania takich przedmiotów, gdyż dają średnio o rząd wielkości większą korelację aniżeli filtry wykrywające krawędzie. W tym pierwszym przypadku można zauważyć, że funkcja korelacji jest tym większa, im wymiary niejednorodności są bliższe wymiarom centralnej części filtru, a zatem istnieje optymalna maska dla każdego wymiaru niejednorodności. Wynika to zresztą z własności filtracji dopasowanej. Filtry wykrywające krawędzie w związku ze swymi własnościami uśredniającymi mają większą odporność na szumy, ale źle wykrywają drobne przedmioty. W tym przypadku korelacja jest tym mniejsza, im większe są wymiary maski w porównaniu z wymiarami niejednorodności.

Osobnym zagadnieniem, ważnym dla śledzenia poruszających się na ekranie przedmiotów, jest zdolność do ich wykrywania przez filtry różnych typów i o róż-

Rys. 16. Maksymalne wartości korelacji k z maskami wykrywającymi krawędzie w zależności od wymiarów obiektu $N \times N$

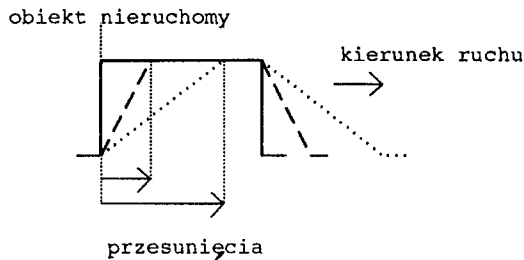


Rys. 17. Maksymalne wartości korelacji k z maskami Laplace'a w zależności od wymiarów obiektu $N \times N$

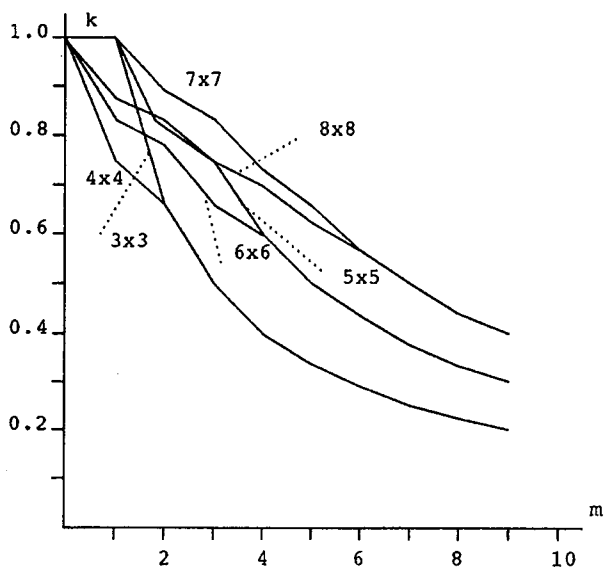
nych wymiarach. Poruszających się w trakcie danej ramki (półobrazu) przedmiot ma bowiem nieostre krawędzie na obrazie. Zjawisko to wynika z tego, że kamera CCD między odczytami całkuje padający na dany piksel strumień światła. Stąd jasność poruszającego się przedmiotu ulega zmianom pokazanym na rys. 18, a krawędzie przedmiotów są tym bardziej rozmyte im większe jest przesunięcie obiektu w trakcie trwania ramki (półobrazu).

Spadek maksymalnej wartości korelacji ruchomego obiektu (w tym wypadku krawędzi pionowej o 100% kontraście) z filtrami wykrywającymi krawędzie pokazano na rys. 19. Podobne dane dla filtrów Laplace'a podano na rys. 20.

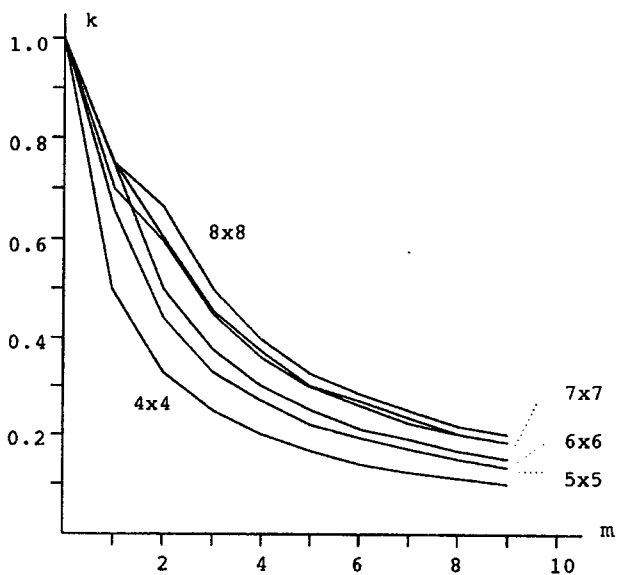
Widać z rysunków, że spadek korelacji może być bardzo znaczny, a korelacja maleje wraz ze wzrostem przesunięcia obiektu wyrażonego w pikselach na ramkę. W ogólności filtry o większych wymiarach dają większe wartości korelacji (są mniej wrażliwe na przesuw obiektów). Podobnie filtry krawędziowe są mniej wrażliwe na przesuw obiektu aniżeli filtry Laplace'a.



Rys. 18. Zmiany jasności poruszających się obiektów



Rys. 19. Maksymalna wartość korelacji k ruchomego obiektu dla filtrów wykrywających krawędzie w zależności od przesunięcia m w pikselach na ramkę



Rys. 20. Maksymalna wartość korelacji k ruchomego obiektu dla filtrów Laplace'a w zależności od przesunięcia m w pikselach na ramkę

Podsumowując rozważania tego rozdziału należy stwierdzić, że nie można wskazać jednego filtra najlepszego we wszystkich zastosowaniach. Wybór filtra musi być zawsze podyktowany rodzajem obiektów, które chcemy zlokalizować na obrazie. Spośród dwóch typów zaprezentowanych filtrów uniwersalnych, filtry krawędziowe lepiej nadają się do lokalizacji większych obiektów nawet na zaszumionych obrazach, podczas gdy filtry Laplace'a są lepsze w umiejscawianiu małych przedmiotów na obrazach niezaszumionych.

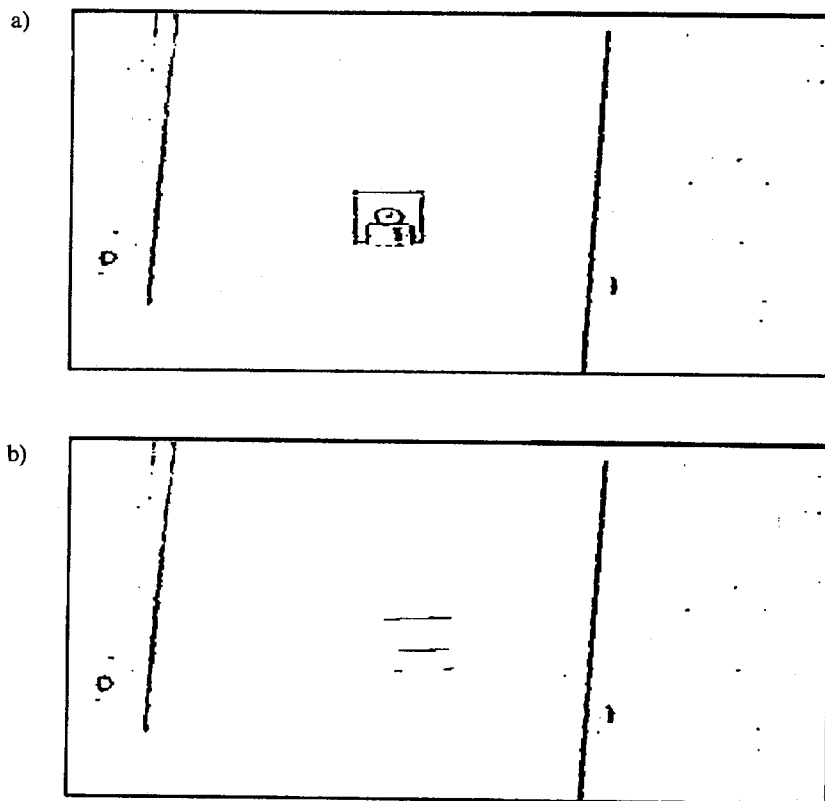
5. UZYSKANE REZULTATY

Zasadnicze znaczenie dla jakości pracy urządzenia okazał się mieć wybór filtra dwuwymiarowego oraz dobór poziomu progowego, powyżej którego korelacja uznawana była za korelację z krawędzią. Jeśli chodzi o pierwsze zagadnienie, to poziom szumów obrazu telewizyjnego wykluczał stosowanie filtrów Laplace'a o małych wymiarach. W przypadku stosowania takich filtrów układ wykrywał krawędzie w przypadkowych miejscach, uniemożliwiając poprawną pracę urządzenia. Dlatego ograniczono się do masek o największych wymiarach. Z kolei poziom progowy był dobierany indywidualnie dla każdego filtra dla uzyskania najlepszych rezultatów dla danej klasy obrazów. Przykłady obrazów krawędziowych uzyskanych za pomocą masek Laplace'a pokazano na rys. 7.

Innym problemem, poruszonym już w rozdziale 4, był spadek wartości sygnału na wyjściu filtra przy ruchomych obiektach. Miało to zasadniczy wpływ przy obróbce obrazów z takimi obiektami. Szybko poruszający się obiekt po prostu znikał z ekranu pomocniczego monitora. Zjawisko to zilustrowano na rys. 21: na rys. a) dyskietka jest nieruchoma podczas gdy na rys. b) porusza się w kierunku poziomym (dyskietkę powieszono na nitce i puszczono jak wahadło).

Na rys. 21b widać, że krawędzie prostopadłe do kierunku ruchu zniknęły z obrazu, podczas gdy krawędzie równoległe zostały zachowane. Znikanie niektórych krawędzi przy ruchu obiektu powodowało zawężenie okna śledzącego do części obiektu, a nawet całkowite jego zgubienie. Szybki ruch przedmiotu, nawet przy całkowitej widoczności jego krawędzi, powodował także opóźnienie okna śledzącego w stosunku do ruchu obiektu. Było to spowodowane tym, że w bieżącej ramce przeszukiwane było otoczenie położenia obiektu z poprzedniej ramki, podczas gdy śledzony obiekt zdążył już zmienić swoją pozycję. Rozwiązaniem tego problemu może być wykorzystanie predykcji położenia obiektu na podstawie jego uprzedniego ruchu.

Bardzo ważne w pracy urządzenia było właściwe zogniskowanie kamery i ustawienie odpowiedniej jasności obrazu. Przy braku zogniskowania (obraz nieostry) krawędzie nie były wykrywane poprawnie, podczas gdy średnia jaskrawość obrazu miała istotny wpływ na wartości uzyskiwanych korelacji.



Rys. 21. Obraz obiektu nieruchomego a) i szybko poruszającego się w kierunku poziomym b)

PODSUMOWANIE

Omawiany system pracował poprawnie, tj. umożliwiał wykrycie i śledzenie kontrastowych przedmiotów znajdujących się na w miarę jednolitym tle. Błędy śledzenia pojawiały się jedynie w przypadku szybkiego ruchu obiektu i jak już wspomniano, mogą one być przynajmniej częściowo wyeliminowane przy pomocy zastosowania predykcji położenia obiektu (np. filtracja Kalmana [11]). Ze względu na szумы zastosowanie filtrów Laplace'a o małych wymiarach nie dawało dobrych rezultatów. Z kolei filtry wykrywające krawędzie miały lepsze właściwości szumowe, ale potrzeba wykrywania krawędzi leżących wzdłuż obydwu kierunków prostopadłych prowadzi do konieczności zdublowania tych filtrów.

Przy mało kontrastowych obiektach lub niejednorodnym tle, układ nie pracował poprawnie, gdyż fragmenty tła dawały również przekroczenia poziomu progowego na wyjściu filtru. Przez to okno obserwacyjne rozszerzało się i obejmowało również wspomniane fragmenty tła, a w przypadku silnie niejednorodnego tła, zaczynało obejmować prawie cały ekran. Można próbować eliminacji takich zakłóceń na

drodze programowej przez zerowanie izolowanych pikseli. Innym wyjściem w tym przypadku jest zastosowanie filtru dopasowanego do poszukiwanego obiektu. W omawianym urządzeniu maksymalne wymiary takiego filtru nie są duże (8×8), jednak zastosowane w nim układy PDSP 16488 umożliwiają stosunkowo łatwe tworzenie filtrów o większych wymiarach (16×16 , 32×32) [8].

Układ nie był projektowany do jednego konkretnego rozwiązania, ale jest na tyle elastyczny, iż przez zmianę parametrów filtrów i odpowiedniego oprogramowania może znaleźć zastosowanie w wielu dziedzinach. Jako przykłady można podać tu sterowanie procesami produkcyjnymi, kontrolę jakości, kierowanie pojazdami, zastosowania wojskowe itd. [2].

BIBLIOGRAFIA

1. J. Siuzdak: *Wybrane metody segmentacji obrazów*. Elektronika, 1993, vol. 34, nr 2, str. 15–18
2. J. Siuzdak: *Wybrane zastosowania szybkiej segmentacji obrazów*. Przegląd Telekomunikacyjny, 1997, vol. 70, nr 1, str. 29–33
3. X. Zhang, G. Sexton: *Automatic pedestrian counting using image processing techniques*. Electronics Letters, 1995, vol. 31, nr 11, str. 863–865
4. C.K. Lee, W.S. Ng: *Real-time image segmentation system for IC die patterns*. Int. J. Electronics, 1995, vol. 78, nr 5, str. 959–973
5. D. Kuan i in.: *Autonomous Robotic Vehicle Road Following*. IEEE Trans. Pattern Analys. & Machine Intellig., 1998, vol. 10, nr 5, str. 648–658
6. S.M. Smith, J.M. Brady: *ASSET-2: Real-Time Motion Segmentation and Shape Tracking*. IEEE Trans. Pattern Analys. & Machine Intellig., 1995, vol. 17, nr 8, str. 814–820
7. P.K. Sahoo i in.: *A survey of thresholding techniques*. Computer Vision, Graphics & Image Processing 1988, vol. 41, str. 233–260
8. GEC Plessey: *Digital Video & Digital Signal Processing IC Handbook*, 1993
9. D. Middleton: *An Introduction to Statistical Communication Theory*. McGraw-Hill Inc. New York, 1960
10. T.Y. Young, K.S. Fu: *Handbook of Pattern Recognition and Image Processing*. Academic Press, Inc., Orlando, 1986.
11. B.D.O. Anderson, J.B. Moore: *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, 1979

J. SIUZDAK

REAL-TIME IMAGE SEGMENTATION SYSTEM

S u m m a r y

In the paper, there is described the operating principle of the image processing system that makes it possible to extract objects in real time. After A/D conversion the image intensity undergoes 2-D filtering in order to obtain object edges. This edge image is then transferred to the RAM of an IBM PC computer, where it is processed further. The system enables one to track the moving object position in real time. The selection of 2-D filter and the influence of noise and object movement are also discussed.

Key words: image processing system, 2-D filters.

Nowy rodzaj filtru dla detekcji krawędzi na obrazie

JERZY SIUZDAK, TOMASZ CZARNECKI

Instytut Telekomunikacji, Politechnika Warszawska

Otrzymano 1997.06.19

Autoryzowano 1997.09.21

Zaprezentowano nowy rodzaj filtru dwuwymiarowego przydatnego do wykrywania krawędzi przedmiotów na obrazach. Filtr ten maksymalizuje swoją korelację z dowolnie skierowaną krawędzią, przy jednoczesnej optymalizacji własności szumowych. W przypadku implementacji cyfrowej filtru błąd systematyczny określenia położenia krawędzi obiektu jest nieznaczny. Dla przykładu dla maski 8×8 nie przekracza 2 piksli. Zaprezentowano cyfrową implementację filtru. Pokazano efekt zastosowania filtru na przykładowym obrazie.

Słowa kluczowe: obróbka obrazu, segmentacja, filtracja dwuwymiarowa.

1. WSTĘP

Detekcja krawędzi jest jedną z form wstępnej obróbki obrazu. W jej trakcie uwydatniane zostają krawędzie obiektów. Może to być wykorzystane w dalszej obróbce obrazu, np. przy segmentacji lub kompresji zawartości informacyjnej. Wykrywania krawędzi dokonuje się zazwyczaj przez korelację obrazu z grupą masek [1]. Przegląd tych metod można znaleźć w pracy [2]. Przy sprzętowej obróbce obrazu wymagają one jednak dosyć złożonych układów [3]. Poniżej prezentujemy detektor krawędzi wykorzystujący tylko jeden filtr dwuwymiarowy. Filtry takie są dosyć łatwo dostępne w postaci układów scalonych [4], co zapewnia stosunkowo prostą implementację sprzętową.

2. TEORIA

Założmy, że rozpatrujemy pojedynczy obraz monochromatyczny i przyjmiemy, że rozkład jasności tego obrazu L dany jest funkcją dwóch zmiennych $g(x, y)$:

$$L = g(x, y) \quad \text{dla} \quad -L_x < x < L_x \quad -L_y < y < L_y, \quad (1)$$

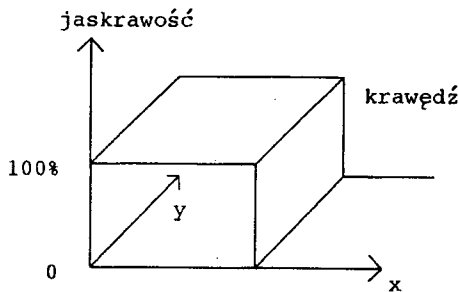
gdzie $2L_x, 2L_y$ są wymiarami obrazu wzdłuż odpowiednich współrzędnych. Jeśli filtr jest określony przez funkcję $f_{xy}(x, y)$, to w wyniku korelacji obrazu z tym filtrem otrzymamy nowy obraz, którego jasność $T(x, y)$ wyraża się zależnością

$$T(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x', y') f_{xy}(x' - x, y' - y) dx' dy'. \quad (2)$$

Dalsza część pracy jest poświęcona znalezieniu funkcji f_{xy} określającej filtr, która spełnia następujące postulaty:

1. Maksymalna korelacja filtru z idealną krawędzią nie zależy od kierunku tejże krawędzi.
2. Korelacja filtru ze stałym tłem jest zerowa.
3. Funkcja f_{xy} maksymalizuje wartość korelacji z idealną krawędzią, a jednocześnie filtr posiada dużą odporność na szumy.

Wyjaśnienia wymaga tu pojęcie idealnej krawędzi. Jest to prostoliniowa krawędź przedmiotu o 100% kontraście w stosunku do tła. Narysowano ją na rys. 1.



Rys. 1. Idealna krawędź

Zdefiniujmy matematycznie warunki 1... 3.

1. Jasne jest, że jeżeli maksymalna korelacja filtru ma być niezależna od kierunku krawędzi to funkcja $f_{xy}(x, y)$ powinna mieć symetrię kołową. Oznacza to, że jeżeli przedstawimy wspomnianą funkcję we współrzędnych biegunowych, to nie zależy ona od współrzędnej kątowej, tzn.

$$f_{xy}(x, y) = f(r, \varphi) = \begin{cases} f(r) & \text{dla } r \leq a \\ 0 & \text{dla } r > a. \end{cases} \quad (3)$$

Zerowanie się funkcji $f(r)$ dla $r > a$ wynika ze względów praktycznych, filtr

powinien mieć ograniczoną dziedzinę. W dalszej części ograniczymy się do poszukiwania postaci funkcji $f(r)$ dla $r < a$, która spełnia postawione założenia.

2. Warunek zerowania się odpowiedzi filtru na stałą wartość tła prowadzi do zależności

$$\int \int_{r \leq a} f_{xy}(x, y) dx dy = 0 \quad (4)$$

czyli

$$J_1 = \int_0^a r f(r) dr = 0 \quad (5)$$

Jak już wspomniano rozpatrywane filtry powinny, niezależnie od postaci funkcji $f(r)$, wykazywać jednakowe właściwości szumowe. Dla określenia tych własności posłużymy się pojęciem wariancji σ^2 sygnału na wyjściu filtru pobudzonego szumem białym o dwuwymiarowej gęstości widmowej mocy równej N_0 . Jeżeli pole szumów jest jednorodne, to wspomniana wariancja wyraża się zależnością

$$\begin{aligned} \sigma^2 = E \left[\int \int_{r \leq a} n(x, y) f_{xy}(x, y) dx dy \int \int_{r' \leq a} n(x', y') f_{xy}(x', y') dx' dy' \right] = \\ \int \int_{r \leq a} \int \int_{r' \leq a} E[n(x, y) n(x', y')] f_{xy}(x, y) f_{xy}(x', y') dx dy dx' dy' = \end{aligned} \quad (6)$$

$$N_0 \int \int_{r \leq a} \int \int_{r' \leq a} \delta(x - x', y - y') f_{xy}(x, y) f_{xy}(x', y') dx dy dx' dy' =$$

$$N_0 \int \int_{r \leq a} f_{xy}(x, y) dx dy = 2\pi N_0 \int_0^a r f^2(r) dr.$$

Tutaj $E()$ jest wartością oczekiwaną wyrażenia w nawiasach, zaś δ jest funkcją delta Diraca. Zatem dla maksymalizacji stosunku kwadratu wartości korelacji filtru z idealną krawędzią do wariancji wyjścia przy pobudzeniu szumem białym należy przyjąć określoną (stałą) wartość wariancji σ^2 i przy tym założeniu maksymalizować wartość bezwzględną korelacji filtru z idealną krawędzią.

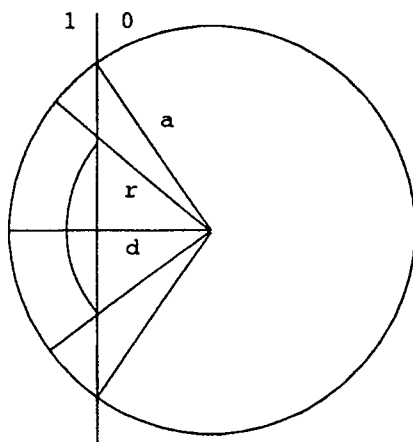
Na podstawie zależności (6) pierwszy warunek przyjmuje postać

$$J_2 = \int_0^a r f^2(r) dr = 1. \quad (7)$$

Na podstawie rys. 2 nietrudno jest wyznaczyć korelację filtru z krawędzią i drugi warunek jest następujący:

$$I = \int_a^d r f(r) 2 \arccos \frac{d}{r} dr = \min. \quad (8)$$

Tutaj d jest odległością krawędzi od środka filtru (patrz rys. 2). Zatem nasze zadanie sprowadza się do znalezienia funkcji $f(r)$, która minimalizuje funkcjonal I (8) przy



Rys. 2. Korelacja filtru z idealną krawędzią

ograniczeniach danych funkcjami J_1 (5) i J_2 (7). Warunek konieczny Kuhna-Tuckera przyjmuje wtedy postać [5], [6]:

$$\nabla I + \lambda_1 \nabla J_1 + \lambda_2 \nabla J_2 = 0, \quad (9)$$

gdzie ∇ oznacza gradient funkcjonu, zaś λ_1, λ_2 są mnożnikami Lagrange'a [6]. Z warunków zagadnienia wynika, że jeśli istnieje tylko jedna taka funkcja dla której jest spełniony warunek (9), to jest ona jednocześnie szukanym rozwiązaniem. Ponieważ funkcjonały I i J_1 są liniowe, a funkcjonal J_2 — kwadratowy, ich gradienty można stosunkowo łatwo wyliczyć [5]:

$$\nabla I = \begin{cases} \int_d^a rh(r) 2 \arccos \frac{d}{r} \frac{dr}{r} & \text{dla } r \geq d \\ 0 & \text{dla } r < d \end{cases} \quad (10)$$

$$\nabla J_1 = \int_0^a rh(r) dr$$

$$\nabla J_2 = \int_0^a 2rf(r)h(r) dr$$

gdzie $h(r)$ jest dowolną funkcją o jednostkowej normie w przedziale $(0, a)$. Z zależności (9) i (10) nietrudno jest znaleźć ogólną postać funkcji $f(r)$:

$$f(r) = \begin{cases} -\frac{\lambda_1}{2\lambda_2} & \text{dla } r \leq d \\ -\frac{\lambda_1 + 2 \arccos \frac{d}{r}}{2\lambda_2} & \text{dla } a \geq r \geq d. \end{cases} \quad (11)$$

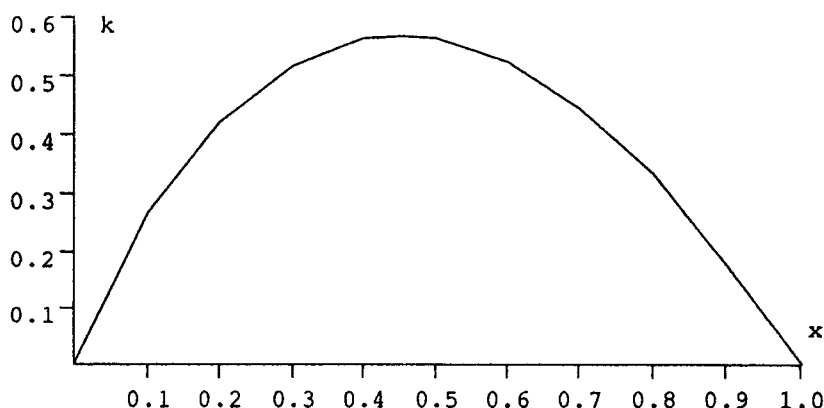
Mnożniki Lagrange'a można wyznaczyć wstawiając (11) do zależności (5), (7) i licząc odpowiednie całki. W wyniku obliczeń dostajemy

$$\begin{aligned}\lambda_1 &= 2(x\sqrt{1-x^2} - \arccos x), \\ \frac{\lambda_2}{a^2} &= -x^2(\ln x + 0.5(1-x^2)), \\ \text{gdzie } x &= \frac{d}{a}.\end{aligned}\quad (12)$$

Wartość funkcjonau (8), czyli wartość korelacji z idealną krawędzią, zależy od parametru $x = d/a$. Z wzorów (8), (11), (12) otrzymujemy

$$\frac{I}{a} = -\sqrt{2}\sqrt{-2x^2\ln x - x^2(1-x^2)}.\quad (13)$$

Przebieg powyższej zależności pokazano na rys. 3. Z rysunku widać, że korelacja przyjmuje wartość ekstremalną dla $x = 0.45$. Taki wynik wskazuje na to, że minimalizacja korelacji filtru z krawędzią występuje dla odległości krawędzi od



Rys. 3. Zależność sygnału na wyjściu filtru k od położenia krawędzi: $x = d/a$

środku filtru równej mniej więcej 1/4 średnicy filtru. Oznacza to, że w przypadku oceny położenia krawędzi za pomocą filtru popełniany jest pewien błąd systematyczny. Nie jest on jednak zbyt duży, gdyż rozmiary filtrów są ograniczone. Przebieg samej funkcji $f(r)$ pokazano na rys. 4 dla $x = 0.45$.

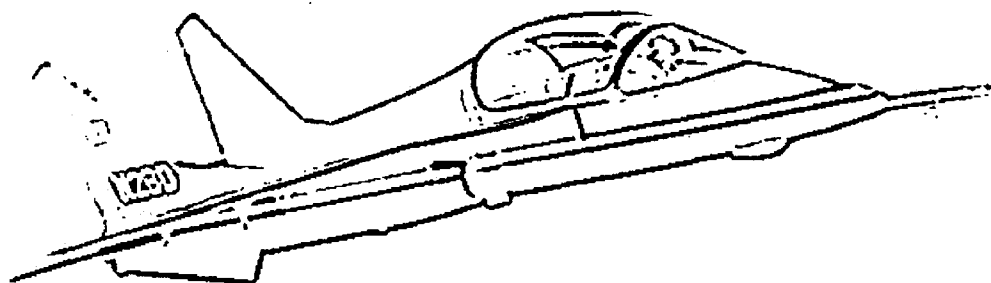
3. ZASTOSOWANIE

W obróbce cyfrowej obrazu stosuje się filtry (maski) o skończonych wymiarach; konieczna jest również dyskretyzacja samego filtru. Dane przykładowego filtru przybliżającego funkcję $f(r)$ z rys. 4 pokazano w Tabeli 1, zaś efekt zastosowania

a)



b)



Rys. 5. Wynik działania filtru z Tabeli 1: a) obraz wyjściowy, b) obraz przekształcony przez filtr i poddany komparacji z progiem

Tabela 1

Filtr cyfrowy 8×8 przybliżający filtr optymalny

0	0	-1.5	-1.5	-1.5	-1.5	0	0
0	-1.5	-0.25	-0.25	-0.25	-0.25	-1.5	0
-1.5	-0.25	1	2.5	2.5	1	-0.25	-1.5
-1.5	-0.25	2.5	2.5	2.5	2.5	-0.25	-1.5
-1.5	-0.25	2.5	2.5	2.5	2.5	-0.25	-1.5
-1.5	-0.25	1	2.5	2.5	1	-0.25	-1.5
0	-1.5	-0.25	-0.25	-0.25	-0.25	-1.5	0
0	0	-1.5	-1.5	-1.5	-1.5	0	0

tego filtru do kilku obrazów — na rys. 5. Obrazy te zostały otrzymane przez dokonanie korelacji z maską z Tabeli 1 i porównanie wyników z progami o wartości dobranej empirycznie. Widać, że filtr dobrze spełnia swoje zadanie, gdyż obraz wyjściowy jest obrazem krawędzi obiektów na obrazie.

PODSUMOWANIE

Zaprezentowano nowy rodzaj filtru dwuwymiarowego przydatnego do wykrywania krawędzi na obrazach. Filtr ten maksymalizuje swoją korelację z dowolnie skierowaną krawędzią, przy jednoczesnej optymalizacji własności szumowych. W przypadku implementacji cyfrowej błąd systematyczny określenia położenia krawędzi nie przekracza 2 pikseli przy wymiarach maski nie przekraczających 8×8 .

BIBLIOGRAFIA

1. R.C. Gonzales, P. Wintz: *Digital Image Processing*. Addison-Wesley, Reading, Massachusetts, 1987
2. J. Siuzdak: *Wybrane metody segmentacji obrazów*. Elektronika 1993, vol. 34, nr 2, str. 15–18
3. J. Siuzdak: *System segmentacji obrazu i śledzenie obiektów w czasie rzeczywistym*. Kwart. Elektr. i Telekom. 1997, t. 43, z. 4, ss.
4. GEC Plessey Semiconductors: *Digital Video & Digital Signal Processing-IC Handbook*. 1993
5. L.E. Franks: *Teoria sygnałów*. PWN, Warszawa 1975
6. W. Findeisen, J. Szymanowski, A. Wierzbicki: *Teoria i metody obliczeniowe optymalizacji*. PWN, Warszawa 1980

J. SIUZDAK, T. CZARNECKI

A NOVEL FILTER FOR EDGE DETECTION

S u m m a r y

Described is a novel type of 2-D filter which may be used for object edge detection in image processing. The filter maximizes its correlation with an edge of any direction while having the best noise properties. When digitally implemented it has a small systematic error of edge position determination. The error does not exceed 2 pixels for a mask of 8×8 pixels size. A digital implementation of the filter is presented. An example of the filter application on an image is shown.

Key words: image processing, segmentation, 2-D filtering.

Projektowanie szerokopasmowych zwrotnic mikrofalowych z zastosowaniem tożsamości Kurody

HENRYK GRUCHAŁA

Wojskowa Akademia Techniczna, Warszawa

Otrzymano 1997.05.14

Autoryzowano 1997.09.24

W artykule przedstawiono metodykę transformacji filtrów dolnoprzepustowych i pasmowo-przepustowych, jako elementów składowych zwrotnic mikrofalowych. Do transformacji tej zastosowano tożsamości Kurody. Przeprowadzono dyskusję struktur zwrotnic z zastosowaniem filtrów pasmowo-przepustowych. Opisano szeregowe połączenie obwodów mikrofalowych realizowanych w technice NLP. Zaprezentowano przykłady szerokopasmowych zwrotnic z zastosowaniem szeregowego połączenia filtrów oraz ekwiwalentne im układy z zastosowaniem przesuwników fazy.

Słowa kluczowe: filtry dolnoprzepustowe, filtry pasmowo-przepustowe, zwrotnice mikrofalowe, tożsamości Kurody.

1. WPROWADZENIE

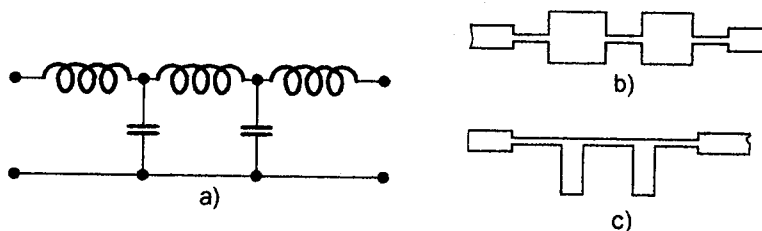
W technice linii paskowych (mikropaskowych) najczęściej stosowane są filtry dolnoprzepustowe, lub pasmowoprzepustowe. Filtry górnoprzepustowe stosowane są bardzo rzadko. Wynika to ze skali trudności podczas ich wytwarzania. W pewnych przypadkach filtry pasmowoprzepustowe mogą zastępować filtry górnoprzepustowe. Niektóre źródła zaliczają filtry pasmowoprzepustowe o pasmie powyżej dwóch oktaw do filtrów górnoprzepustowych. Przeprowadzona analiza poparta wynikami praktycznie wykonanych układów dowiodła, że najszersze pasma przepustowe wśród mikrofalowych filtrów pasmowoprzepustowych posiadają filtry ze sprzężeniami bezpośrednimi. Filtry te są jednocześnie elektrycznymi ekwiwalentami pasmowoprzepustowych filtrów mikrofalowych złożonych z odcinków linii transmisyjnych (w tym linii sprzężonych) o jednakowych długościach. Strukturę tych filtrów można transformować stosując drugą tożsamość Kurody. Transformację należy prowadzić

w taki sposób, by uzyskać realizowalną strukturę układu. W podobny sposób transformować można struktury filtrów dolnoprzepustowych. Znajduje tu zastosowanie pierwsza tożsamość Kurody.

Ponieważ filtry są podstawowymi elementami zwrotnic, w skład których wchodzi filtry dolno- i górnoprzepustowe lub dolno- i pasmowoprzepustowe albo tylko filtry pasmowoprzepustowe, podczas projektowania stosuje się tożsamości Kurody dla każdego z tych filtrów. Wyprowadzenie zależności analitycznych dotyczących przekształceń Kurody załączono w dodatku.

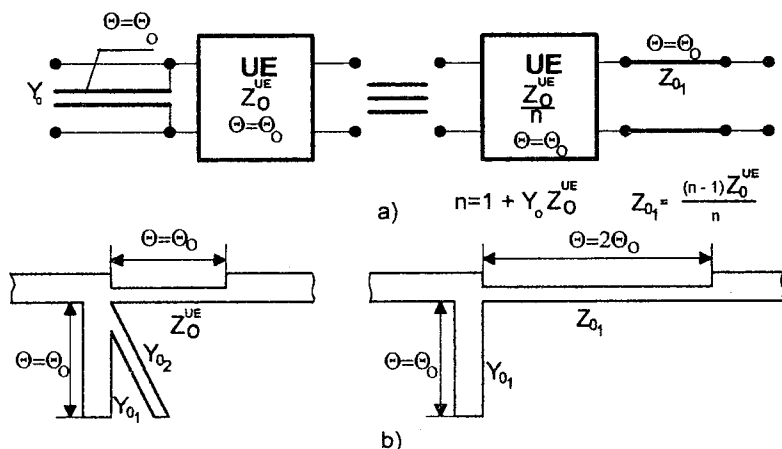
2. TRANSFORMACJA STRUKTUR FILTRÓW

Filtr dolnoprzepustowy o charakterystyce maksymalnie płaskiej lub równomiernej falistej składa się z szeregowo włączonych indukcyjności i równolegle włączonych pojemności, jak przedstawiono to na rys. 1a. Jego ekwiwalentami w zakresie mikrofal są układy przedstawione na rys. 1b i c. W wyniku zastosowania odcinków linii o skończonych długościach, parametry elektryczne tych filtrów różnią się nieco od odpowiadających im parametrom układów na stałych skupionych. Długości linii wchodzących w skład filtrów wpływają na nachylenie zboczy charakterystyk przenoszenia, poziom zafalowania i szerokość pasma zaporowego. Filtr przedstawiony na rys. 1c może być poddany transformacji zgodnie z rys. 2 [9]. Transformacja ta dotyczy struktur filtrów o jednakowych długościach elektrycznych linii szeregowych i równoległych. Linie szeregowo spełniają rolę elementów jednostkowych (UE) [9, 4].



Rys. 1. Pięcioelementowy dolnoprzepustowy filtr na stałych skupionych (a) i jego realizacja za pomocą kaskadowo połączonych odcinków linii wysoko- i niskoowowych (b) oraz szeregowych linii wysokoomowych i rozwartych linii równoległych (c)

W zależności od przyjętej wartości współczynnika transformacji „ n ” uzyskać można różne wartości impedancji linii transmisyjnych połączonych kaskadowo z elementem jednostkowym. Szczególnym przypadkiem jest transformacja ze współczynnikiem $n = 2$. Wówczas impedancja charakterystyczna linii połączonej kaskadowo z elementem jednostkowym jest równa impedancji tego elementu jednostkowego. Wywołuje to dwukrotne wydłużenie linii realizującej element jednostkowy, jak przedstawiono to na rys. 2b.



Rys. 2. Zastosowanie tożsamości Kurody (a) do transformacji struktury filtra dolnoprzepustowego (b)

Współczynnik transformacji n przyjmuje wartość $n = 2$, kiedy $Y_{02} = \frac{1}{Z_0^{UE}}$ (rys. 2b). Wówczas zachodzą zależności:

$$Z_{01} = \frac{Z_0^{UE}}{2} \quad (1a)$$

$$Z_0^{UE} = \frac{Z_0^{UE}}{2} \cdot \quad (1b)$$

Transformacja ze współczynnikiem $n = 2$ zachodzi, kiedy

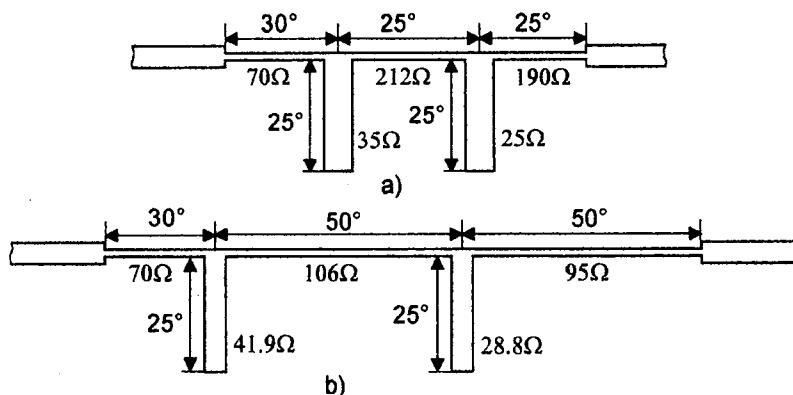
$$Y_{01} = Y_0 - \frac{1}{Z_0^{UE}} \quad (2)$$

W wyniku zastosowania przedstawionej wyżej transformacji zachodzi:

- dwukrotne wydłużenie i dwukrotne zmniejszenie impedencji charakterystycznych linii włączonych szeregowo,
- zwiększenie impedencji charakterystycznych linii włączonych równolegle.

Podczas transformacji struktury filtra można stosować inne wartości współczynnika $n \neq 2$. Wówczas linie włączone szeregowo (połączone kaskadowo) składać się będą z dwóch odcinków o różnych impedancjach.

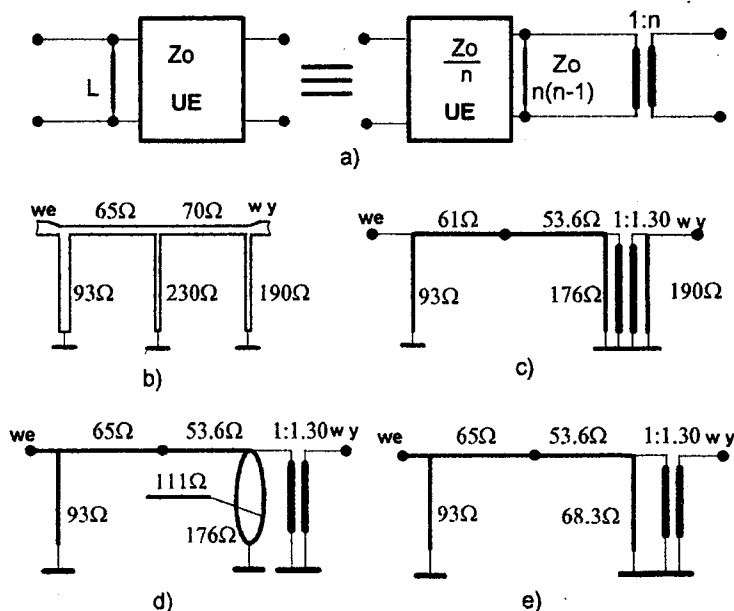
Dla przykładu rozpatrzono strukturę filtra dolnoprzepustowego o długościach elektrycznych linii na górnej częstotliwości granicznej $\Theta = 25^\circ$ i o impedancjach charakterystycznych, jak na rys. 3a. Występujące tu linie „szeregowo” o impedancjach charakterystycznych równych odpowiednio 212Ω i 190Ω są w technice NLP praktycznie nierealizowalne. Stosując zależności (1) i (2) zgodnie z procedurą przedstawioną na rys. 2 uzyskano linie o impedancjach charakterystycznych mieszczących się w przedziale $(28.8 - 106) \Omega$, co pozwala na pełną ich realizację w ww.



Rys. 3. Transformacja struktury filtra dolnoprzepustowego z zastosowaniem procedury przedstawionej na rys. 2; a) układ przed transformacją, b) układ po transformacji

technice (rys. 3b). Przeprowadzona transformacja nie zmieniła szerokości pasma zaporowego filtra.

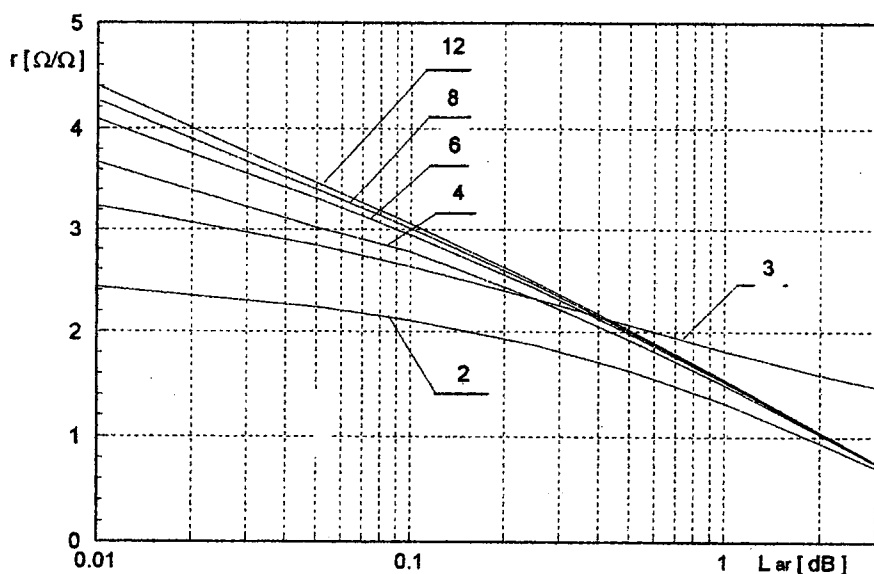
Transformację struktury filtra pasmowoprzepustowego przedstawiono na rys. 4. Linie realizujące równoległe rezonatory, o impedancjach charakterystycznych równych odpowiednio $230\ \Omega$ i $190\ \Omega$ w technice linii paskowych są nierealizowalne.



Rys. 4. Druga tożsamość Kurody (a) oraz przykład transformacji układu (b), etapy transformacji (c, d) oraz postać końcowa (e)

Transformując środkowy rezonator w kierunku wyjścia układu (w prawo) maleje impedancja charakterystyczna linii realizującej ten obwód oraz wzrasta wartość impedancji wyjściowej pełnego układu (rys. 4c). Z kolei przeniesienie rezonatora wyjściowego (poprzez transformator) do płaszczyzny włączenia przetransformowanego drugiego rezonatora, odpowiada równoległemu połączeniu linii realizujących te rezonatory (rys. 4d). W układzie wyjściowym przedstawionym na rys. 4e wartości realizowanych impedancji charakterystycznych zawierają się w przedziale (53.6 – 93) Ω .

Podczas transformacji ulega zmianie wartość impedancji wyjściowej filtru. Te własności znajdują zastosowanie podczas projektowania filtrów jako elementów składowych zwrotnic. W filtrach tych wyłącza się dwa wyjściowe rezonatory, w wyniku czego uzyskuje się charakterystyki impedancji wyjściowej odpowiadające charakterystynom górnoprzepustowych filtrów jednostronnie obciążonych (dla zbocza wznoszącego) i filtrom dolnoprzepustowym obciążonych jednostronnie (dla zbocza opadającego). W strukturze takiej zachodzi zmiana



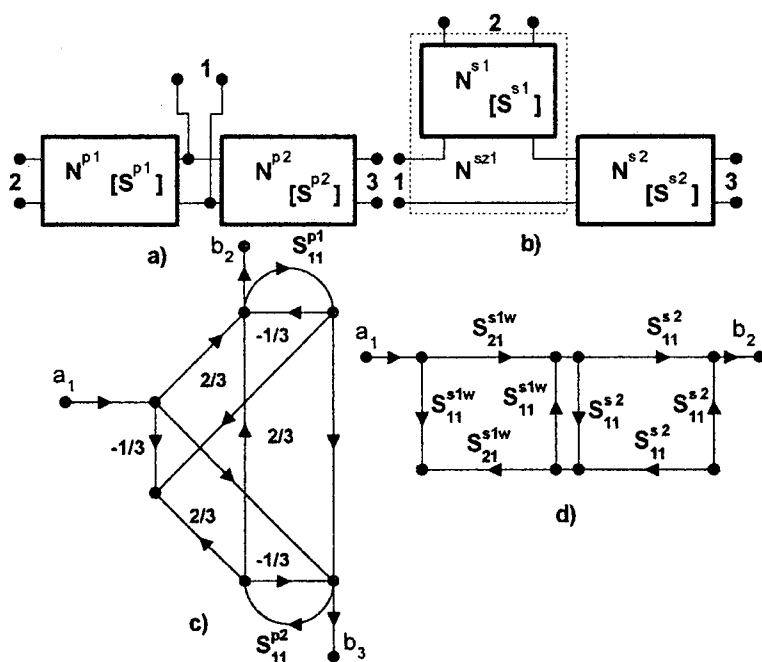
Rys. 5. Zależności unormowanych impedancji wyjściowych filtrów ze sprzężeniami bezpośrednimi po wyłączeniu dwóch rezonatorów w funkcji poziomu zafalowania i rzędu filtru

impedancji wyjściowej. Jej unormowane wartości w funkcji rzędu filtru i poziomu zafalowania przedstawiono na rys. 5. Transformację impedancji do poziomu $r=1$ uzyskuje się poprzez zastosowanie drugiej tożsamości Kurody [9, 3]. W układach zwrotnic obwody transformujące dodatkowo spełniać mogą rolę przesuwników fazy.

3. RÓWNOWAŻNE POŁĄCZENIA DWUWROTNIKÓW

Podstawowymi elementami zwrotnic, w tym i zwrotnic mikrofalowych, są filtry o odpowiednio dobranych charakterystykach przenoszenia [6, 4]. Filtry wchodzące w skład układów rozdziału częstotliwości (zwrotnic) mogą być połączone szeregowo lub równolegle. Połączenia te mogą być ekwiwalentne przy odpowiednio dobranych charakterystykach filtrów. Poniżej przedstawiono warunki, którym powinny odpowiadać dwuwrotniki (filtry), wchodzące w skład układu, by charakterystyki odbiciowe układu były jednakowe (w przypadku połączenia szeregowego lub równoległego).

W odniesieniu do zwrotnic rozważania można ograniczyć do analizy charakterystyki odbiciowej we wrotach wejściowych. Można postawić tezę, że dla zadanej charakterystyki dwóch dwuwrotników połączonych równolegle, istnieją dwa dwuwrotniki połączone szeregowo zapewniające uzyskanie takiej samej charakterystyki



Rys. 6. Połączenie dwóch dwuwrotników: równoległe (a) i ekwiwalentne szeregowe (b) oraz ich grafy przepływu sygnału (c) i (d)

odbiciowej we wrotach wejściowych układu. Równoważne połączenie dwóch dwuwrotników oraz ich grafy przepływu sygnału [5, 8] przedstawiono na rys. 6. Zastosowane tu oznaczenia N^{p1} i N^{p2} oraz N^{s1} i N^{s2} dotyczą układów przeznaczonych do połączenia równoległego (N^{p1} , N^{p2}) i połączenia szeregowego (N^{s1} , N^{s2}). Wypadkowy współczynnik odbicia w wrotach wejściowych (1) w przypadku połączenia równoległego, na podstawie grafu przedstawionego na rys. 6c, wynosi:

$$S_{11}^p = - \frac{1 - S_{11}^{p1} - S_{11}^{p2} - 3S_{11}^{p1}S_{11}^{p2}}{2 + S_{11}^{p1} + S_{11}^{p2} - S_{11}^{p1}S_{11}^{p2}}. \quad (4)$$

Połączenie szeregowe dwóch dwuwrotników jest równoważne połączeniu kaskadowemu dwuwrotnika zastępczego N^{sz1} z dwuwrotnikiem N^{s3} (tu dwuwrotnik N^{s1} potraktowano jako jednowrotnik, nie wpływa na wartość wejściowego współczynnika odbicia). Macierz rozproszenia dwuwrotnika N^{sz1} ma postać:

$$[S^{sz1}] = \frac{1}{3 - S_{11}^{s1}} \begin{bmatrix} 1 + S_{11}^{s1} & 2 - S_{11}^{s1} \\ 2 - S_{11}^{s1} & 1 + S_{11}^{s1} \end{bmatrix}. \quad (5)$$

Graf przepływu sygnału (rys. 6.4c) odpowiadający szeregowemu połączeniu dwuwrotników pozwala na wyznaczenie wejściowego współczynnika odbicia we wrotach wejściowych układu, który wynosi:

$$S_{11}^s = \frac{1 + S_{11}^{s1}}{3 - S_{11}^{s1}} + \frac{\left(2 \frac{1 - S_{11}^{s1}}{3 - S_{11}^{s1}}\right)^2 S_{11}^{s2}}{1 - \frac{1 + S_{11}^{s1}}{3 - S_{11}^{s1}} S_{11}^{s2}}. \quad (6a)$$

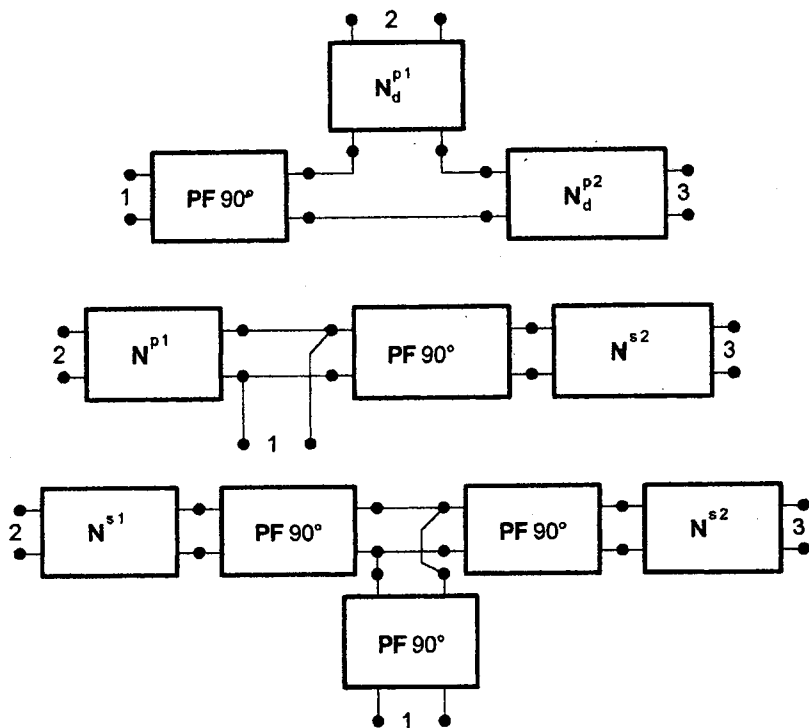
Po przekształceniach zależność ta ma postać:

$$S_{11}^s = \frac{1 + S_{11}^{s1} + S_{11}^{s2} - 3S_{11}^{s1}S_{11}^{s2}}{3 - S_{11}^{s1} - S_{11}^{s2} - S_{11}^{s1}S_{11}^{s2}}. \quad (6b)$$

Porównując wzory (4) i (6) można wyciągnąć wniosek, że połączenie szeregowe dwuwrotników równoważne jest połączeniu równoległemu dwuwrotników dualnych z włączonym dodatkowo (na wejściu układu) przesuwnikiem fazy 90° *. Zatem układy odpowiadające równoległemu połączeniu dwóch dwuwrotników mogą mieć strukturę, jak na rys. 7. Zastosowane tu dwuwrotniki dualne N_d^{pi} odpowiadają dwuwrotnikom przeznaczonym do połączenia równoległego obwodów.

Użycie przesuwników fazy rozbudowuje strukturę układu i ogranicza pasmo pracy układu do pasma przenoszenia przesuwników fazy. W przypadku układów wąskopasmowych rolę przesuwników może spełniać ćwierćfalowy odcinek linii transmisyjnej. Zastosowanie inwerterów impedancji/admitancji jest możliwe w układach, w których wypadkowe impedancje charakterystyczne (gałęzi równoległej inwertora i równoległego rezonatora wejściowego) są dodatnie.

* Przesuwnik opisany jest współczynnikami rozproszenia równymi: $S_{11} = S_{22} = 0$, $S_{12} = S_{21} = e^{-j\pi/2}$.

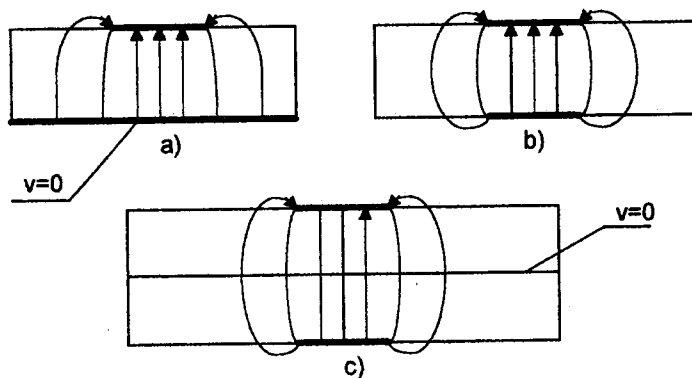


Rys. 7. Równoważne połączenia dwóch dwuwrotników odpowiadające połączeniu równoległemu dwuwrotników. Oznaczenia: $PF\ 90^\circ$ — przesuwnik fazy 90° , N_d^{pi} — dwuwrotnik dualny względem dwuwrotnika przeznaczony do połączenia szeregowego

4. SZEREGOWE POŁĄCZENIE OBWODÓW MIKROFALOWYCH W TECHNICIE PASKOWEJ

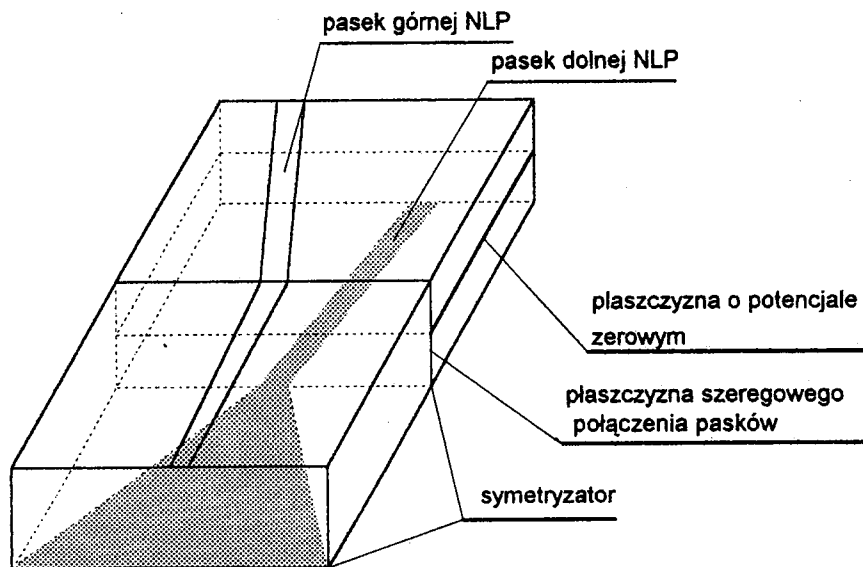
Niesymetryczna linia paskowa (rys. 8a) charakteryzuje się niejednakowym rozkładem prądów w pasku i w przewodniku tworzącym masę. Zatem przewodniki tworzące taką prowadnicę falową nie mogą pobudzać dwóch innych pasków („gorących”) o potencjałach różnych od zera. Ograniczenie to nie dotyczy dwuprzewodowej linii taśmowej o przekroju, jak na rys. 8b. Tu uzyskujemy jednakowy rozkład prądów w pasku górnym i w pasku dolnym, co pozwala na pobudzenie dwóch układów o niezależnych paskach „gorących”. W środku dielektryka rozdzielającego paski, potencjał jest równy zero. Odpowiada to występowaniu w tej płaszczyźnie ścianki elektrycznej. Zatem umiejscowienie w płaszczyźnie symetrii płaszczyzny o potencjale zerowym nie zaburzy rozkładu pola elektrycznego (rys. 8c).

W tej strukturze występują dwie prowadnice falowe składające się z paska górnego i dolnego oraz wspólnej masy. Linie te względem siebie połączone są szeregowo. Chcąc zatem zrealizować połączenie dwóch układów mikrofalowych



Rys. 8. Rozkład linii sił pola elektrycznego w niesymetrycznej linii paskowej (a), w linii symetrycznej dwuprzewodowej (b) i w linii symetrycznej dwuprzewodowej z ścianką elektryczną (masą) w środku struktury (c)

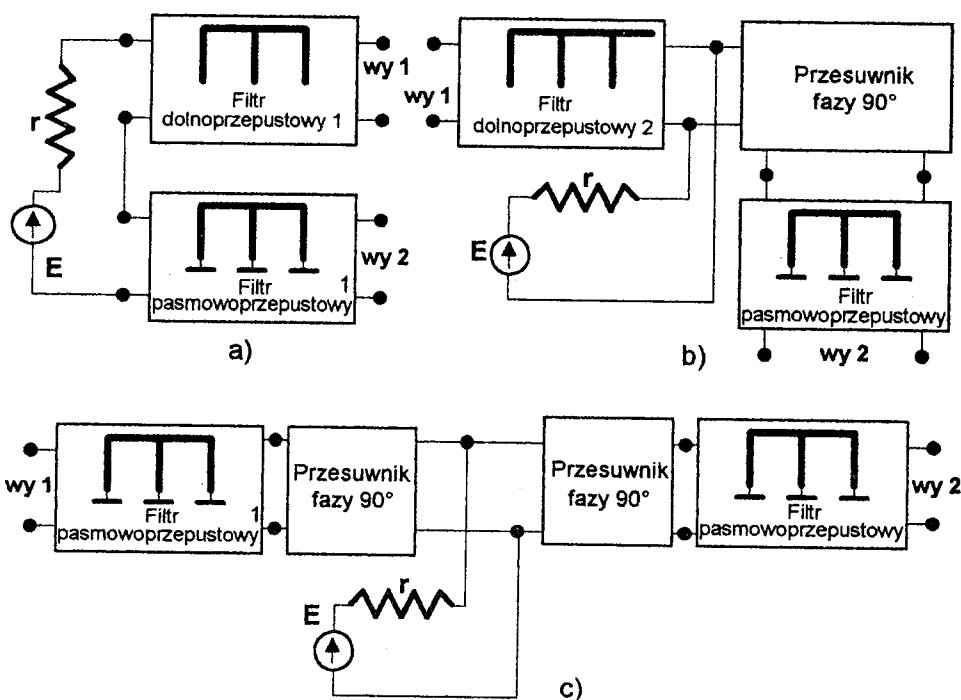
wykonanych w technice niesymetrycznej linii paskowej (NLP), należy na wejściu układu zastosować symetryzator, jak przedstawiono to na rys. 9. Wyjście symetryzatora połączone jest z strukturą dwuwarstwową o wspólnej masie i paskach umiejscowionych na przeciwnych płaszczyznach. W płaszczyźnie połączenia symetryzatora i struktury dwuwarstwowej występuje szeregowe połączenie dwóch przewodów falowych.



Rys. 9. Połączenie dwuprzewodowej linii symetrycznej z dwoma niesymetrycznymi liniami paskowymi połączonymi szeregowo

5. STRUKTURY DIPLEKSORÓW MIKROFALOWYCH Z ZASTOSOWANIEM FILTRÓW PASMOWOPRZEPUSTOWYCH

Biorąc pod uwagę nieskomplikowaną technologię wytwarzania oraz parametry elektryczne analizowanych filtrów, optymalną wydaje się być zwrotnica mikrofalowa składająca się z filtrów dolnoprzepustowych i pasmowoprzepustowych lub dwóch filtrów pasmowoprzepustowych. Struktury dipleksorów z zastosowaniem filtrów dolnoprzepustowych i pasmowoprzepustowych przedstawiono na rys. 10. Zastosowane oznaczenia filtrów dolnoprzepustowych dotyczą ich struktur: 1 — filtr z pierwszą równoległą pojemnością, 2 — filtr z pierwszą szeregową indukcyjnością.



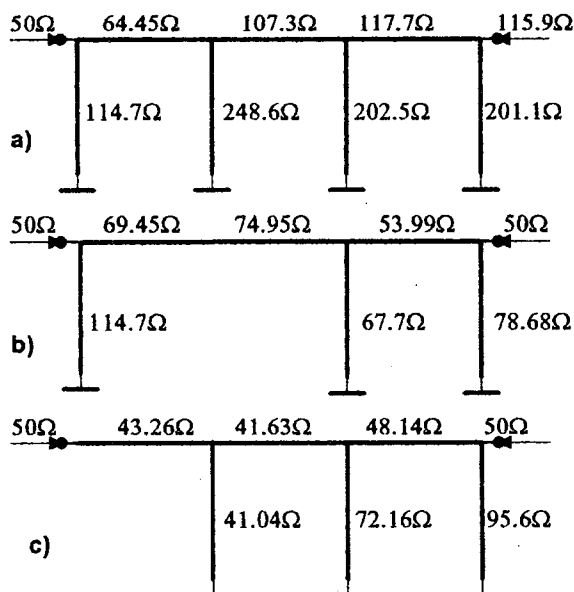
Rys. 10. Struktury zwrotnic z zastosowaniem filtrów pasmowoprzepustowych: a) z szeregowym połączeniem filtrów, b) z równoległym połączeniem filtrów i jednym przesuwnikiem fazy, c) z dwoma filtrami pasmowoprzepustowymi i dwoma przesuwnikami fazy

Zwrotnice z szeregowym połączeniem filtrów: dolnoprzepustowego i pasmowoprzepustowego (rys. 10a) zapewniają rozdział częstotliwości w szerokim pasmie [3]. Szerokość pasm wydzielanych kanałów częstotliwości zdeterminowana jest przez parametry elektryczne zastosowanych filtrów. Na charakterystykę wejściową zwrotnicy wpływają dodatkowo charakterystyki symetryzatora zastosowanego na wejściu układu. Szerokość wydzielanych pasm częstotliwości w układach z zastosowaniem przesuwników fazy 90° (rys. 8b i c) jest dodatkowo ograniczone przez pasma przenoszenia zastosowanych przesuwników. Praktycznie maksymalna szerokość

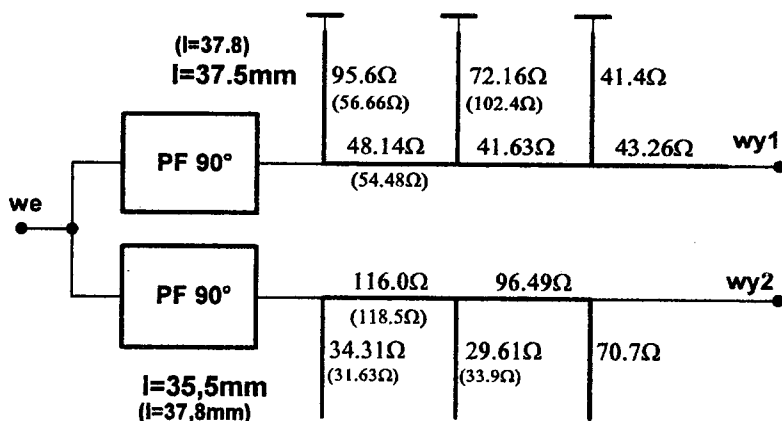
pasma przesuwników fazy przekracza dwie oktawy $((f_2/f_1)_{\max} = \sim 6)$ [7]. Ponieważ przesuwniki fazy praktycznie mają dolną częstotliwość pasma pracy różną od zera, w układzie przedstawionym na rys. 10b zachodzi dodatkowe ograniczenie pasm dolnego kanału. W przypadku zastosowania w układzie zwrotnicy tylko filtrów pasmowoprzepustowych efekt ten nie występuje. Mogą tu też znaleźć zastosowanie transformatory impedancji i korektory fazy projektowane w oparciu o [1, 2]. Dla wąskich pasm przenoszenia filtrów rolę przesuwników pełnić mogą ćwierćfalowe odcinki linii.

6. PRZYKŁADY ZWROTNIC

Rozpatrzmy zwrotnicę złożoną z filtru dolnoprzepustowego i pasmowoprzepustowego. Niech stopień filtru dolnoprzepustowego i poziom zafalowania oraz długość elektryczna linii na górnej częstotliwości granicznej wynoszą odpowiednio: $n = 5$, $L_{ar} = 0.25$ dB, $\Theta = 45^\circ$. Filtrowi temu odpowiada filtr pasmowoprzepustowy czwartego rzędu o poziomie zafalowania $L_{ar} = 0.25$ dB i współczynniku pokrycia pasma $f_2/f_1 = 3$. Prototyp filtru spełniający te założenia przedstawiono na rys. 11a. Filtr ten należy poddać transformacji w taki sposób, by impedancja wyjściowa wynosiła 50Ω oraz by wartości impedancji charakterystycznych linii mieściły się w obszarach ich realizowalności. Wyniki takiej transformacji przedstawiono na rys. 11b i c. Zatem zwrotnica złożona z dwóch rozpatrywanych filtrów może mieć strukturę, jak

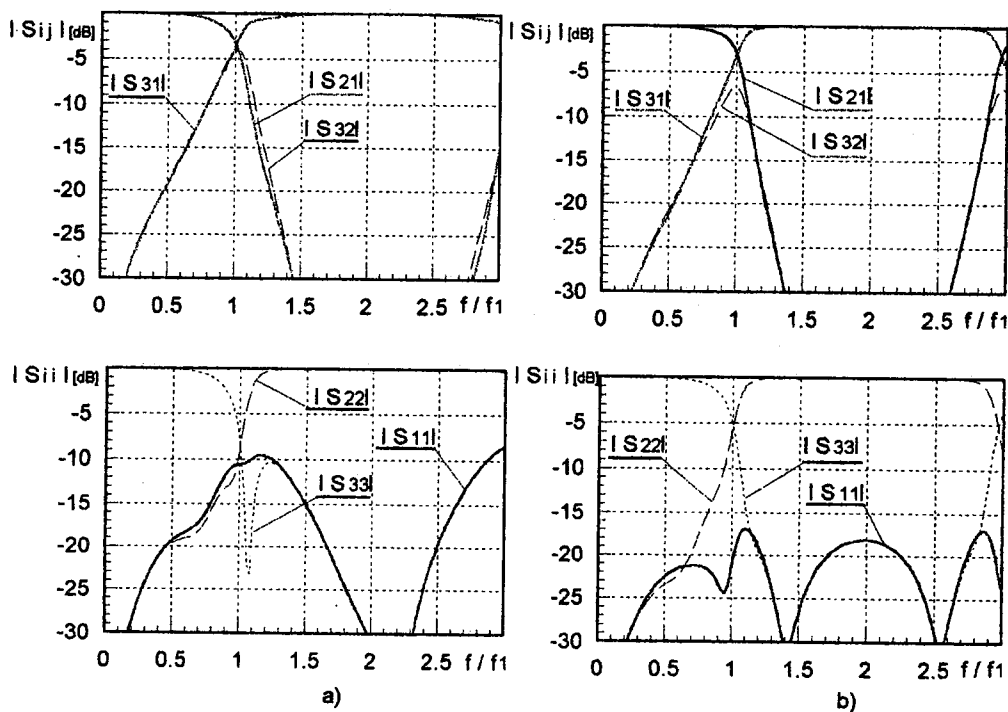


Rys. 11. Pasmowoprzepustowy filtr prototyp czwartego rzędu o $L_{ar} = 0.25$ dB i $f_2/f_1 = 3$ (a) oraz jego struktury ekwiwalentne po transformacji (b i c)



Rys. 12. Zwrotnica szerokopasmowa pokrywająca pasmo (0–3) GHz z częstotliwością rozdzielu $f_1 = 1$ GHz

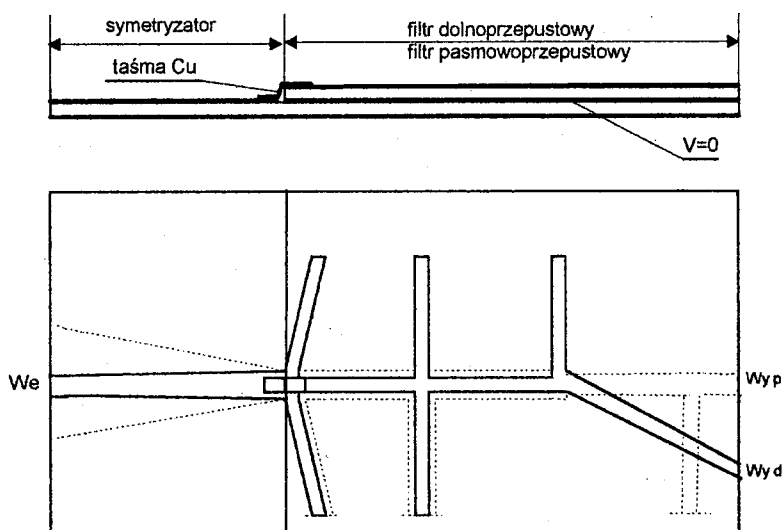
na rys. 12. Charakterystyki tej zwrotnicy przedstawiono na rys. 13a. Zwrotnica pokrywa pasmo (0–3) GHz. Straty odbiciowe sięgają wartości -9 dB, co odpowiada poziomowi zafalowania $L_{ar} = 0.58$ dB. Poddając optymalizacji parametry zwrotnicy można uzyskać charakterystyki, jak na rys. 13b. Tu maksymalny poziom



Rys. 13. Charakterystyki zwrotnicy przedstawionej na rys. 12 przed (a) i po optymalizacji (b)

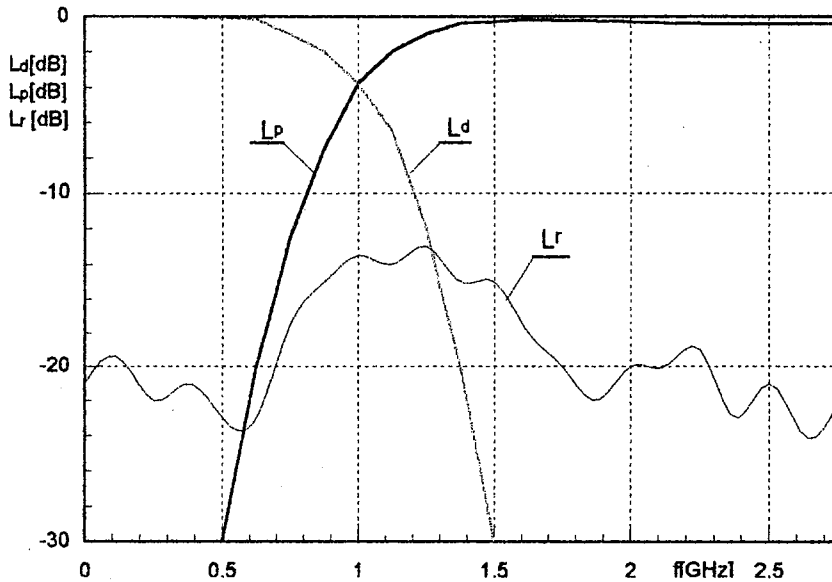
strat odbiciowych wynosi -17 dB, co odpowiada poziomowi zafalowania $L_{ar} = 0.08$ dB. Podczas optymalizacji uziemienniu podlegały tylko trzy pierwsze elementy każdego z filtrów. Ich wartości po optymalizacji przedstawiono na rys. 12 w nawiasach.

W rozpatrywanym przypadku pasmo zwrotnicy zostanie dodatkowo ograniczone do pasm przenoszenia zastosowanych przesuwników fazy. Ograniczenie to nie występuje przy szeregowym połączeniu filtrów. Układ realizujący szeregowe połączenie układów mikrofalowych w technice NLP, opisany w pkt. 4 pozwala na bezpośrednie połączenie filtrów (bez przesuwników fazy). Praktycznie wykonany układ zwrotnicy przedstawiono na rys. 14. Na wejściu zwrotnicy zastosowano symetryzator przekształcający NLP w symetryczną linię dwuprzewodową. Sygnał we wrotach wyjściowych pobudza wrota wejściowe filtrów umieszczonych symetrycznie względem masy. Wyjścia filtrów są wrotami wyjściowymi zwrotnicy oznaczonymi na rysunku odpowiednio przez w'_d i w''_p . Wrota wyjściowe zwrotnicy rozsunięto względem siebie na odległość umożliwiającą podłączenie złącz mikrofalowych. Zwrotnicę wykonano na podłożu LX 1030.



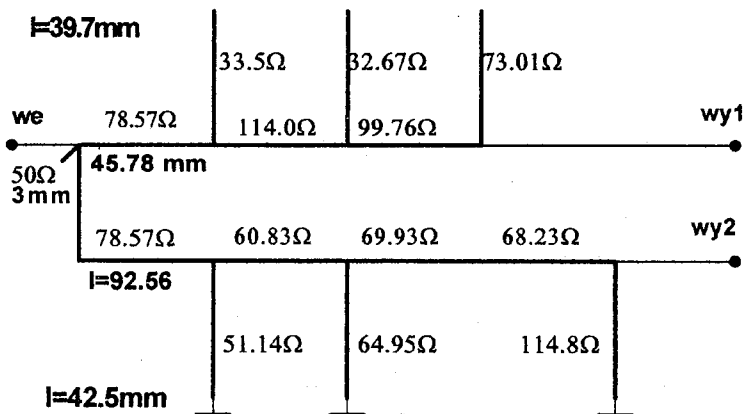
Rys. 14. Struktura szerokopasmowej zwrotnicy z szeregowym połączeniem filtrów w technice niesymetrycznej linii paskowej

Charakterystyki praktycznie wykonanej zwrotnicy przedstawiono na rys. 15. Przecięcie charakterystyk transmisyjnych zachodzi na poziomie 3.8 dB. Straty wtrąceniowe nie przewyższają wartości 0.5 dB zaś straty odbiciowe 12 dB. Straty wtrąceniowe wynikają przede wszystkim ze strat w dielektryku i z strat będących efektem promieniowaniem energii. Nieznacznie podwyższony poziom strat odbiciowych wywołany jest przez symetryzator. Typowa wartość WFS zastosowanego symetryzatora zawiera się w przedziale $(1.2-1.5)$. Zatem wypadkowa charakterystyka odbiciowa jest zgodna z założeniami teoretycznymi. Zwrotnica ta pokrywa pasmo $(0-2.8)$ GHz. Jeź parametry dowodzą słuszności przyjętych rozwiązań.

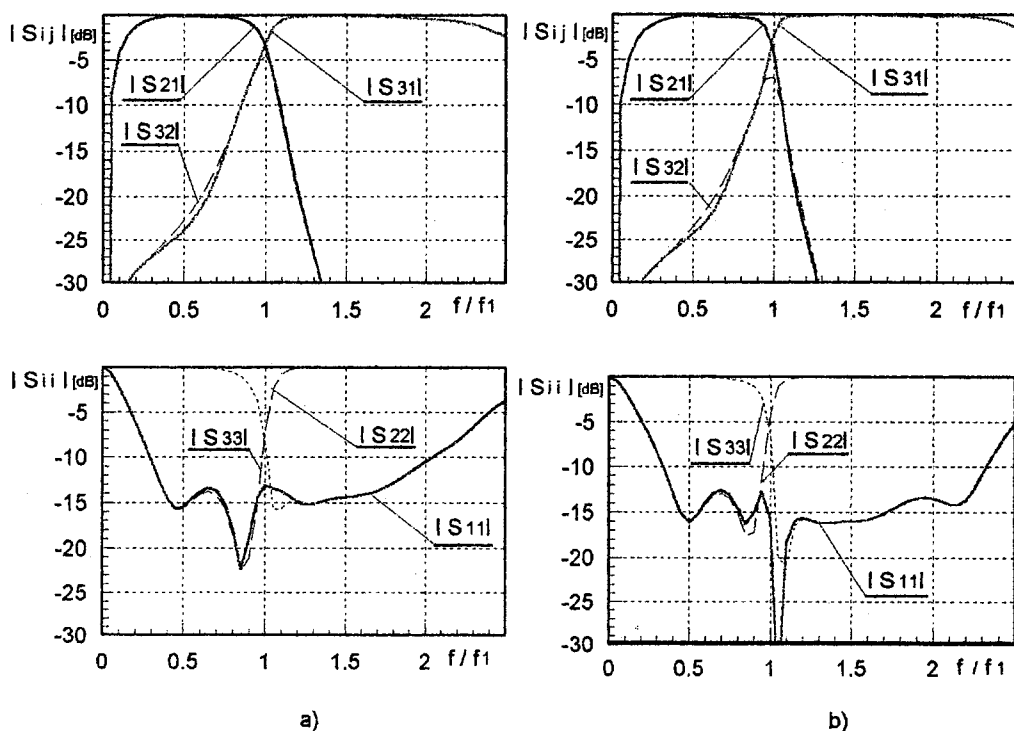


Rys. 15. Charakterystyki praktycznie wykonanej zwrotnicy dwukanałowej z szeregowym połączeniem układów

Wykorzystując rozpatrywane wyżej filtry, w układzie zwrotnicy rolę przesuw-
ników fazy spełniać mogą odcinki linii transmisyjnych, jak przedstawiono to na
rys. 16. Charakterystyki tej zwrotnicy przedstawiono na rys. 17. Przed optymaliza-
cją dla przyjętego poziomu zafalowania zwrotnica pokrywa pasmo dwu okta-
w $B = (0.32 - 1.88)$ GHz. Po optymalizacji pasmo poszerza się do wartości współczyn-
nika pokrycia $f_2/f_1 = \sim 6$. Zwrotnica ta pracuje poprawnie od dolnej częstotliwości



Rys. 16. Szerokopasmowa dwukanałowa zwrotnica mikrofalowa z liniami transmisyjnymi zastępującymi przesuwniki fazy



Rys. 17. Charakterystyki zwrotnicy szerokopasmowej z liniami transmisyjnymi przed (a) i po optymalizacji (b)

odcinka wynoszącej odpowiednio 0.32 GHz i 0.38 GHz. Występowania dolnej częstotliwości odcinka, różnej od zera wynika z bocznikującego wpływu filtra pasmowoprzepustowego na parametry filtra dolnoprzepustowego poniżej tej częstotliwości. Na charakterystykach transmisyjnych wzajemny wpływ filtrów objawia się spadkiem nachylenia zbocza charakterystyki filtra pasmowoprzepustowego poniżej częstotliwości 0.5 GHz. W układzie zwrotnicy zastosowano równoległą pojemność włączoną równolegle do wrót wejściowych w celu zmniejszenia wypadkowej susceptancji wejściowej.

PODSUMOWANIE

W skład szerokopasmowych zwrotnic mikrofalowych mogą wchodzić: filtry dolnoprzepustowe, górnoprzepustowe lub pasmowoprzepustowe. Zastosowanie tożsamości Kurody w odniesieniu do filtrów dolno- i pasmowoprzepustowych pozwala na realizację filtrów praktycznie o każdej wymaganej szerokości pasma przepustowego. W przypadku filtrów pasmowoprzepustowych zachodzi czasami potrzeba zamiany rodzaju filtra, co może wystąpić w przypadku filtrów wąskopasmowych.

W artykule przedstawiono metodykę transformacji struktur filtrów w odniesieniu do jednego rodzaju mikrofalowego pasmowoprzepustowego tj. filtru ze sprzężeniami bezpośrednimi. Ponieważ jest on ekwiwalentem elektrycznym innych pasmowoprzepustowych filtrów mikrofalowych, zaproponowana transformacja pozwala na projektowanie innego rodzaju filtrów mikrofalowych. Transformacja ta upraszcza proces projektowania filtrów w wybranej technice. Wyznaczone wartości elementów zapewniają uzyskanie punktu startu (podczas optymalizacji) leżącego blisko minimum globalnego. Podczas optymalizacji parametrów zwrotnic stosowano metodę gradientową. Optymalizację prowadzono pod kątem uzyskania wymaganego poziomu dopasowania we wrotach wejściowych zwrotnicy.

BIBLIOGRAFIA

1. I.N. Borbov: *Sintez korrektorov svc na peredajuschich liniach*. Radioelektr. Mos. 1980, No. 3, s. 11–16
2. A.P. Gorbacev: *Kaskadnyje diferencjalnoje fazovrascatieli diapazona svc*. Radioelektr. Moskwa 1984, Nr 11, s. 14–19
3. H. Gruchała: *Szerokopasmowy diplexer mikrofalowy z szeregowym połączeniem filtrów w technice mikropaskowej*. Elektronika 5/95, Warszawa 1995, s. 19–21
4. J.A.G. Malherbe: *Microwave Transmission Line Filters*. Artech House 1977
5. S.J. Mason, H.J. Zimmermann: *Electronic Circuits, Signals and Systems*. John Wiley & Sons. Inc. N.Y. 1960
6. G. Matthaei, L. Young, E.M.t. Jones: *Microwave Filters, Impedance-Matching Networks and Coupling Structures*. Art. House, 1980
7. V.P. Meschanov, I.V. Metelnikova, V.D. Yupikin, G.G. Chumaeyvskaya: *A New Structure of Microwave Ultrawide — Band Differential Phase Shifter*. IEEE Trans. on MTT 1994, vol. 42 No. 5, pp. 762–765
8. M.A. Silajev, S.F. Brancev: *Prolozenie matric i grafov k analizu svz ustrojstv*. Sov. Radio, Moskva 1970
9. R.J. Wenzel: *Exact Design of TEM Microwave Networks Using Quarter — Wave Lines*. IEEE Trans. on MTT, 1964, vol. 12, No. 1, pp. 94–111

H. GRUCHAŁA

DESIGNING OF BROADBAND MICROWAVE MULTIPLEXERS WITH USING OF KURODA'S IDENTITIES

S u m m a r y

Transformation method of low pass and band pass filters has been presented in this paper. Kuroda's identities have been used. The transformation permits to get realisable structures of these filters. Discussion of multiplexer structures consisting band pass filters has been made. Series connection of microwave circuits in microstrip technique has been described. Equivalent circuits of series connection of circuits have been shown. Experiment results of broadband multiplexer with series connection of circuits have been presented.

Key words: low pass and band pass filters, microwave multiplexers, Kuroda's identities.

Simple hardware implementations of modifiable bit permutations based on matrix description of permutation

ALBERT SADOWSKI

Instytut Podstaw Elektroniki, Politechnika Warszawska

Received 1997.09.20

Authorized 1997.10.08

In the paper we discuss a class of encryption/decryption schemes called the substitution-permutation networks (SPNs) or the cascade algorithms. First we briefly describe some extension of such ciphers — the position permutations. Position permutations are some realization of the idea of modifiable bit permutations. In this paper we describe an implementation aspect of this idea. We present matrix description of the bit permutations used in the substitution-permutation networks and based on this descriptions few simple, hardware implementations of the modifiable permutations. Presented solutions are digital circuits constructed with use of standard gates, open collector gates and three-state gates. They are easy to implement and effective in work; because of this they may be used in various applications.

Key words: Substitution-permutation network, bit permutation, bit permutation implementations.

1. INTRODUCTION

One of the most important advantages of the cascade algorithms is the high speed of their performance. These algorithms allow to relatively fast encrypt and decrypt messages; because of this, after adjusting so called *communication session key*, for example with use of the methods of public key cryptography, just the cascade algorithms are used to encrypt exchanged informations. The fastest implementations of the cascade algorithms are hardware implementations. Nowadays semiconductor integrated circuitus used to encrypt the streams of bits are produced.

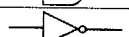
In [16] we described some modification of the substitution-permutation network. This modification consists in extending every iteration of the algorithm with so

called *position permutations*. The differential cryptanalysis ([2], [3], [4], [5]) is made with use of the differential characteristics. For the substitution-permutation network in which the position permutations are applied creating the differential characteristics as described in [5] is impossible. It is important that introducing the position permutations in the SPN does not change the type of the algorithm; we still have a number of substitution rounds connected by the permutations of a special kind (see [16]). The SPNs with the position permutations are some realization of the idea of networks with modifiable permutations. We hope that this idea will be developed in many manners.

In this paper we present an implementation aspect of this solution. The implementation of the position permutations requires a skill of constructing the circuits realizing modifiable permutations. In the paper we present the propositions of such circuits based on a matrix description of permutation. Presented methods allow to construct the permutations which can be modified in any way, not only by permuting positions. Because of this they can be implemented not only in the substitution-permutation networks with the position permutations but also in any other designs where the modifiable permutations are needed (not only in cryptography).

Presented in this paper methods of hardware implementation of modifiable permutations are the methods which result in simple digital circuits when realized; owing to this, such implementations are effective — this was an intention of the author. Probably, for many readers, they will be the very intuitively methods but author admitted that their presentation is a very good pretext to present formal, matrix description of the bit permutation. The author's aim was to emphasize the fact that in this case (and also in many others) engineer practice directly issues from mathematical theory and is in close relation with it.

In this paper we use following denotations and symbols:

Function	Denotation	Symbol
logical sum of bits a and b (OR)	$a + b$	
logical product of bits a and b (AND)	$a * b$	
negation of bit a (NOT)	$\sim a$	
negated logical product of bits a nad b (NAND)	$\sim(a * b)$	

Presented above logical functions (operations) can be extended to be more than two-argument operations. For example, the OR operations of three bits, a, b and c, we denote:

$$a + b + c$$

and we use a symbol:



2. BACKGROUND

In this paper we discuss an N -bit SPN consisting of a number of rounds of S-boxes connected by bit permutations (Fig. 1). The S-boxes are $p \times p$ bijective mappings and the permutations belong to a special set Π of permutations for which no two outputs of an S-box are connected to one S-box in the next round. There are $n = N/p$ S-boxes in each round. The round (or iteration) keys are in general

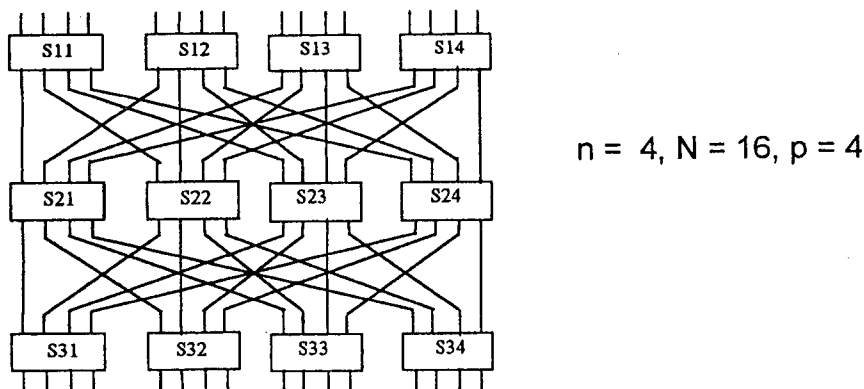


Fig. 1. A sample substitution-permutation network

independent on each other. The keying is done by xoring the iteration keys with the inputs to the rows of the S-boxes in each round and by xoring result of the last round with the last iteration key. In the networks with position permutations we moreover use keying the permutation.

3. POSITION PERMUTATIONS

An *S-box position* is a subset of these bits of particular word of the SPN, which come from — for S-boxes output word — or go to — for S-boxes input word — one S-box. Therefore in particular word bits $1..p$ (counting from the left) create first S-box position, bits $p + 1, ..., 2p$ second one, etc.. We will use the natural numbers $(1, 2, ...)$ to denote the S — box positions.

Next we briefly describe the position permutation. (More precise description is given in [16]). Position permutation is a transformation which moves every position of its input word onto a new location at the output (in particular onto the same location). Therefore, it is a transformation which changes the arrangement of the positions in the word (in particular to the same arrangement). We can think of the position permutation as a function which assigns a new position number to each S-box position. We can say that this transformation mixes whole S-box positions of the word. Sample position permutation is presented in Fig. 2.

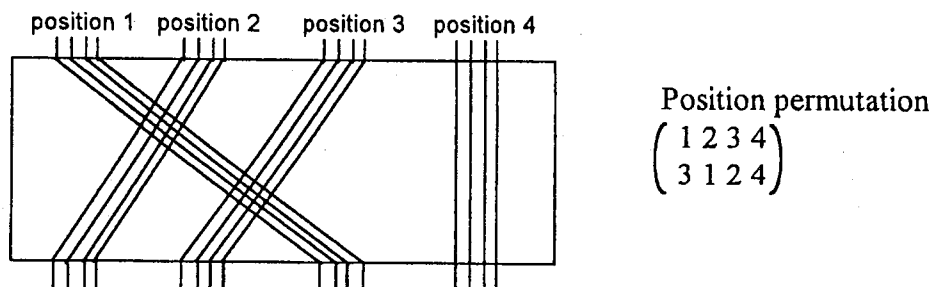


Fig. 2. An example of position permutation

Let P_p represent the position permutation of the set of S-box positions of one word. We will apply the permutation P_p to the output of the row of S-boxes in each round of the SPN (Fig. 3). Moreover we will introduce a dependence of used in particular iteration position permutation on the iteration key. Used in the round key will be determining which position permutation from the set of $n!$ possible permutations is to be used. We assume that each permutation is determined by the same number of keys (in practice it may be approximately the same number of keys). The differential characteristics are constructed to make the cryptanalysis of the algorithm, it means — to find the iteration keys. But to construct the characteristic we need to know the position permutations used in the rounds, so we need to know the used keys. From this follows an important theorem.

Theorem 1. For any iteration in which the position permutation was applied to the outputs of the S-boxes, it is impossible, for given S-boxes output xor, to determine unambiguously the round output xor if the used iteration key is unknown. For given

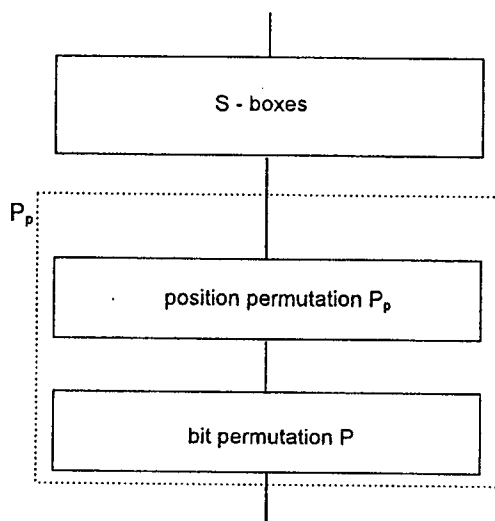


Fig. 3. One round of the SPN with position permutation

S-boxes output xor the number of possible round output xor values (for all possible keys) is equal to $n!/(n-k)!$, where k is the number of the S-boxes with non-zero input xor in considered round (we omit the proof, see [16]).

Hence, it is impossible to construct the differential characteristic as defined in [5] and determine the probability of its occurrence for the substitution-permutation network in which the position permutations were used.

4. MATRIX DESCRIPTION OF PERMUTATION

For the networks with established permutations, the bit permutations were realized by adequate wiring the inputs and outputs of successive rounds of substitutions. For the networks with the modifiable permutations we must have a possibility of easy change the permutations connecting successive rounds. In this section we present a matrix method of description of the permutations of the SPN; on the grounds of this method it is possible to easily implement modifiable permutations in hardware.

It is known that bit permutation P is a linear (homomorphous) mapping. In this mapping an N -bit input vector $\mathbf{I}$ is transformed into an N -bit output vector $\mathbf{O}$:

$\mathbf{I}, \mathbf{O} \in V_N$, where V_N is an N -dimensional vector space over the field $GF(2)$,

$P: V_N \ni \mathbf{I} \rightarrow \mathbf{O} = P(\mathbf{I}) \in V_N$.

Therefore the bit permutation P is a one-one linear transformation of the vector space V_N onto itself, or an automorphism of V_N . However not every automorphism of the space V_N is the permutation from the set Π (a definition of this set is given section 2) of permutations used in SPNs. (For example; an identity mapping belongs to $\text{Aut}(V_N)$ but it does not belong to the set Π). Therefore we have:

$$\Pi \subset \text{Aut}(V_N).$$

Because the bit permutation P is the linear mapping we can write it as a matrix called a *linear mapping matrix*. We write the vectors $\mathbf{I}, \mathbf{O}$ as column vectors: $\mathbf{I} = [i_1, i_2, \dots, i_N]^T$ and $\mathbf{O} = [o_1, o_2, \dots, o_N]^T$. A matrix $\mathbf{P}$ of the linear mapping P is a square $N \times N$ matrix. Now we can present the bit permutation of the substitution-permutation network as follows:

$$\begin{bmatrix} o_1 \\ \vdots \\ o_N \end{bmatrix} = \begin{bmatrix} p_{1,1} & \dots & p_{1,N} \\ \vdots & & \vdots \\ p_{N,1} & \dots & p_{N,N} \end{bmatrix} \begin{bmatrix} i_1 \\ \vdots \\ i_N \end{bmatrix}. \quad (4)$$

As we noted earlier, not every automorphism of the vector space V_N is the bit permutation from the set Π . Therefore the matrix $\mathbf{P}$ has some special properties.

Property 1. There is exactly one 1 in every row of the matrix $\mathbf{P}$.

Proof. Consider k -th bit of the vector $\mathbf{O} \rightarrow o_k$, thus o_k represents the k -th output of the bit permutation. This output is connected with exactly one input of the bit permutation, consider — with the m -th input. Therefore

$$\forall \mathbf{I} \in V_N \quad o_k = 1 \Leftrightarrow i_m = 1. \quad (5)$$

For the vector $\mathbf{I}$ with zeros at all positions but m -th one, we obtain 1 at the output o_k if, and only if, $p_{k,m} = 1$. For every vector $\mathbf{I}$ with zero at the m -th position (remaining bits — arbitrary), we obtain zero at the output o_k if, and only if, in the row $\mathbf{P}_{(k)}$ all bits but m -th one are equal to zero. Hence, to have the bit o_k depend only on the bit i_m or to have the condition (5) satisfied, we must have

$$(p_{k,1} = 1) \wedge (p_{k,t} = 0, \text{ for } t \neq m, t \in \{1, 2, \dots, N\}).$$

From this follows that in every row of the matrix $\mathbf{P}$ there is only one 1 and it is in the column number of which is equal to the number of the bit of input which is connected with this bit of output number of which is equal to the number of the row. $\square$

Therefore the k -th row of the matrix $\mathbf{P}$ describes the k -th output of the permutation and the location of the only 1 in this row describes with which input this output is connected.

Property 2. There is only one 1 in every column of the matrix $\mathbf{P}$.

Proof. Considering similar as in the proof of the Property 1 we obtain the corollary that m -th column of the matrix $\mathbf{P}$ describes m -th input of the permutation P . 1 is in this row of this column number of which is equal to the number of the output of P with which the m -th input is connected. Because there is only one such output, the m -th column contains one 1. $\square$

Fact 2. The matrix $\mathbf{P}$ is an invertible matrix. The inverse of $\mathbf{P}$ is a transpose of $\mathbf{P}$: $\mathbf{P}^{-1} = \mathbf{P}^T$.

Proof. Formal proof is obvious; it is easy to show that if $\mathbf{P}$ is the matrix containing only one non-zero element in every row and in every column then $\mathbf{P}$ is the invertible matrix and the inverse of it is its transposition. But let's try to prove it another way.

We can think of the matrix $\mathbf{P}$ as a description of the permutation P ; because the permutation P is the one-one mapping then it exists an inverse mapping of $P \rightarrow P^{-1}$. Hence for any vector $X \in V_N$ we have:

$$P^{-1}(P(X)) = X \quad \text{and} \quad P(P^{-1}(X)) = X.$$

In the matrix description:

$$\mathbf{P}^{-1} \mathbf{P} \mathbf{X} = \mathbf{X} \quad \text{and} \quad \mathbf{P} \mathbf{P}^{-1} \mathbf{X} = \mathbf{X}.$$

Hence

$$\mathbf{P}^{-1} \mathbf{P} = \mathbf{P} \mathbf{P}^{-1} = \mathbf{E}, \quad \mathbf{E} \text{ represents an identity, square } N \times N \text{ matrix.}$$

Therefore $\mathbf{P}$ is the invertible matrix.

Let $\mathbf{I}$ and $\mathbf{O}$ represent input and output vectors, respectively, of the permutation P . In permutation P^{-1} , $\mathbf{I}$ is the image of the vector $\mathbf{O}$. In the matrix $\mathbf{P}$ we had the row number k describing k -th bit of the vector $\mathbf{O}$; in that row we had one at the m -th position, where m is the number of the bit of the vector $\mathbf{I}$ to which the bit o_k is connected. In the transformation P^{-1} , k -th bit of the input is connected to m -th bit of the output. Therefore k -th column of the matrix $\mathbf{P}^{-1}$ is equal to a transpose of the k -th row of $\mathbf{P}$:

$$\mathbf{P}^{-1(k)} = (\mathbf{P}_{(k)})^T, \quad k \in \{1, 2, \dots, N\}.$$

From this follows that

$$\mathbf{P}^{-1} = \mathbf{P}^T.$$

□

Next property of the matrix $\mathbf{P}$ is related to the fact that a fundamental requirement for the bit permutations is that they cause the *diffusion* of the bit changes in the network. Because of this only these permutations for which no two outputs of the S-box are connected to one S-box in the next round create the previously mentioned set Π .

Property 3. Every submatrix $\mathbf{P}_{((r-1)p+1, (r-1)p+2, \dots, rp), (s-1)p+1, (s-1)p+2, \dots, sp)}$, $s, r \in \{1, 2, \dots, n\}$ of the matrix $\mathbf{P}$ contains at most one element equal to 1 (see Fig. 4).

Proof. The rows $(r-1)p+1, \dots, rp$ describe the outputs of the same numbers of the permutation P . These outputs go to one S-box of some iteration w (to the S-box number r). Consider submatrix $\mathbf{P}_{((r-1)p+1, (r-1)p+2, \dots, rp)}$. The columns number $1, \dots,$

$$\mathbf{P} = \begin{array}{c} \text{row\#} \end{array} \begin{array}{c} \text{col\#} \end{array} \begin{array}{c} 1 \quad p \quad p+1 \quad 2p \quad (n-1)p+1 \quad np \end{array} \begin{bmatrix} 1 & & & & & \\ \vdots & & & & & \\ p & & & & & \\ p+1 & & & & & \\ \vdots & & & & & \\ 2p & & & & & \\ \vdots & & & & & \\ (n-1)p+1 & & & & & \\ \vdots & & & & & \\ np & & & & & \end{bmatrix}$$

Fig. 4. Square $p \times p$ submatrixes of the matrix $\mathbf{P}$

p of this submatrix describe how the inputs of the S-box r of the iteration w depend on the outputs of the first S-box of the previous iteration $w-1$, the columns $p+1, \dots, 2p$ — on the outputs of the second S-box, etc., the columns $(n-1)p+1, \dots, np$ — on the outputs of the n -th S-box. Because $P \in \Pi$, every input of the S-box r of the round w comes from different S-box of the round $w-1$. Therefore in the submatrix $P_{\langle (s-1)p+1, (s-1)p+2, \dots, sp \rangle, \langle (r-1)p+1, (r-1)p+2, \dots, rp \rangle}$, exactly one element is equal to 1 — if the S-box r of the round w has connection with S-box s of the round $w-1$, or all elements are zeros — if these S-boxes are not connected with each other. $\square$

Therefore we can partition the whole matrix P into square $p \times p$ submatrixes, each one of which describes connections between two S-boxes of successive iterations.

Next simple remark is a ground of the presented in further part of this chapter method of hardware implementation of modifiable permutations. Because of this it deserves a special attention.

Fact 3. The output vector O of the bit permutation P is a linear combination of successive columns of the matrix P of coefficients equal to successive bits of the input vector:

$$O = i_1 P^{(1)} + i_2 P^{(2)} + \dots + i_n P^{(n)}. \quad (6)$$

5. IMPLEMENTATION OF MODIFIABLE BIT PERMUTATIONS

In the previous section we noted that we can write the output vector of the bit permutation as the linear combination of successive columns of the matrix P of coefficients equal to successive bits of the input vector. We can use this fact in simple hardware realization of the modifiable bit permutations. A multiplication of the vector of the linear space over the field $GF(2)$ by a scalar of the same field is equivalent to executing for every bit of this vector a logical operation AND, one argument of which is the bit of the vector and the other one — the scalar. Such solution is presented on Fig. 5a). Summing the vectors $i_k P^{(k)}$, $k = 1, 2, \dots, N$ is executing an operation EXOR (EXclusive OR — an addition modulo 2) of corresponding bits of them (an addition in $GF(2)$ is an „ordinary” addition modulo 2). Because in any row of all the vectors $i_k P^{(k)}$, $k = 1, 2, \dots, N$ there is at most one 1, the operation EXOR can be replaced with an operation OR — a little easier to implement in the integrated circuit. This solution is presented on Fig. 5b).

If we implement the columns of the matrix P as modifiable (writable) registers then by changing their contents we will be able to change the permutation realized by the circuit. Therefore we presented a simple logical circuit which may be used to implement modifiable permutations. The words written to the control registers are the words in a code „1 of N ” (the code „1 of N ” is used to encode the values from the set $\{1, 2, \dots, N\}$ with use of N bits: a value $a \in \{1, 2, \dots, N\}$ coded in this code gives an N -bit word in which there is 1 at the a -th position and zeros at the remaining ones). However, it may occur that, for some reasons, it is more convenient to have the values of these words coded in a natural binary code. In such situation the circuit

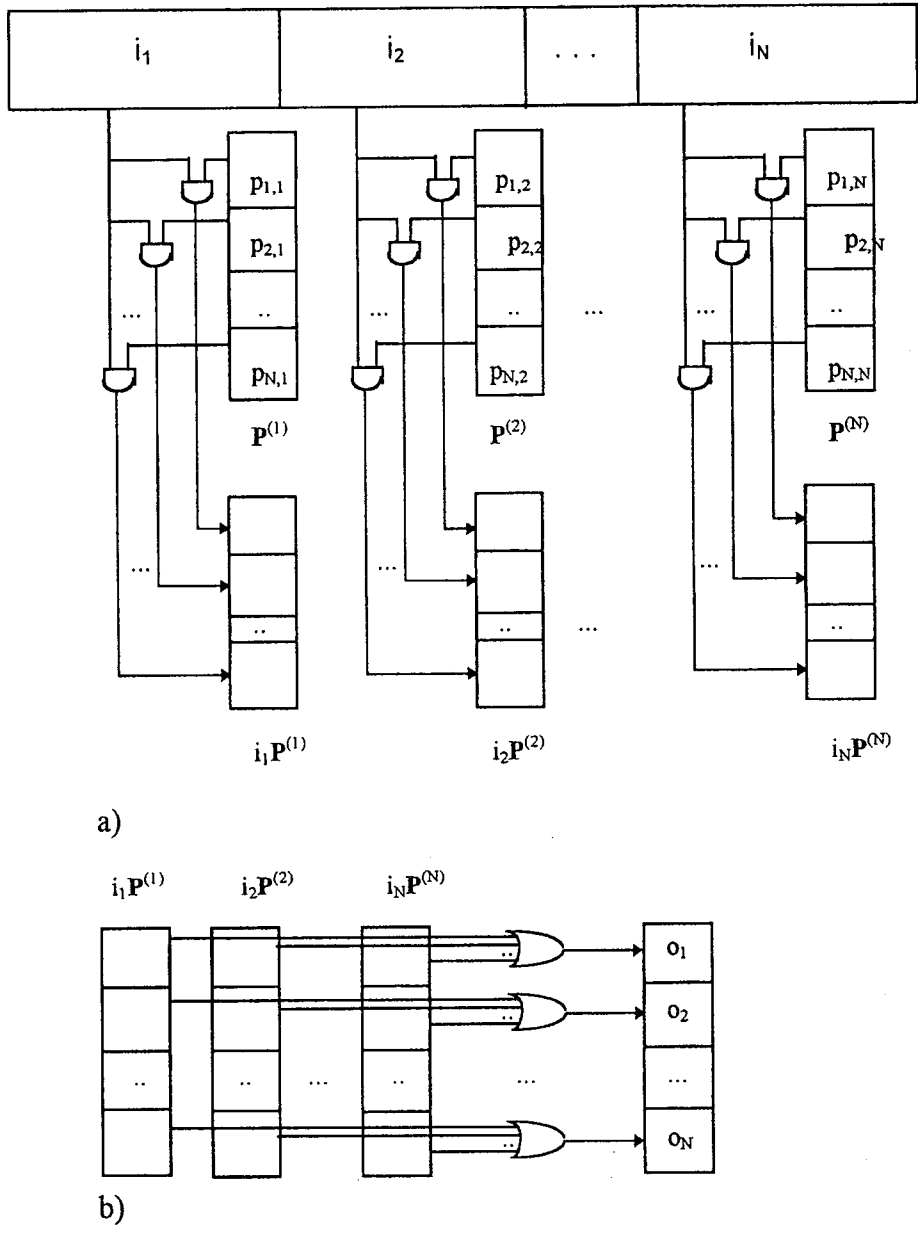


Fig. 5. Hardware implementation of modifiable bit permutations

presented on the Fig. 5 may be extended as presented on Fig. 6 — a case for $N = 8$. (We present only this part of the circuit which realizes the connection of k -th input with outputs; the parts realizing the connections of the remaining inputs are constructed in the same way).

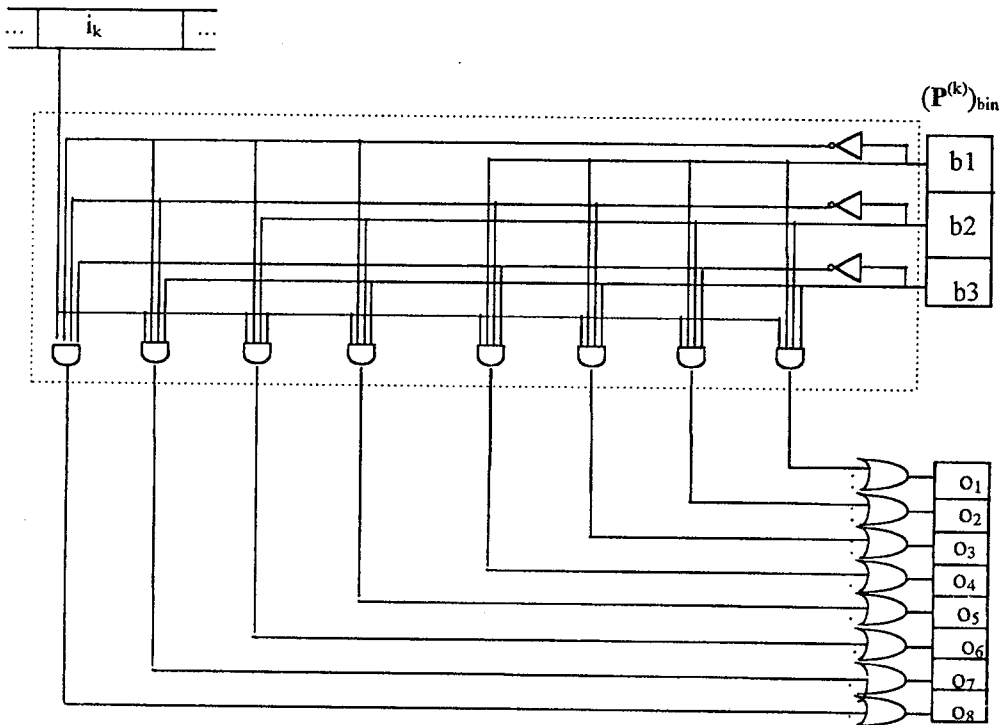


Fig. 6. A modification of the circuit from Fig. 5

The binary coded value determining in which row of the column $P^{(k)} 1$ would be placed is given as a word $(P^{(k)})_{bin}$ (in this case the rows would be numbered starting with 0). Then, appropriately for this value, one of the AND gates is opened and k -th input of the permutation is connected with appropriate output. For example, if we set a binary value 000 in the word $(P^{(k)})_{bin}$ the most left gate will be opened, if we set a value 111 — the most right.

A part of the circuit from Fig. 6 which is selected in a dashed-line box is a popular in digital-circuit engineering *demultiplexer circuit* with one-bit input. The demultiplexer with one-bit input (Fig. 7) is used to transmit input bit x to one of N outputs; the number of this output is equal to an actual value of some control word A called an *address*. All the remaining outputs are equal to zeros. According to what was written above, the described modifiable permutation circuit may be presented as on Fig. 8 (a case for $N = 4$).

A digital circuit working „inversely” to the demultiplexer is a *multiplexer*. The one-bit output multiplexer is used to select one of many input bits and to transmit it to the output (Fig. 9a)). A input A is called an *address input*. It is used to determine the number of input which is to be transmitted to the output. The address is coded in

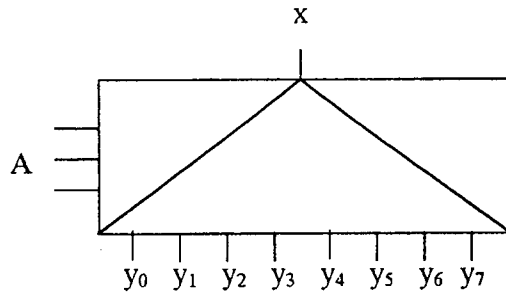


Fig. 7. A demultiplexer with one bit

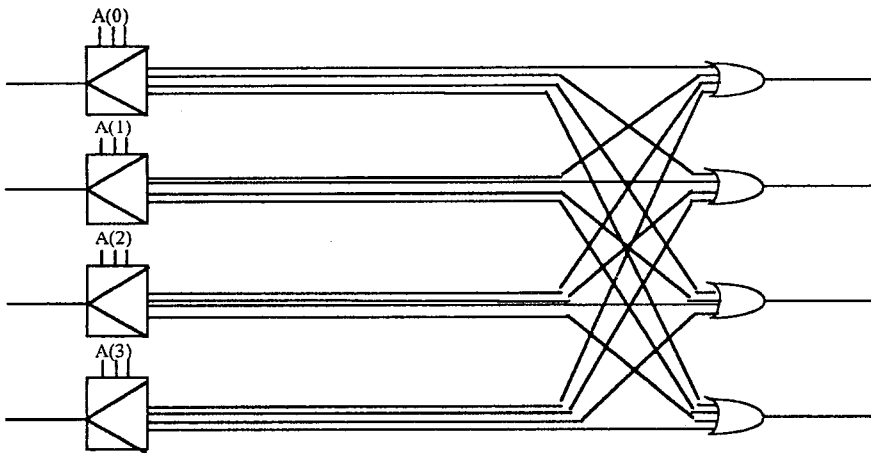


Fig. 8. The implementation of the modifiable permutations with use of the demultiplexers

natural binary code. The internal structure of the multiplexer, for $N = 4$ inputs, is presented on Fig. 9b).

The N -bit modifiable bit permutation may also be realized with use of the multiplexers. Such implementation is presented on Fig. 10 (for $N = 4$ bit permutation).

In the presented in further part of this chapter circuits which implement the modifiable bit permutations on the grounds of the atrix description of bit permutation we omit the parts of this circuits which are used to change the code of the control words from the binary code to the code „1 of N ” (for simplicity; we showed such solution above). Therefore, as the control words there will always be given the binary vectors creating appropriate row or appropriate column of the permutation description matrix P (according to the Fig. 5). We remind that expressed in the code „1 of N ” value of the k -th column of the matrix P determines the number of the

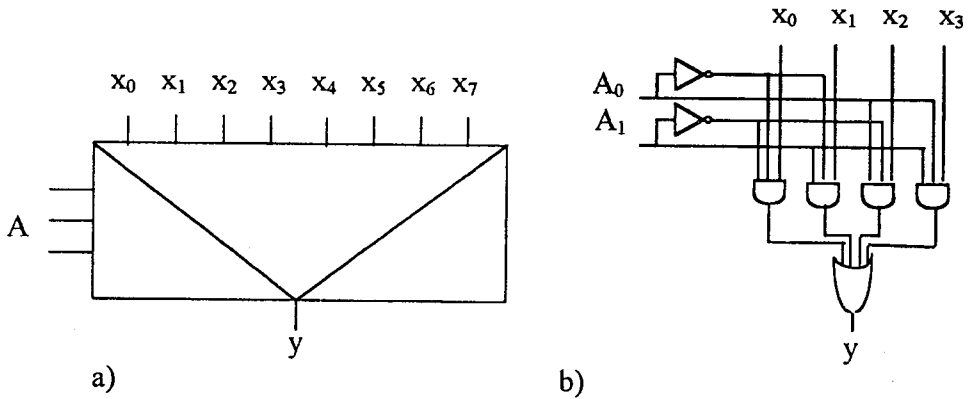


Fig. 9. A symbol (a)) and the internal structure of the multiplexer (b)) (for $N = 4$ inputs)

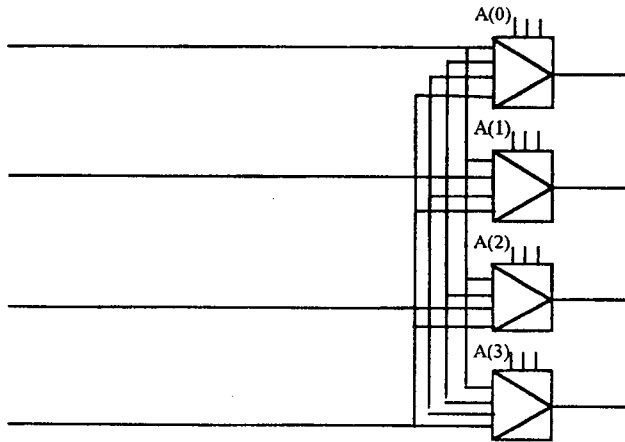


Fig. 10. The implementation of the modifiable permutation with use of the multiplexers (for 4 bit permutation)

output of the permutation with which k -th input is connected, and similarly; expressed in the code „1 of N ” value of the 1-th row of the matrix P determines the number of the input of the permutation with which 1-th output is connected.

In the following two sections we present next two kinds of the modifiable permutations implementation circuits. Design of this circuits is also based on the matrix description of the permutation. In the first one of them the *open collector NAND gates* are used and in the second one — the *three-states NAND gates*. The circuits have a very simple structure and are very similar to the circuit presented on Fig. 5.

6. IMPLEMENTATION WITH USE OF THE OPEN COLLECTOR GATES

In the next method of implementation of modifiable bit permutations we use the open collector NAND gates. These components are intended for connecting the outputs; such connecting relies on attaching the outputs of the components to one wire and connecting an external collector resistor R — see Fig. 11. The internal structure of the open collector NAND gate and a precise description of the work of the NAND gate (from the electronic circuits theory point of view) can be easily found in the literature.

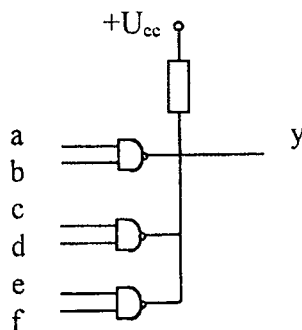


Fig. 11. Connecting the outputs of the open collector NAND gates

We have 1 at the output y if, and only if, none of the output transistors of the gates is active (none conducts), therefore connecting the outputs corresponds to realizing a function (we omit „*” in logical products)

$$y = (\sim(ab)) (\sim(cd)) (\sim(ef)),$$

or according to the de Morgans law

$$y = \sim(ab + cd + ef).$$

A realization of the N -bit modifiable permutation with use of the open collector NAND gates is presented on Fig. 12 (for $N = 3$).

To the input $A(i)$ we give the value — coded in the code „1 of N ” — determining which one of the N gates connected to the i -th output is to be opened. A signal from the input of the permutation given to the second input of this opened gate will be transmitted to the output. The output transistors of the remaining gates, it means, of these to which zero bits of the word $A(i)$ are connected, will always be turned off independently on the state of their inputs connected with the inputs of the permutation. Owing to this, these gates will not transmit the signals from these inputs to the output. This way the connection of appropriate output with appropriate input is realized. It is easy to note that i -th control word $A(i)$ represent i -th row of the matrix P , $P_{(i)}$, $i \in \{1, 2, \dots, N\}$.

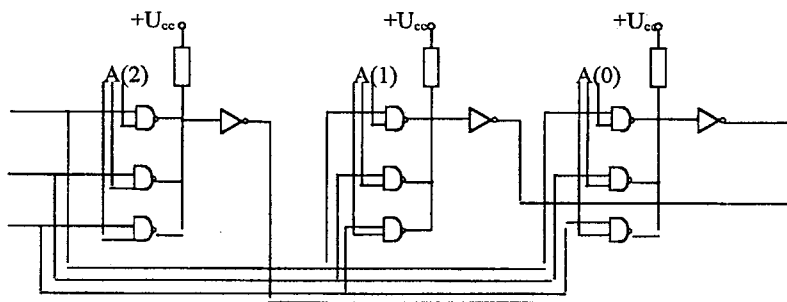


Fig. 12. A realization of the modifiable permutations with use of the open collector NAND gates

7. IMPLEMENTATION WITH USE OF THE THREE-STATE GATES

A very similar to the described in the previous section modifiable permutations implementation circuit may be realized with use of the *three-states gates*. A symbol of the three-states NAND gate is presented on Fig. 13.

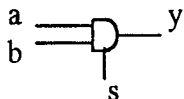


Fig. 13. A symbol of the three-states NAND gate

The component of this kind works following way: if $s = 0$, we have $y = \sim(ab)$. However, if $s = 1$ the output is in a disconnected state and does not depend on input signals. Then the value of the output is indeterminate and the gate has a large output impedance. The outputs of the three-state gates may be connected to the common wire but at a given moment at most one gate may be active (it means: may have 0 or 1 at the output). The remaining gates have to be „turned off” by giving appropriate control signals.

A modifiable permutations circuit constructed with use of the three-state NAND gates is presented on the Fig. 14 (for $N = 3$).

As we noted, the circuit is very similar to the circuit with the open collector NAND gates; very similar are the circuits topologies, the same is the number of gates and the role that the control words $A(i)$ play.

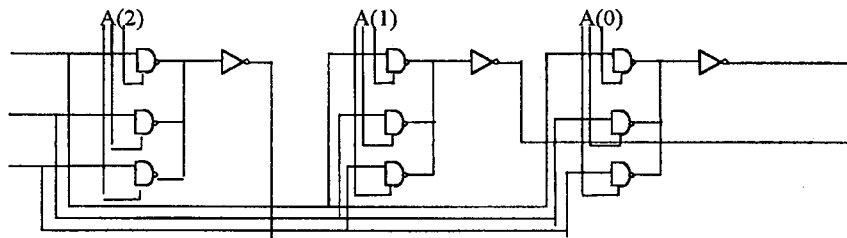


Fig. 14. A realization of the modifiable permutations with use of the three-state NAND gates

8. SIMPLE IMPLEMENTATION OF POSITION PERMUTATIONS

In this section we describe an adaptation of the solutions presented in previous sections for realization of the position permutation (see chapter 2). the matrix P describing position permutation has some special structure. Every submatrix $P_{\langle\langle s-1 \rangle\rangle_{p+1, \dots, sp}^{p+1, \dots, sp}}$, $s, r \in \{1, 2, \dots, N\}$ of the matrix P is either an identity matrix — if r is an image of the position s in permutation P_p (P_p represent position permutation), or a null matrix — in the other case:

$$P_{\langle\langle s-1 \rangle\rangle_{p+1, \dots, sp}^{p+1, \dots, sp}} = \begin{cases} E_{pxp}, & \text{if } P_p(s) = r, s, r \in \{1, 2, \dots, n\}, E_{pxp} \text{ represents} \\ & \text{square, } p \times p \text{ unitary matrix,} \\ [0]_{pxp}, & \text{in the other case.} \end{cases}$$

Simple matrix P describing position permutation is presented on Fig. 15a. The description of the position permutation can also be given as a square $n \times n$ matrix Q . The position permutation matrix Q which describes permutation

$$P = \begin{array}{c} \text{col\#1} \quad p \quad p+1 \quad 2p \quad (n-1)p+1 \quad np \\ \text{row\#} \end{array} \begin{bmatrix} 1 & & & & & \\ & \text{"0"} & & \text{"0"} & & \dots & & 1 \\ & & & & & & & 1 \\ p+1 & 1 & & \text{"0"} & & \dots & & \text{"0"} \\ 2p & & 1 & & & & & \\ & & & & & & & \\ (n-1)p+1 & & & & 1 & & & \\ & \text{"0"} & & & & \dots & & \text{"0"} \\ np & & & & & & 1 & \end{bmatrix}$$

a)

Position permutation matrix Q

$$Q = \begin{bmatrix} 0 & 0 & \dots & 1 \\ 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 1 & \dots & 0 \end{bmatrix}_{n \times n}$$

$$P_p = \begin{pmatrix} 1 & 2 & \dots & n \\ 2 & n & \dots & 1 \end{pmatrix}$$

b)

Fig. 15. Sample matrixes describing position permutation

presented on Fig. 15a) is presented on Fig. 15b). On the grounds of the matrix Q we can realize the logical circuit presented on Fig. 16. This circuit is a modification of the circuit from the Fig. 5. The columns of the matrix Q correspond to the writable registers $Q^{(1)}, Q^{(2)}, \dots, Q^{(n)}$. We write to the register $Q^{(k)}$ the number $P_p(k)$ in the code „1 of n ".

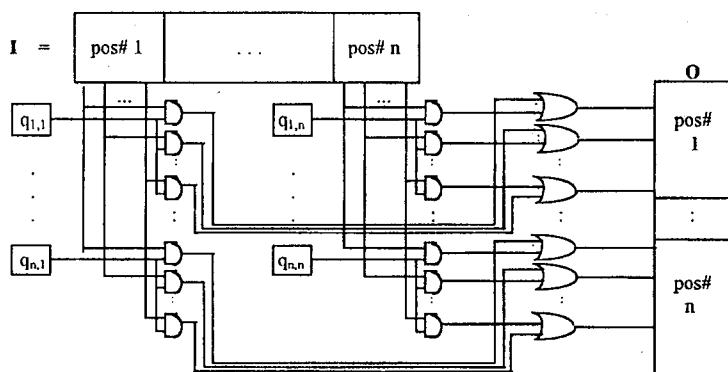


Fig. 16. Hardware implementation of position permutations

SUMMARY

At the beginning of this paper we shortly describe the position permutations and presented the matrix description of the bit permutation. Next we presented a few hardware implementations of modifiable bit permutations designed on the grounds this description. These implementations are very similar to each other and are the circuits of a simple structure. They are constructed with use of the ordinary gates, the open collector gates or the three-state gates. For the N -bit permutation the number of gates of each one of the circuits is equal to $N^2 + N$. In the situation in which the control words of the permutation circuit are to be expressed in the natural binary code we have about $N \log_2 N$ additional gates (see Fig. 6). The actual values of the number of gates for the permutations of various number of bits are presented in Table 1.

Table 1
The number of gates of the implementation circuit for the permutations of various number of bits

N	no. of gates for code „1 of N ” ($N^2 + N$)	no. of gates for binary code ($N^2 + N + N \log_2 N$)
4	20	28
8	72	96
16	272	336
32	1056	1216
64	4160	4544
128	16512	17408

We can realize the position permutation with use of one of the circuits presented in sections 3.3, 3.4 or 3.5 — in such case we have a possibility of any modification of the bit permutation (including permuting positions), or — if we establish the bit permutations in the network and we only want to permute positions — we can use simpler circuit presented in section 3.6. Presented in the chapter hardware solutions are easy to implement in the integrated circuit (and not only) and they ensure effective (fast) performance of the algorithm. They are alternative solutions of one problem; the choice of one of them depends on real construction requirements.

REFERENCES

1. C.M. Adams: *On immunity against Biham and Shamir's differential cryptanalysis*. Information Processing Letters 1992, 41(2), p. 77–80
2. I. Ben-Aroya, E. biham: *Differential Cryptanalysis of Lucifer*. Journal of Cryptology, 1996, Vol. 9, No. 1, p. 21–34
3. E. Biham, A. Shamir: *Cryptanalysis of Snefru, Khafre, REDOC-II, LOKI and Lucifer*. Advances in Cryptology — CRYPTO'91. Springer-Verlag, 1992, pp. 156–171
4. E. Biham, a. Shamir: *Differential cryptanalysis of the Full 16-Round DES*. Advances in Cryptology — CRYPTO'92. Springer-Verglag, 1993, pp. 487–496
5. E. Biham, A. Shamir: *Differential Cryptanalysis of DES-like Cryptosystems*. Journal of Cryptology, 1991, Vol. 4, No. 1, p. 3–72
6. L. Brown, M. Kwan, J. Pieprzyk, J. Seberry: *Improving Resistance to Differential Cryptoanalysis and the Redesign of LOKI*. Advances in Cryptology — ASIACRYPT'91, Springer-Verlag, 1992, pp. 36–50
7. L. Brown, J. Seberry: *On the design of permutation P in DES type cryptosystems*. Advances in Cryptology — EUROCRYPT'89. Springer-Verlag, 1990, pp. 696–705
8. C. Charnes, J. Pieprzyk: *Linear Nonequivalence versus Nonlinearity*. Advances in Cryptology — AUSCRYPT'92. Springer-Verlag, 1993, pp. 156–164
9. H.M. Heys, S.E. Tavares: *Substitution-Permutation Networks Resistant to Differential and Linear Cryptanalysis*. Journal of Cryptology, 1996, Vol. 9, No. 1, p. 1–20
10. K. Kim: *Construction of DES-like S-boxes Based on Boolean Functions Satisfying the SAC*. Advances in Cryptology — ASIACRYPT'91. Springer-Verlag, 1992, pp. 59–72
11. L.R. Knudsen: *Cryptanalysis of LOKI91*. Advances in Cryptology — AUSCRYPT'92. Springer-Verlag, 1993, pp. 196–208
12. L.R. Knudsen: *Iterative Characteristics of DES and s^2 -DES*. Advances in Cryptology — CRYPTO'92. Springer-Verlag, 1993, pp. 497–514
13. M. Matsui: *Linear cryptanalysis method for DES cipher*. Advances in Cryptology — EUROCRYPT'93. Springer-Vergla, 1994, pp. 386–397
14. K. Nyberg, L.R. Knudsen: *Provable Security Against Differential Cryptanalysis*. Advances in Cryptology — CRYPTO'92. Springer-Verlag, 1993, pp. 566–574
15. M. Sivabalan, S.E. Tavares, L.E. Peppard: *On the Design of SP Networks from an Information Theoretic Point of View*. Advances in Cryptology — CRYPTO'92. Springer-Verlag, 1993, pp. 260–279
16. A. Sadowski: *Key-Controlled Bit Permutations in Substitution Permutation Networks*. Kwart. Elektr. i Telekom., 1997, Vol. 43, z. 3, ss. 333–351

A. SADOWSKI

**PROSTE SPRZĘTOWE IMPLEMENTACJE MODYFIKOWALNYCH PERMUTACJI
BITÓW OPARTE NA MACIERZOWYM OPISIE PERMUTACJI****S t r e s z c z e n i e**

W artykule opisano pewną klasę procedur szyfracji/deszyfracji zwaną sieciami podstawieniowo-permutacyjnymi, lub algorytmami kaskadowymi. Opisano w skrócie pewne rozszerzenie tych szyfrów — permutacje pozycji. Są one pewną realizacją pomysłu modyfikowalnych permutacji bitów. Opisano implementacyjny aspekt tego pomysłu i zaprezentowano macierzowy opis permutacji bitów używanych w sieciach podstawieniowo-permutacyjnych, a także kilka opartych na tym opisie prostych sprzętowych implementacji modyfikowalnych permutacji. Zaprezentowane rozwiązania są układami cyfrowymi skonstruowanymi przy użyciu standardowych bramek, bramek trójszanowych, lub bramek z otwartym kolektorem. Są łatwe w implementacji i efektywne w działaniu, w związku z tym mogą mieć różnorodne zastosowania.

Słowa kluczowe: Permutacje bitów, sieci podstawieniowo-permutacyjne, implementacje permutacji bitów.

Dynamiczne zniekształcenia intermodulacyjne we wzmacniaczach mocy m.cz.

MIECZYŚLAW GŁOWACKI, JÓZEF STANCLIK

Instytut Telekomunikacji i Akustyki, Politechnika Wrocławska

MARIAN PIERZCHAŁA

Przedsiębiorstwo HILMAR, Wrocław

Otrzymano 1997.08.13

Autoryzowano 1998.01.02

W pracy zostały omówione modele wzmacniacza i sygnału sterującego. Na podstawie przyjętych modeli została wykonana analiza warunków powstawania dynamicznych zniekształceń intermodulacyjnych we wzmacniaczach mocy m.cz. z ujemnym sprzężeniem zwrotnym i kompensacją wyprzedzeniem fazowym w bloku sprzężenia zwrotnego oraz opóźnieniem fazowym w torze wzmocnienia. Zostały wyprowadzone kryteria projektowe, których spełnienie gwarantuje niewystępowanie dynamicznych zniekształceń intermodulacyjnych w stopniach wejściowych wzmacniaczy z obu typami kompensacji. Zostało wykazane, że maksymalna wielkość napięcia sterującego stopień wejściowy wzmacniacza w stanach przejściowych jest mniejsza w przypadku zastosowania kompensacji wyprzedzeniem fazowym, niż w takim samym wzmacniaczu, w który jednak została zastosowana kompensacja opóźnieniem fazowym. Wynika z tego, że zastosowanie kompensacji wyprzedzeniem fazowym zapewnia mniejszą podatność na przesterowanie stopnia wejściowego wzmacniacza w stanach przejściowych, niż zastosowanie kompensacji opóźnieniem fazowym.

Słowa kluczowe: wzmacniacz mocy małych częstotliwości, dynamiczne zniekształcenia intermodulacyjne.

1. WPROWADZENIE

Dynamiczne zniekształcenia intermodulacyjne (*transient intermodulation distortion* — TIM) powstają w wyniku nieliniowej pracy wzmacniacza w stanach

przejściowych. We wzmacniaczach z globalną pętlą ujemnego sprzężenia zwrotnego najczęściej są one wynikiem przesterowania stopnia wejściowego [5].

Na możliwość przesterowania pierwszego stopnia wzmacniacza mocy m.cz. objętego pętlą ujemnego sprzężenia zwrotnego w stanach przejściowych zwrócili uwagę w 1966 roku Don G. Daugherty i Richard A. Greiner [4]. Praktyczne znaczenie tego problemu wzrosło, gdy do budowy wzmacniaczy zostały użyte tranzystory bipolarne. Okazało się wówczas, że tranzystorowe wzmacniacze o bardzo dobrych parametrach nie gwarantowały doskonałej jakości brzmienia odtwarzanego dźwięku. Subiektywna ocena jakości dźwięku była niska szczególnie w przypadku niektórych wzmacniaczy z silnym ujemnym sprzężeniem zwrotnym.

Termin „dynamiczne zniekształcenia intermodulacyjne” został wprowadzony przez Matti Otalę w 1970 roku [10]. Od tego czasu ukazało się wiele publikacji, w których zostały wyjaśnione przyczyny i mechanizmy powstawania dynamicznych zniekształceń intermodulacyjnych, zostały sformułowane zalecenia projektowe, których przestrzeganie umożliwia uniknięcie powstawania zniekształceń we wzmacniaczach z ujemnym sprzężeniem zwrotnym. Szczególnie istotne okazały się tu prace Petera Gardego (np. [5], [6]). Ważność prac Otali i Gardego polega na użyciu do badania dynamicznych zniekształceń intermodulacyjnych we wzmacniaczach mocy analizy w dziedzinie czasu zamiast tradycyjnie stosowanej analizy w dziedzinie częstotliwości.

Znaczna większość publikacji jest poświęcona badaniu dynamicznych zniekształceń intermodulacyjnych we wzmacniaczach z ujemnym sprzężeniem zwrotnym i kompensacją opóźnieniem fazowym (*lag compensation*). Tylko niewiele prac (np. [1], [2], [8], [9]) dotyczy wzmacniaczy z kompensacją wyprzedzeniem fazowym (*lead compensation*). Warunki, których spełnienie prowadzi do uniknięcia powstawania zniekształceń we wzmacniaczach z tą prostą, dobrze znaną i szeroko stosowaną kompensacją są nie zawsze prawidłowo rozumiane, a czasami błędnie interpretowane. Chery [1] twierdził, że jeśli wzmacniacz zostanieysterowany za pomocą sygnału o ograniczonym widmie częstotliwości, to zarówno kompensacja opóźnieniem fazowym w torze wzmocnienia, jak i kompensacja wyprzedzeniem fazowym w członie ujemnego sprzężenia zwrotnego nie wpływa na maksymalną wielkość napięcia sterującego stopień wejściowy wzmacniacza mocy. Wniosek ten jest jednak błędny, co zostanie wykazane w niniejszej pracy.

W pracy przedstawiono analizę warunków powstawania dynamicznych zniekształceń intermodulacyjnych we wzmacniaczach z ujemnym sprzężeniem zwrotnym i z kompensacją wyprzedzeniem fazowym w bloku ujemnego sprzężenia zwrotnego oraz we wzmacniaczach z kompensacją opóźnieniem fazowym w torze wzmocnienia. Podstawą analizy jest model wzmacniacza, w którym została uwzględniona nieliniowość stopnia wejściowego oraz właściwości pasmowe stopni (pojedyncze bieguny transmitancji stopnia pośredniego i wyjściowego). Wzmacniacz jest sterowany za pomocą fali prostokątnej filtrowanej przez filtr dolnoprzepustowy umieszczony na wejściu wzmacniacza. Wykazano, że rodzaj kompensacji wpływa w znacznym stopniu na maksymalną wielkość napięcia sterującego stopień wejściowy wzmacniacza i w ten sposób wpływa na podatność wzmacniacza na powodowanie

dynamicznych zniekształceń intermodulacyjnych. Zostanie wykazane, że maksymalne napięcie sterujące stopień wejściowy wzmacniacza w przypadku zastosowania kompensacji wyprzedzeniem fazowym jest znacznie mniejsze, niż w przypadku kompensacji opóźnieniem fazowym. Na podstawie wykonanej analizy sformułowano kryteria, których spełnienie gwarantuje, że w stopniu wejściowym wzmacniacza nie wystąpią dynamiczne zniekształcenia intermodulacyjne.

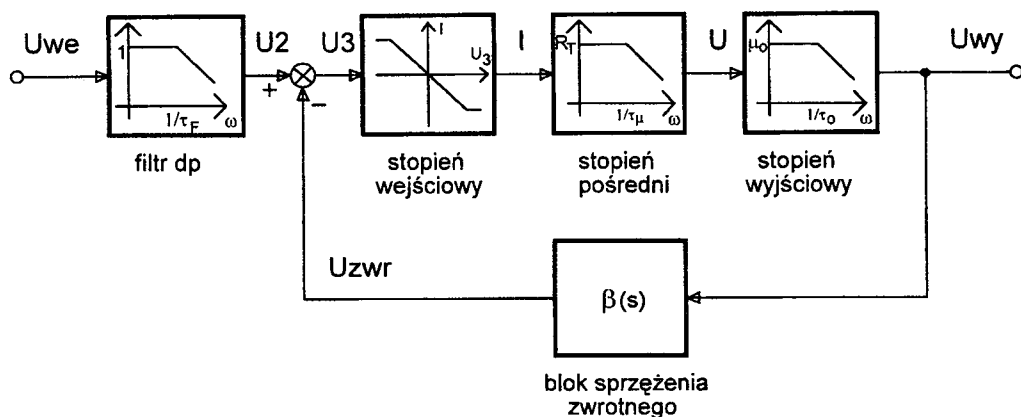
Część przedstawionych w tej pracy wyników była prezentowana na konferencji 94th Convention of the Audio Engineering Society w Berlinie [9]. Autorzy wyrażają wdzięczność uczestnikom dyskusji za cenne sugestie, których uwzględnienie przyczyniło się do podniesienia jakości tego opracowania.

2. DYNAMICZNE ZNIEKSZTAŁCENIA INTERMODULACYJNE WE WZMACNIACZACH Z KOMPENSACJĄ WYPRZEDZENIEM FAZOWYM

2.1. MODEL WZMACNIACZA

Typowy wzmacniacz mocy m.cz. składa się z trzech stopni wzmocnienia objętych ujemnym sprzężeniem zwrotnym i filtra dolnoprzepustowego na wejściu. Model takiego wzmacniacza, który stanowi podstawę do analizy warunków powstawania dynamicznych zniekształceń intermodulacyjnych, jest przedstawiony na rys. 1.

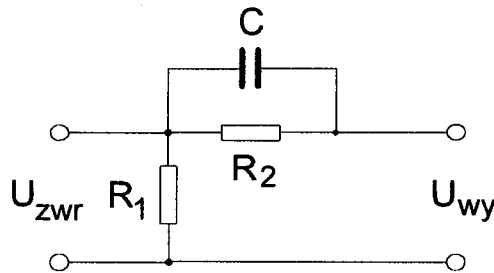
Stopień wejściowy wzmacniacza jest opisany za pomocą transkonduktancji $-G_T$ niezależnej od częstotliwości. Transmitancje stopni: pośredniego i wyjściowego są funkcjami jednobiegunowymi: transrezystancja stopnia pośredniego ma postać $-R_T/(1 + s\tau_\mu)$, a transmitancja napięciowo-napięciowa stopnia wyjściowego $\mu_0/(1 + s\tau_0)$. Przyjęcie takiego rodzaju transmitancji dobrze spełnia warunek wzaje-



Rys. 1. Schemat blokowy wzmacniacza mocy m.cz. z kompensacją wyprzedzeniem fazowym w bloku sprzężenia zwrotnego

mnego nieobciążania się poszczególnych bloków. Pominięcie ograniczania pasma przenoszenia stopnia wejściowego i aproksymacja funkcjami jednobiegunowymi transmitancji pozostałych stopni okazała się konieczna wobec trudności obliczeniowych i interpretacyjnych (zaproponowany model jest dokładniejszy i bardziej złożony od modeli dotąd opublikowanych np. [1]). Szczegółowe informacje na temat tego modelu, takie, jak: określenie klasy struktur wzmacniaczy mocy, dla których model może być stosowany, identyfikacja modelu, jego weryfikacja, ocena dokładności i zakresu zastosowań, zostały przedstawione w pracy [11].

Stopień wejściowy wzmacniacza jest sterowany ze źródła sygnału przez filtr dolnoprzepustowy pierwszego rzędu o transmitancji $1/(1 + s\tau_F)$. Drugie wejście tego stopnia jest sterowane napięciem zwrotnym, uzyskiwanym z wyjścia bloku ujemnego sprzężenia zwrotnego. Jego transmitancja jest funkcją z pojedynczym zerem i pojedynczym biegunem (w publikowanych dotąd pracach biegun tej transmitancji był pomijany, np. [1], [8]). Typowa struktura bloku ujemnego sprzężenia zwrotnego z obwodem kompensacji wyprzedzeniem fazowym jest pokazana na rys. 2.



Rys. 2. Schemat ideowy bloku ujemnego sprzężenia zwrotnego

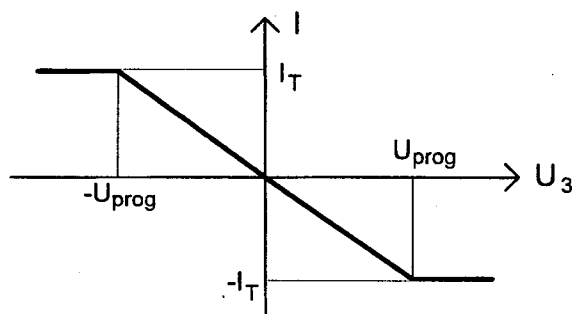
Transmitancja tego bloku jest równa:

$$\beta(s) = \frac{U_{zwr}}{U_{wy}} = \frac{R_1}{R_1 + R_2} \frac{1 + sR_2C}{1 + s \frac{R_1R_2}{R_1 + R_2}C} = \beta \frac{1 + s\tau_L}{1 + s\tau_P} \quad (1)$$

gdzie: $\beta = \frac{R_1}{R_1 + R_2}$, $\tau_L = R_2C$, $\tau_P = \frac{R_1R_2}{R_1 + R_2}C$, przy czym $\tau_P = \beta\tau_L$.

Przyczyną powstawania dynamicznych zniekształceń intermodulacyjnych jest przesterowanie stopnia wejściowego wzmacniacza w stanach przejściowych. Charakterystyka przejściowa stopnia wejściowego jest aproksymowana za pomocą funkcji:

$$I = \begin{cases} -G_T \cdot U_{prog}, & U_3 > U_{prog} \\ -G_T \cdot U_3, & |U_3| \leq U_{prog} \\ G_T \cdot U_{prog}, & U_3 < -U_{prog} \end{cases} \quad (2)$$



Rys. 3. Aproxymowana charakterystyka przejściowa stopnia wejściowego

Została ona przedstawiona na rys. 3. Przyjęcie innego sposobu aproksymacji tej charakterystyki np. aproksymacja za pomocą funkcji tangens hiperboliczny prowadzi do bardzo znacznych trudności obliczeniowych (porównaj [2]).

2.2. SYGNAŁ TESTOWY

Dynamiczne zniekształcenia intermodulacyjne powstające we wzmacniaczach mocy m.cz. są związane z ograniczaniem maksymalnej szybkości zmian sygnału wyjściowego wzmacniacza i zależą od rodzaju i parametrów sygnału pobudzającego wzmacniacz. Rzeczywisty sygnał akustyczny cechuje się ograniczoną maksymalną szybkością zmian. Unormowana maksymalna szybkość zmian sygnału muzycznego jest oceniana w wyniku różnych badań jako 3 do 26 V/ms/V [3]. Dla porównania unormowana szybkość zmian czystego sygnału sinusoidalnie zmiennego o częstotliwości 20 kHz wynosi 126 V/ms/V.

Sygnał testowy w opracowanej przez Otalę metodzie pomiaru dynamicznych zniekształceń intermodulacyjnych cechuje się również ograniczoną szybkością zmian, co jest wynikiem przepuszczania go przez filtr dolnoprzepustowy o górnej częstotliwości granicznej 100 kHz lub 30 kHz. Unormowana szybkość zmian fali prostokątnej przefiltrowanej przez filtr dolnoprzepustowy o częstotliwości granicznej 100 kHz wynosi 1250 V/ms/V.

W pracy [7] jako sygnał testowy przyjęto falę trapezoidalną o małej częstotliwości powtarzania. Chociaż takie pobudzenie dobrze modeluje sygnały rzeczywiste, to wyniki analizy wzmacniacza są trudne do interpretacji i nie są jednoznaczne, ponieważ zależą od przyjętego nachylenia czoła sygnału pobudzającego.

Najlepszym sygnałem testowym do badania warunków powstawania dynamicznych zniekształceń intermodulacyjnych jest fala prostokątna [5]. Jeżeli przy takim pobudzeniu dynamiczne zniekształcenia intermodulacyjne nie występują, to napewno nie wystąpią również przy każdym innym sygnale o takiej samej amplitudzie. Sygnały zakłócające, trzaski oraz niepożądane sygnały o wielkich częstotliwościach (np. fale nośne stacji radiowych, sygnał generatora podkładu magnetofonu, podnoś-

na pilota stereo itp.), które mogą się pojawić na wejściu wzmacniacza mocy, cechują się również bardzo dużymi szybkościami zmian.

Najtrudniejsze warunki pracy stopnia wejściowego (największe napięcie sterujące ten stopień w stanach przejściowych) wystąpią, gdy częstotliwość powtarzania fali prostokątnej pobudzającej wzmacniacz będzie tak mała, że następny skok pojawi się dopiero, kiedy wzmacniacz osiągnie stan ustalony po poprzednim skoku. Napięcie sterujące stopień wejściowy (U_3 na rys. 1) można wtedy wyznaczyć dla czasu nieujemnego, uważając sygnał U_{we} jako sumę dwóch pobudzeń [5]:

1) skoku napięcia od 0 do $-U_{1m}$, który wystąpił tak dawno, że w chwili $t = 0$ układ znajduje się w stanie ustalonym,

2) skoku napięcia od 0 do $2U_{1m}$ w chwili $t = 0$.

Taki sygnał pobudzający zostanie więc przyjęty do analizy warunków powstawania dynamicznych zniekształceń intermodulacyjnych.

2.3. ANALIZA WARUNKÓW POWSTAWANIA DYNAMICZNYCH ZNIEKSZTAŁCEŃ INTERMODULACYJNYCH

Wzmacniacz mocy z kompensacją wyprzedzeniem fazowym o strukturze przedstawionej na rys. 1 stanowi układ dynamiczny, nieliniowy i jego analiza w ogólnym przypadku jest problemem złożonym. Przeprowadzenie takiej analizy nie jest jednak niezbędne z punktu widzenia celu tej pracy. Najbardziej bowiem interesujące jest określenie warunków, w których dynamiczne zniekształcenia intermodulacyjne w stopniu wejściowym wzmacniacza nie wystąpią. Ponieważ zniekształcenia te są wynikiem przesterowania stopnia wejściowego w stanach przejściowych, więc aby ich uniknąć musi być spełniony podstawowy warunek:

$$\max_{t \geq 0} |U_3(t)| \leq U_{prog} \quad (3)$$

Celem analizy jest więc wyznaczenie maksymalnej wielkości napięcia sterującego stopień wejściowy wzmacniacza pobudzonego falą prostokątną $U_{3max} = \max_{t \geq 0} |U_3(t)|$

i określenie warunków projektowych, których spełnienie gwarantuje, że we wzmacniaczu nie wystąpią dynamiczne zniekształcenia intermodulacyjne, wynikające z przesterowania stopnia wejściowego.

Przy przyjętych założeniach związek pomiędzy sygnałami U_3 i U_{we} przedstawiony jest za pomocą zależności:

$$\frac{U_3}{U_{we}} = \frac{(1 + s\tau_\mu)(1 + s\tau_p)(1 + s\tau_0)}{(1 + s\tau_F)(1 + \mu\beta + sq + s^2p + s^3\tau_\mu\tau_p\tau_0)} \quad (4)$$

gdzie: $\mu = G_T R_T \mu_0$, $q = \tau_\mu + \tau_0 + \tau_p + \mu\beta\tau_L$, $p = \tau_\mu\tau_p + \tau_\mu\tau_0 + \tau_p\tau_0$.

Analiza warunków powstawania dynamicznych zniekształceń intermodulacyjnych na podstawie tej transmitancji jest złożona, a interpretacja wyników trudna. Te

trudności można ominąć w wyniku przyjęcia założeń co do położenia zera transmitancji bloku sprzężenia zwrotnego.

Założmy, że kompensacja wyprzedzeniem fazowym jest wprowadzona w taki sposób, że $\tau_L = \tau_0$. W takim przypadku zero $1/\tau_L$ kompensuje biegun $1/\tau_0$ zaś transmitancja (4) przyjmuje postać:

$$\frac{U_3}{U_{we}} = \frac{(s + 1/\tau_\mu)(s + 1/\tau_p)s_3}{(s + s_1)(s + s_2)(s + s_3)} \quad (5)$$

gdzie:

$$s_1 = \frac{1}{2\tau_\mu\tau_p} \left[\tau_\mu + \tau_p + \sqrt{(\tau_\mu - \tau_p)^2 - 4\mu\beta\tau_\mu\tau_p} \right],$$

$$s_2 = \frac{1}{2\tau_\mu\tau_p} \left[\tau_\mu + \tau_p - \sqrt{(\tau_\mu - \tau_p)^2 - 4\mu\beta\tau_\mu\tau_p} \right], \quad s_3 = \frac{1}{\tau_F}.$$

Bieguny s_1 i s_2 są rzeczywiste, jeśli $\mu\beta \leq \mu\beta_{\max} = \frac{(\tau_\mu - \tau_p)^2}{4\tau_\mu\tau_p}$ i mogą być zapisane jako:

$$s_1 = \frac{1+a}{2\tau_p} + \frac{1-a}{2\tau_\mu}, \quad s_2 = \frac{1-a}{2\tau_p} + \frac{1+a}{2\tau_\mu}, \quad \text{gdzie } a = \sqrt{1 - \frac{\mu\beta}{\mu\beta_{\max}}}.$$

Inne transmitancje w schemacie blokowym przedstawionym na rys. 1 są następujące:

– transmitancja toru wzmocnienia:

$$\frac{U_{wy}}{U_3} = \frac{\mu}{(1 + s\tau_\mu)(1 + s\tau_0)}, \quad (6)$$

– transmitancja pętli ujemnego sprzężenia zwrotnego:

$$\frac{U_{zwr}}{U_3} = \frac{\mu\beta}{(1 + s\tau_\mu)(1 + s\tau_p)}, \quad (7)$$

– transmitancja wzmacniacza z zamkniętą pętlą i filtrem:

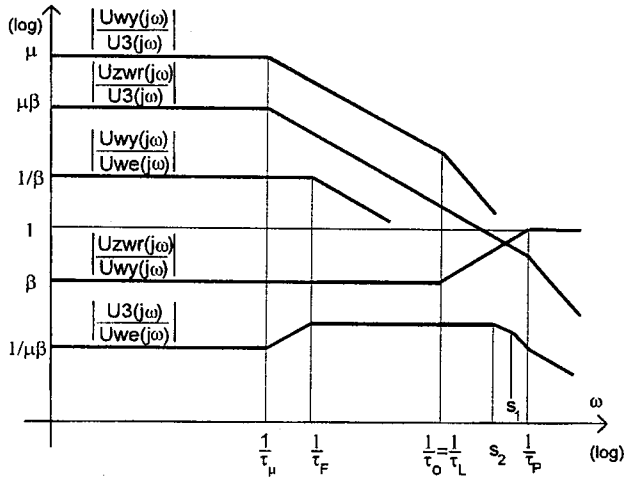
$$\frac{U_{wy}}{U_{we}} = \frac{\mu(1 + s\tau_p)}{(1 + s\tau_F)(1 + s\tau_0)[(1 + s\tau_\mu)(1 + s\tau_p) + \mu\beta]}. \quad (8a)$$

Gdy $|s\tau_p| \ll 1$ i $\mu\beta \gg 1$, to:

$$\frac{U_{wy}}{U_{we}} \approx \frac{1}{\beta(1 + s\tau_F)(1 + s\tau_0)(1 + s\tau_\mu/\mu\beta)} \quad (8b)$$

a gdy jeszcze $|s\tau_0| \ll 1$ i $|s\tau_\mu/\mu\beta| \ll 1$, to:

$$\frac{U_{wy}}{U_{we}} \approx \frac{1}{\beta(1 + s\tau_F)} \quad (8c)$$



Rys. 4. Asymptotyczne charakterystyki amplitudowe wzmacniacza z kompensacją wyprzedzeniem fazowym w bloku sprzężenia zwrotnego

Asymptotyczne charakterystyki Bodego ilustrujące zależności modułów transmitancji (6), (7) i (8) oraz (5) i (1) są przedstawione na rys. 4. Przy konstruowaniu tych charakterystyk zostały uwzględnione typowe wielkości parametrów wzmacniacza mocy m.cz. $\mu \approx 10^4 \div 10^5$, $\beta \approx 10^{-2} \div 10^{-1}$, $\tau_\mu \approx 7 \div 100 \mu\text{s}$, $\tau_F \approx 2 \div 5 \mu\text{s}$, $\tau_0 \approx 30 \div 100 \text{ ns}$. Na tej podstawie można sformułować nierówności obowiązujące we wzmacniaczach mocy m.cz.:

$$\beta \ll 1 \ll \mu\beta \ll \mu, \quad \frac{1}{\tau_\mu} < \frac{1}{\tau_F} \ll s_2 < s_1 < \frac{1}{\tau_P} \quad (9)$$

Uwzględnienie nierówności (9) okaże się niezbędne przy obliczaniu maksymalnej wielkości napięcia $U_3(t)$. Napięcie to uzyskamy na podstawie transformaty odwrotnej transmitancji (5) przy przyjęciu jako pobudzenia fali prostokątnej o amplitudzie U_{1m} i bardzo małej częstotliwości powtarzania (por. p. 2.2):

$$U_3(t) = U_{1m}[A + 2(Be^{-s_1 t} + Ce^{-s_2 t} + De^{-s_3 t})] \quad (10)$$

gdzie:

$$A = \frac{1}{1 + \mu\beta}, \quad B = \frac{-s_3 \left(s_1 - \frac{1}{\tau_\mu}\right) \left(s_1 - \frac{1}{\tau_P}\right)}{s_1(s_2 - s_1)(s_3 - s_1)},$$

$$C = \frac{-s_3 \left(s_2 - \frac{1}{\tau_\mu}\right) \left(s_2 - \frac{1}{\tau_P}\right)}{s_2(s_1 - s_2)(s_3 - s_2)}, \quad D = \frac{-\left(s_3 - \frac{1}{\tau_\mu}\right) \left(s_3 - \frac{1}{\tau_P}\right)}{(s_1 - s_3)(s_2 - s_3)}.$$

Funkcja (10) przyjmuje wartości $U_3(0) = \frac{-U_{1m}}{1 + \mu\beta}$ oraz $U_3(\infty) = \lim_{t \rightarrow \infty} [U_3(t)] =$

$= \frac{U_{1m}}{1 + \mu\beta}$. Aby ułatwić interpretację wyników zostaną uwzględnione nierówności (9). Wtedy:

$$\begin{aligned} \mu\beta_{max} &\approx \frac{\tau_\mu}{4\tau_p}, \quad s_1 \approx \frac{2\mu\beta}{(1-a)\tau_\mu}, \quad s_2 \approx \frac{2\mu\beta}{(1+a)\tau_\mu}, \quad A \approx \frac{1}{\mu\beta}, \\ B &\approx \frac{(1-a)^2}{4a\mu\beta} \frac{\tau_\mu}{\tau_F}, \quad C \approx \frac{(1+a)^2}{4a\mu\beta} \frac{\tau_\mu}{\tau_F}, \quad D \approx \frac{1}{\mu\beta} \left(\frac{\tau_\mu}{\tau_F} - 1 \right). \end{aligned} \quad (11)$$

Napięcie $U_3(t)$ osiąga wartość maksymalną dla czasu t^* . Wtedy jego pochodna jest równa zero, jednak analityczne rozwiązanie równania $\frac{dU_3(t)}{dt} = 0$ nie istnieje.

Aby oszacować to rozwiązanie uwzględnimy warunki (9):

$$(1+a) \cdot \exp \left[-\frac{2\mu\beta}{(1+a)} \cdot \frac{t^*}{\tau_\mu} \right] - (1-a) \cdot \exp \left[-\frac{2\mu\beta}{(1-a)} \cdot \frac{t^*}{\tau_\mu} \right] = \frac{2a}{\mu\beta} \cdot \left(\frac{\tau_\mu}{\tau_F} - 1 \right) \quad (12)$$

Jeżeli $a > a_{min}$, określonego przez (13),

$$\frac{1 + a_{min}}{1 - a_{min}} \left[\frac{2a_{min}}{1 + a_{min}} \frac{1}{\mu\beta} \left(\frac{\tau_\mu}{\tau_F} - 1 \right) \right]^{\frac{2a_{min}}{a_{min}-1}} \gg 1 \quad (13)$$

to drugi składnik lewej strony równania (12) może zostać pominięty i czas t^* można oszacować jako:

$$t^* \approx -\frac{1+a}{2\mu\beta} \tau_\mu \ln \left[\frac{2a}{(1+a)\mu\beta} \left(\frac{\tau_\mu}{\tau_F} - 1 \right) \right] \quad (14)$$

(w praktycznych układach wzmacniaczy mocy m.cz. z kompensacją wyprzedzeniem fazowym a_{min} jest tak małe, że jeśli $\mu\beta < 0,95 \cdot \mu\beta_{max}$, to nierówność (13) jest spełniona).

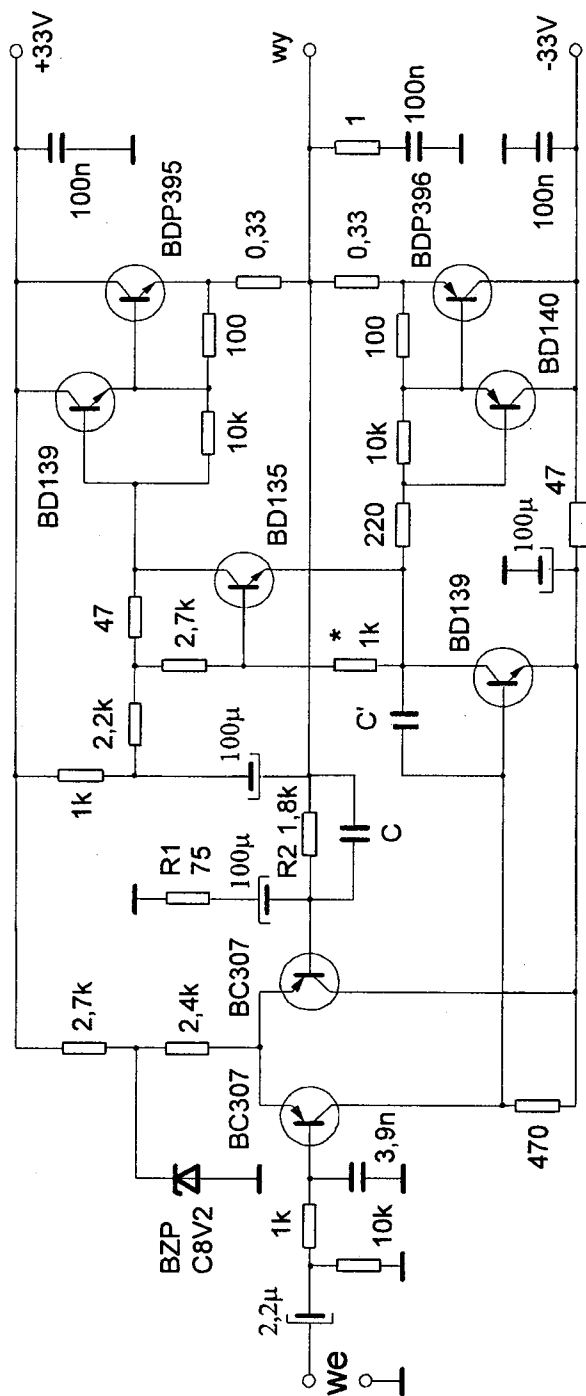
Maksymalna wartość napięcia sterującego stopień wejściowy jest zatem równa $U_{3max} \approx U_3(t^*)$. Jeżeli warunki (13) są spełnione to drugi i trzeci składnik prawej strony równania (10) może być pominięty i U_{3max} można zapisać jako:

$$U_{3max} \approx \frac{U_{1m}}{\mu\beta} \left\{ 1 + 2 \left(\frac{\tau_\mu}{\tau_F} - 1 \right) \left[\frac{2a}{1+a} \frac{1}{\mu\beta} \left(\frac{\tau_\mu}{\tau_F} - 1 \right) \right]^{\frac{1+a\tau_\mu}{2\mu\beta\tau_F}} \right\}, \quad (15)$$

a jeśli $\mu\beta$ przyjmuje typowe wartości (np. $\mu\beta > 0,1 \cdot \mu\beta_{max}$), to wyrażenie (15) można zapisać prościej, jako:

$$U_{3max} \approx \frac{U_{1m}}{\mu\beta} \left(2 \frac{\tau_\mu}{\tau_F} - 1 \right). \quad (16)$$

Na podstawie wzorów (16) i (15) można łatwo zauważyć, że we wzmacniaczu z kompensacją wyprzedzeniem fazowym maksymalna wielkość napięcia sterującego stopień wejściowy może przyjmować niewielkie wartości, jeżeli tylko stała czasu τ_μ



Rys. 5. Schemat ideowy przykładowego wzmacniacza mocy m.c.z.

jest bliska τ_F (pasmo wzmacniacza z otwartą pętlą sprzężenia zwrotnego jest niewiele węższe od pasma filtru wejściowego), a gdy τ_μ zmierza do τ_F (pasmo wzmacniacza z otwartą pętlą zmierza do pasma filtru wejściowego), to $U_{3\max}$ zmierza do $U_{3(\infty)}$ i przebieg nie występuje.

Warunek projektowy, którego spełnienie gwarantuje, że w stopniu wejściowym wzmacniacza z kompensacją wyprzedzeniem fazowym nie wystąpią dynamiczne zniekształcenia intermodulacyjne można zapisać jako:

$$\frac{U_{1m}}{\mu\beta} \left\{ 1 + 2 \left(\frac{\tau_\mu}{\tau_F} - 1 \right) \left[\frac{2a}{1+a} \frac{1}{\mu\beta} \left(\frac{\tau_\mu}{\tau_F} - 1 \right) \right]^{\frac{1+a}{2\mu\beta} \frac{\tau_\mu}{\tau_F}} \right\} \approx \frac{U_{1m}}{\mu\beta} \left(2 \frac{\tau_\mu}{\tau_F} - 1 \right) \leq U_{prog}. \quad (17)$$

2.4. PRZYKŁAD OBLICZENIOWY

W celu zilustrowania uzyskanych rezultatów rozpatrzmy przykładowy wzmacniacz o schemacie ideowym przedstawionym na rys. 5 (porównaj np. [8]).

Dla tego wzmacniacza parametry schematu blokowego z rys. 1 można przyjąć jako równe:

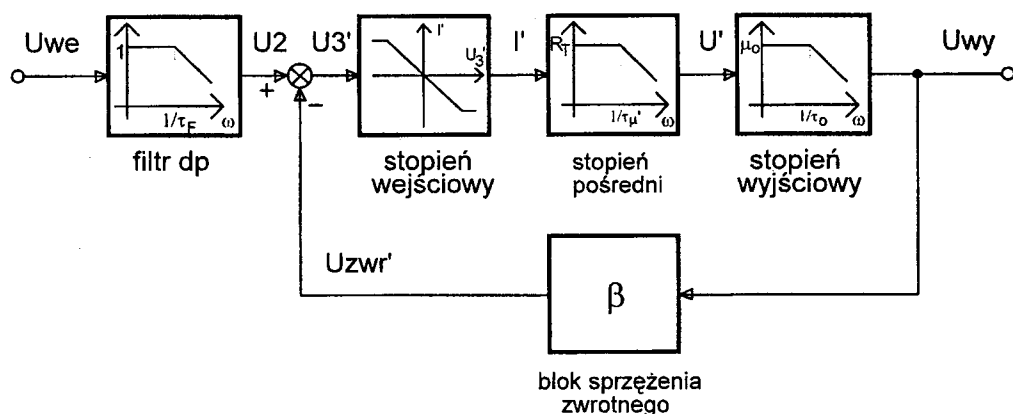
$G_1 = 30 \text{ mS}$, $R_T = 600 \text{ k}\Omega$, $\mu_0 = 0,9$, $R_1 = 75 \text{ }\Omega$, $R_2 = 1,8 \text{ k}\Omega$, $\tau_F = 4 \text{ }\mu\text{s}$, $\tau_\mu = 7 \text{ }\mu\text{s}$, $\tau_0 = 50 \text{ ns}$ i $U_{1m} = 0,775 \text{ V}$. Wzmocnienie pętli otwartej $\mu = G_T R_T \mu_0 = 16200$, współczynnik sprzężenia zwrotnego $\beta = R_1/(R_1 + R_2) = 0,04$, zaś wzmocnienie pętlowe $\mu\beta = 650$. Aby uzyskać kompensację wyprzedzeniem fazowym w torze sprzężenia zwrotnego, przy której $\tau_L = \tau_0$, pojemność kondensatora C (rys. 2) musi być równa $C = \tau_0/R_2 = 27 \text{ pF}$. Wtedy $\tau_p = CR_1 R_2/(R_1 + R_2) = 2 \text{ ns}$, zaś $\mu\beta_{\max} = 875$. W takim przypadku $s_1 = 378 \cdot 10^6$ i $s_2 = 123 \cdot 10^6$ są liczbami rzeczywistymi i nierówności (9) i (13) są spełnione. Na podstawie (14) $t^* = 58,2 \text{ ns}$ i na podstawie (16) $U_{3\max} = 3 \text{ mV}$.

Warunek projektowy, którego spełnienie gwarantuje, że w stopniu wejściowym wzmacniacza nie wystąpią dynamiczne zniekształcenia intermodulacyjne jest spełniony z zapasem, ponieważ dla stopnia wejściowego o strukturze pary różnicowej z tranzystorami bipolarnymi $U_{prog} = I_T/G_T = 52 \text{ mV} \gg U_{3\max} = 3 \text{ mV}$.

3. DYNAMICZNE ZNIEKSZTAŁCENIA INTERMODULACYJNE WE WZMACNIACZACH Z KOMPENSACJĄ OPÓŹNIENIEM FAZOWYM

3.1. ANALIZA WARUNKÓW POWSTAWANIA DYNAMICZNYCH ZNIEKSZTAŁCEŃ INTERMODULACYJNYCH

Struktura wzmacniacza z kompensacją opóźnieniem fazowym w torze wzmocnienia jest przedstawiona na rys. 6. Wzmacniacz ten różni się od przedstawionego na rys. 1 tym, że transmitancja bloku ujemnego sprzężenia zwrotnego nie zależy od częstotliwości (nie występuje tam kondensator C — porównaj rys. 2, więc $\tau_p = \tau_L = 0$). Kompensacja opóźnieniem fazowym w torze wzmocnienia polega na zwiększeniu za pomocą zewnętrznego kondensatora stałej czasu stopnia pośredniego τ_μ do τ'_μ .



Rys. 6. Schemat blokowy wzmacniacza mocy m.c.z. z kompensacją opóźnieniem fazowym w torze wzmocnienia

Aby uzyskać porównywalne warunki we wzmacniaczach z obu rodzajami kompensacji stała czasu τ'_μ powinna być tak duża, aby gwarantować stabilną pracę wzmacniacza z kompensacją opóźnieniem fazowym w takim samym zakresie zmienności wzmocnień pętlowych $\mu\beta$, jak to było w przypadku wzmacniacza z kompensacją wyprzedzeniem fazowym.

Transmitancje pętlowe w obu przypadkach dane są wzorami (6) i (18).

$$\frac{U'_{zwr}}{U'_3} = \frac{\mu\beta}{(1 + s\tau'_\mu)(1 + s\tau_0)} \quad (18)$$

Ponieważ w (6) $\tau_0 = \tau_L$ i $\tau_L = \tau_p/\beta$, to aby uzyskać jednakowe marginesy stabilności w obu wzmacniaczach, musi być spełniony warunek (19).

$$\tau'_\mu = \frac{\tau_\mu}{\beta} \quad (19)$$

Inne transmitancje w tym wzmacniaczu przyjmują postacie:

— transmitancja toru wzmocnienia:

$$\frac{U_{wy}}{U'_3} = \frac{\mu}{(1 + s\tau'_\mu)(1 + s\tau_0)} \quad (20)$$

— transmitancja wzmacniacza z zamkniętą pętlą i filtrem:

$$\frac{U_{wy}}{U_{we}} \approx \frac{\mu}{(1 + s\tau_F)(1 + s\tau_0)(1 + s\beta + s\tau'_\mu)} \quad (21a)$$

a gdy $\mu\beta \gg 1$, to

$$\frac{U_{wy}}{U_{we}} \approx \frac{1}{\beta(1 + s\tau_F)(1 + s\tau_0)(1 + s\tau'_\mu/\mu\beta)}, \quad (21b)$$

a gdy jeszcze $|s\tau_0| \ll 1$ i $|s\tau'_\mu/\mu\beta| \ll 1$, to:

$$\frac{U_{wy}}{U_{we}} \approx \frac{1}{\beta(1 + s\tau_F)} \quad (21c)$$

Związek między sygnałami U'_3 i U_{we} przedstawia transmitancja:

$$\frac{U'_3}{U_{we}} = \frac{(s + 1/\tau'_\mu)(s + 1/\tau_0)s_3}{(s + s'_1)(s + s'_2)(s + s_3)} \quad (22)$$

gdzie:

$$s'_1 = \frac{1}{2\tau'_\mu\tau_0} \left[\tau'_\mu + \tau_0 + \sqrt{(\tau'_\mu - \tau_0)^2 - 4\mu\beta\tau'_\mu\tau_0} \right],$$

$$s'_2 = \frac{1}{2\tau'_\mu\tau_0} \left[\tau'_\mu + \tau_0 - \sqrt{(\tau'_\mu - \tau_0)^2 - 4\mu\beta\tau'_\mu\tau_0} \right], \quad s_3 = \frac{1}{\tau_F}.$$

Bieguny s'_1 i s'_2 są rzeczywiste jeżeli $\mu\beta \leq \mu\beta'_{max} = \frac{(\tau'_\mu - \tau_0)^2}{4\tau'_\mu\tau_0}$ i mogą być zapisane, jako:

$$s'_1 = \frac{1 + a'}{2\tau_0} + \frac{1 - a'}{2\tau'_\mu}, \quad s'_2 = \frac{1 - a'}{2\tau_0} + \frac{1 + a'}{2\tau'_\mu}, \quad \text{gdzie } a' = \sqrt{1 - \frac{\mu\beta}{\mu\beta'_{max}}}.$$

Asymptotyczne charakterystyki Bodego, ilustrujące zależności modułów transmitancji (22), (21), (20) i (18) od pulsacji, zostały przedstawione na rys. 7. Przy konstruowaniu tych charakterystyk zostały uwzględnione typowe wielkości parametrów wzmacniacza mocy m.cz. Prowadzą one do sformułowania następujących nierówności obowiązujących we wzmacniaczach mocy m.cz. z kompensacją opóźnieniem fazowym:

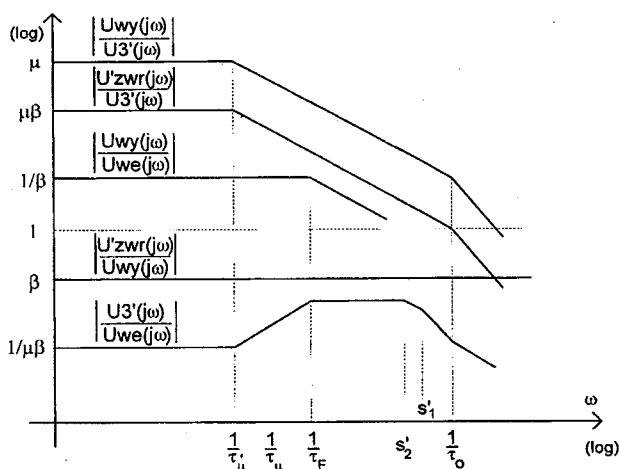
$$\beta \ll 1 \ll \mu\beta \ll \mu, \quad \frac{1}{\tau'_\mu} \ll \frac{1}{\tau_F} \ll s'_2 < s'_1 < \frac{1}{\tau_0}. \quad (23)$$

Nierówności te ułatwią oszacowanie maksymalnej wielkości napięcia $U'_3(t)$. Napięcie to jest wyznaczone z transformaty odwrotnej transmitancji (22) przy przyjęciu pobudzenia identycznego, jak w przypadku analizy wzmacniacza z kompensacją wyprzedzeniem fazowym. Ma ono postać:

$$U'_3(t) = U_{lm} [A + 2(B'e^{-s'_1 t} + C'e^{-s'_2 t} + D'e^{-s'_3 t})] \quad (24)$$

gdzie:

$$A = \frac{1}{1 + \mu\beta}, \quad B' = \frac{-s_3 \left(s'_1 - \frac{1}{\tau'_\mu} \right) \left(s'_1 - \frac{1}{\tau_0} \right)}{s'_1 (s'_2 - s'_1) (s_3 - s'_1)},$$



Rys. 7. Asymptotyczne charakterystyki amplitudowe wzmacniacza z kompensacją opóźnieniem fazowym w torze wzmocnienia

$$C' = \frac{-s_3 \left(s_2' - \frac{1}{\tau_\mu'} \right) \left(s_2' - \frac{1}{\tau_0} \right)}{s_2' (s_1' - s_2') (s_3 - s_2')}, \quad D' = \frac{- \left(s_3 - \frac{1}{\tau_\mu'} \right) \left(s_3 - \frac{1}{\tau_0} \right)}{(s_1' - s_3) (s_2' - s_3)}.$$

A pozostałe oznaczenia są takie same, jak w zależności (10).

Po uwzględnieniu nierówności (23) oraz gdy $a' > a'_{\min}$, które spełnia nierówność:

$$\frac{1 + a'_{\min}}{1 - a'_{\min}} \left[\mu\beta \frac{1 + a'_{\min}}{2a'_{\min}} \frac{\tau_F}{\tau_\mu'} \right]^{\frac{2a'_{\min}}{a'_{\min} - 1}} \gg 1, \quad (25)$$

maksimum sygnału sterującego stopień wejściowy wzmacniacza wystąpi przy czasie $t = t^*$

$$t^* \approx - \frac{1 + a'}{2\mu\beta} \tau_\mu' \ln \left[\frac{2a'}{(1 + a')\mu\beta} \frac{\tau_\mu'}{\tau_F} \right] \quad (26)$$

Wtedy maksymalna wielkość sygnału sterującego $U'_{3\max}$ może być oszacowana jako:

$$U'_{3\max} \approx \frac{2U_{1m}}{\mu\beta} \frac{\tau_\mu'}{\tau_F} \left[\frac{2a'}{1 + a'} \frac{1}{\mu\beta} \frac{\tau_\mu'}{\tau_F} \right]^{\frac{1 + a' \tau_\mu'}{2\mu\beta \tau_F}}, \quad (27)$$

a jeżeli $\mu\beta$ przyjmuje duże wartości (co jest typowe), to

$$U'_{3\max} \approx 2 \frac{U_{1m}}{\mu\beta} \frac{\tau_\mu'}{\tau_F} \quad (28)$$

Warunek projektowy, którego spełnienie gwarantuje, że w stopniu wejściowym wzmacniacza z kompensacją opóźnieniem fazowym nie wystąpią dynamiczne zniekształcenia intermodulacyjne można teraz sformułować jako:

$$\frac{2U_{1m}}{\mu\beta} \frac{\tau'_\mu}{\tau_F} \left[\frac{2a'}{1+a'} \frac{1}{\mu\beta} \frac{\tau'_\mu}{\tau_F} \right]^{\frac{1+a'\tau'_\mu}{2\mu\beta\tau_F}} \approx 2 \frac{U_{1m}}{\mu\beta} \frac{\tau'_\mu}{\tau_F} \leq U_{prog}. \quad (29)$$

3.2. PRZYKŁAD

Rozważmy ten sam wzmacniacz, co w p. 2.4 z kompensacją opóźnieniem fazowym zamiast kompensacji wyprzedzeniem fazowym. Oznacza to pominięcie kondensatora C ($C = 0$) i odpowiednie obniżenie bieguna stopnia pośredniego. Na podstawie (19) $\tau'_\mu = 175 \mu\text{s}$, zaś pozostałe parametry wzmacniacza (G_T , R_T , μ_0 , R_1 , R_2 , τ_p , τ_0 , U_{1m}) mają niezmiennione wartości. Wzmocnienia μ , $\mu\beta$, $\mu\beta_{\max}$ nie uległy zmianie ($\mu\beta'_{\max} = \mu\beta_{\max}$). Bieguny $s'_1 = 15,1 \cdot 10^6$ i $s'_2 = 4,9 \cdot 10^6$ są liczbami rzeczywistymi i nierówności (23) i (25) są spełnione. Na podstawie (26) $t'^* = 630 \text{ ns}$, a z zależności (28) $U'_{3\max} = 104 \text{ mV}$. Zauważmy, że maksymalne napięcie sterujące stopień wejściowy wzmacniacza jest teraz kilkadziesiąt razy większe, niż było w tym samym wzmacniaczu z kompensacją wyprzedzeniem fazowym. Kryterium projektowe (29) nie jest spełnione ($U_{prog} = 52 \text{ mV}$) i stopień wejściowy jest przesterowywany w stanach przejściowych, a więc we wzmacniaczu powstają dynamiczne zniekształcenia intermodulacyjne. Zniekształceń tych można uniknąć, mimo stosowania kompensacji opóźnieniem fazowym, przez powiększenie zakresu liniowej pracy stopnia wejściowego (zastosowanie lokalnego ujemnego sprzężenia zwrotnego dla sygnałów różnicowych w stopniu wejściowym z tranzystorami bipolarnymi, zastosowanie tranzystorów unipolarnych, itp. — porównaj [5], [6]).

3.3. PORÓWNANIE WARUNKÓW POWSTAWANIA DYNAMICZNYCH ZNIEKSZTAŁCEŃ INTERMODULACYJNYCH WE WZMACNIACZACH Z KOMPENSACJĄ WYPRZEDZENIEM I OPÓŹNIENIEM FAZOWYM

Zależności określające maksymalną wielkość napięcia sterującego stopień wejściowy wzmacniacza z kompensacją wyprzedzeniem fazowym (15) i (16), oraz napięcia sterującego ten stopień w przypadku tego samego wzmacniacza z kompensacją opóźnieniem fazowym (27) i (28) umożliwiają ilościową ocenę wpływu rodzaju kompensacji na podatność wzmacniacza na generowanie dynamicznych zniekształceń intermodulacyjnych w jego stopniu wejściowym. Porównując wielkości przepięć (różnic pomiędzy maksymalnymi wartościami napięć sterujących stopień wejściowy, a ich wartością w stanie ustalonym), na podstawie zależności (16) i (28) otrzymujemy:

$$\frac{U'_{3\max} - U_3(\infty)}{U_{3\max} - U_3(\infty)} \approx \frac{2 \frac{U_{1m}}{\mu\beta} \frac{\tau'_\mu}{\tau_F}}{\frac{U_{1m}}{\mu\beta} \left(2 \frac{\tau_\mu}{\tau_F} - 1 \right) - \frac{U_{1m}}{\mu\beta}} = \frac{\tau'_\mu}{\tau_\mu - \tau_F} > \frac{\tau'_\mu}{\tau_\mu} = \frac{1}{\beta}. \quad (30)$$

Jak widać przepięcie przyjmuje znacznie mniejsze wartości w przypadku kompensacji wzmacniacza wyprzedzeniem fazowym, niż w przypadku kompensacji opóźnieniem fazowym. Stosunek przepięć przy obu rodzajach kompensacji jest większy, niż $1/\beta$ (kilkudziesięciokrotny) i zmierza do nieskonczoności, gdy τ_μ zmierza do τ_F (oznacza to, że wartość maksymalna napięcia sterującego stopień wejściowy wzmacniacza z kompensacją wyprzedzeniem fazowym zmierza do wartości tego napięcia odpowiadającej stanowi ustalonemu).

4. WNIOSKI I UWAGI KOŃCOWE

Wyniki analizy warunków powstawania dynamicznych zniekształceń intermodulacyjnych we wzmacniaczach mocy m.cz., o modelach przyjętych w tej pracy, wskazują że jest możliwe zapewnienie liniowej pracy stopnia wejściowego wzmacniacza w stanach przejściowych, a więc zagwarantowanie nie występowania dynamicznych zniekształceń intermodulacyjnych w tym stopniu, jeżeli zostaną spełnione warunki projektowe (17) — w przypadku wzmacniacza z kompensacją wyprzedzeniem fazowym, lub (29) — w przypadku wzmacniacza z kompensacją opóźnieniem fazowym. Sformułowanie warunków projektowych w postaci tak prostych zależności stało się możliwe, ponieważ zostały uwzględnione założenia upraszczające (9) i (13) oraz (23) i (25) — odpowiednio, które obowiązują w praktycznych układach wzmacniaczy mocy.

Model wzmacniacza pokazany na rys. 1 jest bardziej złożony i lepiej odpowiada rzeczywistości wzmacniaczowi, niż modele spotykane dotąd w literaturze (por. [1], [8]). W uzupełnionym modelu zostały uwzględnione bieguny transmitancji stopnia pośredniego i wyjściowego wzmacniacza, jak również biegun i zero transmitancji bloku ujemnego sprzężenia zwrotnego.

Z zależności (16) wynika, że maksymalna wielkość napięcia sterującego stopień wejściowy wzmacniacza z kompensacją wyprzedzeniem fazowym może być w stanach przejściowych tylko niewiele większa, niż w stanie ustalonym, jeśli tylko górna częstotliwość graniczna wzmacniacza z otwartą pętlą ujemnego sprzężenia zwrotnego jest niewiele mniejsza, niż częstotliwość graniczna filtra dolnoprzepustowego, umieszczonego na wejściu wzmacniacza. W takim przypadku spełnienie warunku projektowego (17) jest łatwe. Aby wielkość niskoczęstotliwościowego wzmocnienia pętlowego mogła być w tych warunkach dostatecznie duża, drugi biegun transmitancji wzmacniacza $1/\tau_0$ powinien zostać skompensowany przez zero $1/\tau_L$ transmitancji bloku ujemnego sprzężenia zwrotnego.

Spełnienie warunku projektowego (29), dotyczącego wzmacniacza z kompensacją opóźnieniem fazowym, jest znacznie trudniejsze. Warunek ten może być spełniony przez zmniejszenie stałej czasu τ'_μ , co pociąga za sobą konieczność zmniejszenia niskoczęstotliwościowego wzmocnienia pętlowego, albo (i) stałej czasu τ_0 (zastosowanie tranzystorów mocy o wyższej częstotliwości granicznej). Innym sposobem spełnienia warunku (29) jest powiększenie zakresu napięć sterujących, odpowiadających liniowej pracy stopnia wejściowego U_{prog} . Modyfikacje stopnia wejściowego, prowadzące do zwiększenia jego dynamiki, są przedstawione np. w [6].

Uzyskane wyniki uzasadniają twierdzenie, że maksymalna wielkość napięcia sterującego stopień wejściowy wzmacniacza mocy objętego ujemnym sprzężeniem zwrotnym (28) jest w stanach przejściowych znacznie większa w przypadku zastosowania kompensacji charakterystyk częstotliwościowych metodą opóźnienia fazowego w torze wzmocnienia, niż w tym samym wzmacniaczu, w którym jednak została zastosowana kompensacja wyprzedzeniem fazowym w bloku sprzężenia zwrotnego (16). Na podstawie porównania odpowiednich warunków projektowych (17) i (29), a szczególnie na podstawie zależności 930) można stwierdzić, że liniową pracę stopnia wejściowego wzmacniacza w stanach przejściowych jest znacznie łatwiej zapewnić w przypadku zastosowania kompensacji wyprzedzeniem fazowym. Wzmacniacz z taką kompensacją jest więc mniej podatny na powstawanie w nim dynamicznych zniekształceń intermodulacyjnych.

BIBLIOGRAFIA

1. E.M. Cherry: *Transient Intermodulation Distortion — Part I: Hard Nonlinearity* IEEE Trans. Acoust. Speech and Sign. Proc., 1981, Vol. ASSP-29, No. 2, p. 137
2. E.M. Cherry, K.P. Dabke: *Transient Intermodulation Distortion — Part 2: Soft Nonlinearity*. J. Audio Eng. Soc., 1986, Vol. 34, No. 1/2, p. 19
3. R.R. Cordell: *Another View of TIM. Part 1*. Audio, Feb. 1980, Vol. 64, No. 2, pp. 38–49, Part 2, Audio, March 1980, Vol. 64, No. 3, p. 39
4. D.G. Daugherty, R.A. Greiner: *Some Design Objectives for Audio Power Amplifiers*. IEEE Trans. Audio Electroacoust., March 1966, Vol. AU-14, No. 1, p. 43
5. P. Garde: *Transient Distortion in Feedback Amplifiers*. J. Audio Eng. Soc., 1978, Vol. 26, No. 5, p. 314
6. P. Garde: *Amplifier First-Stage Criteria for Avoiding Slew-Rate Limiting*. J. Audio Eng. Soc., 1986, Vol. 34, No. 5, p. 349
7. M. Głowacki, M. Pierzchała, J. Stanlik: *Wpływ filtru dolnoprzepustowego i skończonej szybkości zmian sygnału wejściowego na warunki powstawania dynamicznych zniekształceń intermodulacyjnych we wzmacniaczach mocy z kompensacją biegunem dominującym*. Rozprawy Elektrotechniczne, 1987, TOM XXXIII, z. 1, s. 257
8. M. Głowacki, M. Pierzchała, J. Stanlik: *Warunki eliminacji dynamicznych zniekształceń intermodulacyjnych we wzmacniaczach z ujemnym sprzężeniem zwrotnym oraz kompensacją biegunem i zerem*. Rozprawy Elektrotechniczne, 1989, TOM XXXV, z. 3, s. 555
9. M. Głowacki, M. Pierzchała, J. Stanlik: *Transient Distortion in Feedback Audio Power Amplifiers*. 94th Convention of the Audio Engineering Society. Berlin (Germany), March 1993, paper D6-1
10. M. Otala: *Transient Distortion in Transistorized Audio Power Amplifiers*. IEEE Trans. Audio Electroacoust., 1970, Vol. AU-18, No. 3, p. 234
11. J. Stanlik: *Analiza wybranych struktur wzmacniaczy mocy m.cz. pod kątem zniekształceń*. Raport I-28 PWr, Seria SPR-061, Wrocław 1997, (w przygotowaniu)

M. GŁOWACKI, J. STANCLIK, M. PIERZCHAŁA

TRANSIENT INTERMODULATION DISTORTION IN AUDIO POWER AMPLIFIERS

S u m m a r y

In the paper an amplifier model, as well, as an input signal model are discussed, and an analysis of hard TIMD generation conditions in feedback audio power amplifiers with lead compensation in the feedback network and with lag compensation in forward path is presented. From the analysis design criteria are derived which enable to avoid hard TIMD in the first stage of the amplifiers with both kind of compensation. It has been proved that the peak first-stage error signal amplitude in a lead compensated amplifier is lower than in the same amplifier with lag compensation. It means that lead compensation asserts lower susceptibility to transient overdrive of the first stage of an amplifier than that attainable with lag compensation.

Key words: audio power amplifier, transient intermodulation distortion.

LISTA RECENZENTÓW KWARTALNIKA ELEKTRONIKI I TELEKOMUNIKACJI

Dr inż. *Jarosław Arabas*
Politechnika Warszawska,
Instytut Telekomunikacji

Doc. dr inż. *Zenon Baran*
Politechnika Warszawska,
Instytut Telekomunikacji

Prof. dr hab. inż. *Jerzy Baranowski*
Politechnika Warszawska,
Instytut Podstaw Elektroniki

Prof. dr hab. inż. *Daniel Józef Bem*
Politechnika Wrocławska,
Instytut Telekomunikacji i Akustyki

Dr hab. inż. *Wojciech Burakowski*
Politechnika Warszawska,
Instytut Telekomunikacji

Prof. dr hab. inż. *Andrzej Czajkowski*
Politechnika Łódzka,
Instytut Automatyki

Prof. dr hab. inż. *Zbigniew Czyż*
Przemysłowy Instytut
Telekomunikacji, Warszawa

Dr hab. inż. *Andrzej M. Dąbrowski*
Politechnika Warszawska,
Instytut Telekomunikacji

Prof. dr hab. inż. *Janusz Dobrogowski*
Politechnika Poznańska,
Instytut Elektroniki i Telekomunikacji

Prof. dr hab. inż. *Janusz Dobrowolski*
Politechnika Warszawska,
Instytut Podstaw Elektroniki

Dr inż. *Jarosław Domaszewicz*
Politechnika Warszawska,
Instytut Telekomunikacji

Dr inż. *Przemysław Dymarski*
Politechnika Warszawska,
Instytut Telekomunikacji

Prof. dr hab. inż. *Adam Fioł*
Politechnika Warszawska,
Instytut Radioelektroniki

Prof. dr hab. inż. *Anatol Gosiewski*
Politechnika Warszawska,
Instytut Automatyki i Informatyki Stosowanej

Prof. dr hab. inż. *Krzysztof Grabowski*
Politechnika Gdańska,
Katedra Techniki Mikrofalowej
i Telekomunikacji Optycznej

Dr inż. *Antoni Grzanka*
Politechnika Warszawska,
Instytut Podstaw Elektroniki

Prof. dr hab. inż. *Wojciech Gwarek*
Politechnika Warszawska,
Instytut Radioelektroniki

emer. Prof. dr hab. inż. *Jan Hennel*
Politechnika Warszawska,
Instytut Mikro- i Optoelektroniki

Prof. dr hab. inż. *Krzysztof Holejko*
Politechnika Warszawska,
Instytut Telekomunikacji

Prof. dr hab. inż. *Ryszard Jachowicz*
Politechnika Warszawska,
Instytut Podstaw Elektroniki

Dr inż. *Dariusz Janecki*
Politechnika Świętokrzyska,
Instytut Technologii Elektronowej

Prof. dr hab. inż. *Włodzimierz Janke*
Politechnika Gdańska,
Katedra Elektroniki Ciała Stałego

Prof. dr hab. inż. *Janusz M. Jaworski*
Politechnika Warszawska,
Instytut Elektrotechniki Teoretycznej
i Miernictwa Elektrycznego

Prof. dr hab. inż. *Romuald Józwicki*
Instytut Konstrukcji Przyrządów
Precyzyjnych i Optycznych,
Politechnika Warszawska

Prof. dr hab. inż. *Zbigniew Kaczkowski*
Polska Akademia Nauk,
Instytut Fizyki, Warszawa

Doc. dr hab. inż. *Andrzej Kasprzak*
Politechnika Wrocławska,
Instytut Sterowania i Techniki Systemów

Dr inż. *Jerzy Kęsik*
Politechnika Warszawska,
Instytut Mikro- i Optoelektroniki

Prof. dr hab. inż. *Jerzy Klamka*
Instytut Technologii
Elektronowej, Warszawa

Dr inż. *Paweł Kliś*
Instytut Łączności, Warszawa

Prof. dr hab. inż. *Andrzej Kobus*
Instytut Technologii
Elektronowej, Warszawa

Doc. dr hab. inż. *Adam Kowalewski*
Akademia Górniczo-Hutnicza,
Katedra Automatyki, Kraków

Mgr inż. *Franciszek Kotz*
Instytut Łączności, Warszawa

Prof. dr hab. inż. *Czesław Królikowski*
Politechnika Poznańska,
Instytut Elektroenergetyki

Dr inż. *Zdzisław Kozłowski*
Politechnika Warszawska,
Instytut Radioelektroniki

Prof. dr hab. inż. *Andrzej Kraśniewski*
Politechnika Warszawska,
Instytut Telekomunikacji

Doc. dr hab. inż. *Jerzy Kruszewski*
Politechnika Warszawska,
Instytut Mikroelektroniki i Optoelektroniki

Prof. dr hab. inż. *Adam Kujawski*
Politechnika Warszawska,
Instytut Fizyki

Prof. dr hab. inż. *Małgorzata Kujawińska*
Politechnika Warszawska,
Instytut Mikromechaniki i Fotoniki

Prof. dr hab. inż. *Juliusz Lech Kulikowski*
Polska Akademia Nauk, Warszawa,
Instytut Biocybernetyki
i Inżynierii Biomedycznej

Dr inż. *Anna Lewińska-Romicka*
Politechnika Warszawska,
Centrum Uczelniano-Przemysłowe Metrologii
i Systemów Pomiarowych

Dr inż. *Zbigniew Lisik*
Politechnika Łódzka,
Instytut Elektroniki

Prof. dr hab. inż. *Józef Lubacz*
Politechnika Warszawska,
Instytut Telekomunikacji

Prof. dr hab. inż. *Jerzy Luciński*
Politechnika Łódzka,
Instytut Elektroniki

Prof. dr hab. inż. *Tadeusz Łuba*
Politechnika Warszawska,
Instytut Telekomunikacji

Prof. dr hab. inż. *Jerzy Majewski*
Wyższa Szkoła Morska, Gdynia,
Wydział elektryczny

Prof. dr inż. *Władysław Majewski*
Politechnika Warszawska,
Instytut Telekomunikacji

Prof. dr inż. *Jan Malachowski*
Wojskowa Akademia Techniczna,
Instytut Fizyki Technicznej, Warszawa

Prof. dr hab. inż. *Bogdan Majkusiak*
Politechnika Warszawska,
Instytut Mikro- i Optoelektroniki

Prof. dr inż. *Marek Matczak*
Uniwersytet Szczeciński,
Wydział Mat.-Fiz., Katedra Fizyki

Prof. dr inż. *Andrzej Milewski*
Instytut Tele- i Radiotechniczny,
Warszawa

Prof. dr hab. inż. *Marian Milek*
Politechnika Zielonogórska,
Instytut Metrologii Elektrycznej

Prof. dr hab. inż. *Tadeusz Morawski*
Politechnika Warszawska,
Instytut Radioelektroniki

Prof. dr hab. inż. *Bogdan Mroziiewicz*
Instytut Technologii Elektronowej,
Warszawa

Prof. dr hab. inż. *Jan Mulawka*
Politechnika Warszawska,
Instytut Podstaw Elektroniki

Prof. dr hab. inż. *Andrzej Napieralski*
Politechnika Łódzka,
Instytut Elektroniki

Doc. dr inż. *Czesław Norek*
Instytut Tele- i Radiotechniczny,
Warszawa

Prof. dr hab. inż. *Stanisław Nowak*
Akademia Górniczo-Hutnicza, Kraków

Dr hab. inż. *Jan Ogrodzki*
Politechnika Warszawska,
Instytut Podstaw Elektroniki

Mgr inż. *Aleksander Orłowski*
Instytut Łączności, Warszawa

Prof. dr inż. *Jerzy Osiowski*
Politechnika Warszawska,
Instytut Podstaw Elektroniki

Dr inż. *Andrzej Pacut*
Politechnika Warszawska,
Instytut Automatyki
i Informatyki Stosowanej

Prof. dr hab. inż. *Władysław Paszek*
Politechnika Warszawska,
Instytut Maszyn i Urządzeń Elektrycznych

Dr hab. inż. *Jan Petykiewicz*
Politechnika Warszawska,
Instytut Fizyki

Dr inż. *Andrzej Pfizner*
Politechnika Warszawska,
Instytut Mikroelektroniki i Optoelektroniki

Prof. dr hab. inż. *Marian S. Piekarski*
Politechnika Wrocławska,
Instytut Telekomunikacji i Akustyki

Prof. dr hab. inż. *Jan Purczyński*
Politechnika Szczecińska,
Instytut Elektroniki i Informatyki

Doc. dr inż. *Olgierd Przesmycki*
Politechnika Warszawska,
Instytut Telekomunikacji

Dr hab. inż. *Roman Rykaczewski*
Politechnika Gdańska,
Katedra Systemów Informacyjnych

Dr inż. *Franciszek Seredyński*
Polska Akademia Nauk, Warszawa
Instytut Podstaw Informatyki

Prof. dr inż. *Stanisław Sławiński*
Politechnika Warszawska,
Instytut Telekomunikacji

Prof. dr hab. inż. *Stanisław Skoczowski*
Politechnika Szczecińska,
Wydział Elektryczny

Dr inż. *Jerzy Skwarczyński*
Akademia Górniczo-Hutnicza, Kraków,
Instytut Maszyn i Sterowania Układów
Elektroenergetycznych

Prof. dr hab. inż. *Tadeusz Sobczyk*
Akademia Górniczo-Hutnicza, Kraków
Instytut Maszyn i Sterowania Układów
Elektroenergetycznych

Prof. dr hab. inż. *Marek Stabrowski*
Politechnika Warszawska,
Instytut Elektrotechniki Teoretycznej
i Miernictwa Elektrycznego

Dr inż. *Jerzy Szabatin*
Politechnika Warszawska,
Instytut Podstaw Elektroniki

Prof. dr hab. inż. *Mieczysław Szustakowski*
Wojskowa Akademia Techniczna,
Warszawa
Instytut Fizyki Technicznej

Dr inż. *Tadeusz Świderski*
Wojskowy Instytut Łączności,
Zakład Radiokomunikacji, Zegrze

Prof. dr inż. *Michał Tadeusiewicz*
Politechnika Łódzka,
Instytut Podstaw Elektrotechniki

Dr hab. inż. *Aleksander Urbaś*
Politechnika Warszawska,
Instytut Podstaw Elektrotechniki

Dr hab. inż. *Krzysztof Wesolowski*
Politechnika Poznańska,
Instytut Elektroniki i Telekomunikacji

Dr hab. inż. *Tadeusz Więckowski*
Politechnika Wrocławska,
Instytut Telekomunikacji i Akustyki

Prof. dr hab. inż. *Jacek Wojciechowski*
Politechnika Warszawska,
Instytut Podstaw Elektroniki

Dr hab. inż. *Tomasz Woliński*
Politechnika Warszawska,
Instytut Fizyki

Dr inż. *Włodzimierz Wolski*
Politechnika Wrocławska,
Instytut Telekomunikacji i Akustyki

Prof. dr hab. inż. *Andrzej Wojtkiewicz*
Politechnika Warszawska,
Instytut Podstaw Elektroniki

Dr inż. *Kornel B. Wydro*
Politechnika Warszawska,
Instytut Automatyki
i Informatyki Stosowanej

Prof. dr hab. inż. *Jan Zabrodzki*
Politechnika Warszawska,
Instytut Informatyki

Dr inż. *Krzysztof Zamłyński*
Politechnika Warszawska,
Instytut Podstaw Elektroniki

Dr inż. *Bogdan Zbierchowski*
Politechnika Warszawska,
Instytut Telekomunikacji

Dr inż. *Piotr Zwierko*
Politechnika Warszawska,
Instytut Telekomunikacji

SPIS TREŚCI

M.S. M a k h o u l: Efekty zniekształcenia złożonego sygnału w pasmie częstotliwości SCM AM/VSB CATV CD	443
A. S a d o w s k i: Projektowanie permutacji bitów sieci podstawieniowo-permutacyjnych ...	473
J. S i u z d a k: System segmentacji obrazu i śledzenia obiektów w czasie rzeczywistym	489
J. S i u z d a k, T. C z a r n e c k i: Nowy rodzaj filtru dla detekcji krawędzi na obrazie ...	509
H. G r u c h a ł a: Projektowanie szerokopasmowych zwrotnic mikrofalowych z zastosowaniem tożsamości Kurody	517
A. S a d o w s k i: Proste sprzętowe implementacje modyfikowalnych permutacji bitów oparte na macierzowym opisie permutacji	533
M. G ł o w a c k i, J. S t a n c l i k, M. P i e r z c h a ł a: Dynamiczne zniekształcenia intermodulacyjne we wzmacniaczach mocy m.cz.	551
Lista recenzentów	569

CONTENTS

M.S. M a k h o u l: Intermodulation distortion effects on the bandwidth frequency of the SCM AM/VSB CATV CD system	443
A. S a d o w s k i: Design of bit permutations for substitution-permutation networks	473
J. S i u z d a k: Real-time image segmentation system	489
J. S i u z d a k, T. C z a r n e c k i: A novel filter for edge detection	509
H. G r u c h a ł a: Designing of broadband microwave multiplexers with using of Kurosada's identities	517
A. S a d o w s k i: Simple hardware implementations of modifiable bit permutations based on matrix description of permutation	533
M. G ł o w a c k i, J. S t a n c l i k, M. P i e r z c h a ł a: Transient intermodulation distortion in audio power amplifiers	551
List of the reviewers	569

INFORMACJE DLA AUTORÓW

Redakcja przyjmuje do publikowania prace oryginalne, przeglądowe i monograficzne wchodzące w zakres szeroko pojętej 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 powinny 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 30 stron po około 1800 znaków na stronie.

Wymagania podstawowe: Artykuły należy nadsyłać w maszynopisie pisany 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.

Sposób pisania tekstu: Tekst powinien być pisany bez używania wyróżnień, a w szczególności nie dopuszcza się spacjiowania, podkreślania i 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 spaciowanie (rozstrzelenie), podkreślenie linią ciągłą — pogrubienie, podkreślenie wężykiem — kursywa. Tekst powinien być napisany z podwójnym odstępem między wierszami, tytuły i podtytuły małymi literami. Marginesy z każdej strony powinny mieć około 35 mm. Przy podziale pracy na rozdziały 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 poziomych 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 oryginalny 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 (strzałki, 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 literowe i graficzne powinny być zgodne z wytycznymi IEE (International Electronical Commision) 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 resource allocation framework. XVI International Symposium. Stockholm (Sweden), May 1991, paper A 2.4
- książki: Y.P. Tvidis: Operation and modelling of the MOS transistors. Mc Graw-Hill, New York 1987, p. 141 – 148

Materiał ilustracyjny: Rysunki powinny być wykonane wyraźnie, na papierze gładkim, lub milimetrowym w formacie nie mniejszym niż 9×12 cm. Mogą być także w postaci wydruku komputerowego. Fotografie lub diapozytywy przyjmowane są barwne 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 powinien być umieszczony na oddzielnej stronie.

Streszczenia: Do każdego artykułu musi być dołączone obszerne — do 60 wierszy — streszczenie z tytułem artykułu. Streszczenia (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. 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 ciągu 3 dni od daty otrzymania tekstu z Redakcji i zwrócić osobiście, lub listownie pod adresem Redakcji. 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 ją o zmianie miejsca pracy i adresu prywatnego.

INFORMATION FOR THE AUTHORS OF K.E.T.

The KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI (K.E.T.) — ELECTRONICS AND TELECOMMUNICATION QUARTERLY is a journal of the Committee for Electronics and Telecommunications of the Polish Academy of Sciences and is a direct continuation for 43 years of the quarterly ROZPRAWY ELEKTROTECHNICZNE, that has been published also under the auspices of the Polish Academy of Sciences.

Its pages contain scientific articles concerned with theory and applications of electronics, telecommunication, microelectronics, optoelectronics, radioengineering and medical electronics.

Accordingly, its pages contain scientific articles, that are original works both contributory and concerned with comparative analyses of methods or systems synthetic reviews of a given field, state-of-art discussions of a given field, and presentations of developments in a given technology.

Regular papers are accepted for publication in Polish or English, but title, the extensive summary and figure caption must be bilingual. Bilingual summary must shown the significance of the results shown in the paper, emphasizing limitations and advantages, possible applications and recommendations for further works.

K.E.T. is meant for specialists and students with above subjects.

FORMAL PREPARATION OF MANUSCRIPT

Texts

Texts to be submitted to K.E.T. may be typed on one side of the paper only with double spacing between the lines and a 4 cm margin or using of wordprocessing with text formatters based on MS DOS, APPLE/MAC such as Microsoft word. In any case the text formatter must be specified on the diskette. Each text should be preceded by: a sheet with paper title, name (in full) and surname of the Author/s, their Affiliations and relevant addresses.

All text pages should be numbered. Regular papers should have a length of no more than 30 pages.

Printing types and symbols

For letteral symbols, units and graphical symbols, the recommendations of IEE (International Electrotechnical Commission) and ISO (International Organisation of Standardisation) must be followed.

- In case of manuscripts conventionally typed (i.e. without wordprocessing), letters to be printed in italic should be underlined; letters to be printed in bold should be double interlined.
- 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.
- The serial number of footnotes should be written between round brackets and upper.
- Mathematical details, ancillary to the main discussion of the paper may be included in one or more appendixes, that are printed after conclusions and before references.

In manuscript containing lot of symbols, a glossary may be included after introduction.

References

References are usually gathered at the end of the text and numbered according to the order of quotation in the text. The serial numbers should be written without square brackets (i.e. 1., 2. ...)

Illustrations

The originals of illustrations may be either drawings or photographs (colour or black). Each copy of paper must include two complete set of illustrations.

All rights reserved. No part of the publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means: electronic, photocopying, recording or otherwise, without the prior permission of the copyright owner — Redaction K.E.T.

**POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI**

**KWARTALNIK
ELEKTRONIKI I TELEKOMUNIKACJI**

**ELECTRONICS AND
TELECOMMUNICATIONS
QUARTERLY**

SPIS TREŚCI DO TOMU 43

**WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1998**

SPIS TREŚCI DO TOMU 43 — 1997

Nr 1

Od redakcji	3
S. K o z i e ł: Algorytmy ewolucyjne w zagadnieniach optymalizacji numerycznej z ograniczeniami na zbiorach wypukłych	5
S. K o z i e ł: Optymalizacja wielokryterialna układów elektronicznych z wykorzystaniem metod ewolucyjnych	19
A. B o r y s: O aproksymacji nieliniowych systemów cyfrowych	31
A. B o r y s: Przekształcenia komplementarne i zniekształcenia intermodulacyjne	57
Z. W r ó b e ł: Analiza nulatorowa układów zawierających konweytor prądowy	71
P. A n t o s z c z y s z y n, E. K o r n a t o w s k i, W. L i p i ń s k i: Skuteczne śledzenie cech charakterystycznych twarzy w ruchomych sekwencjach wideo	83
T.M. A d a m o w i c z, M. K o c i k, J. M i z e r a c z y k, J. M e n t e l: Własności laserów podczerwieni w rozpylanych mieszaninach He—Cu przy pobudzaniu wyładowaniem wielkiej częstotliwości	99
W. P o g r i b n y, Z. D r z y c i m s k i, P. K i e d r o w s k i: Kompresja danych oraz analiza ekstremalna z wykorzystaniem liniowej modulacji delta	109
K. W i a t r: Architektura specjalizowanych procesorów potokowych do obróbki wstępnej danych wizyjnych w czasie rzeczywistym	120
S. K u t a, R. W y d m a ń s k i: Dokładny makromodel konweytor prądowego CCII [±]	149
A.Z. A b d a l l a: O projektowaniu 3 dB dzielników mocy opartych o strukturę sprzęgacza pierścieniowego	167

Nr 2

S.K. P a n d e y: Samonastrajający regulator dla niestabilnych modeli	181
J. S i u z d a k, Z. K w a ń s k i: Porównanie szybkości działania wybranych algorytmów detekcji i obróbki krawędzi	197
D. C z e k a j, Z. S u r o w i a k, V.A. A l y o s h i n, V. B i r y u k o v, V.P. D u d k i e v i c h, I.M. S e m, E.V. S v i r y d o v: Samoistna polaryzacja cienkowarstwowych sensorów piro- i piezoelektrycznych	221
J. D u d e k, D. C z e k a j, Z. S u r o w i a k, A.V. G o r i s h, Y.N. K o p t e v, A.A. K u p r i e n k o, A.E. P a n i c h: Wielowarstwowe przetworniki piezoelektryczne o wysokiej stabilności temperaturowej	233
M.S. M a k h o u l: Koherentna modulacja cyfrowa wielopoziomowa dla systemów CSM ...	239
A. C y s e w s k a - S o b u s i a k: Metrologiczne aspekty pulsooksymetrii	257
J. H a l a w a: Wyznaczanie współczynników modelu zastępczego a dobór nastaw regulatorów PI	279

Nr 3

H. K u n e r t: Identyfikacja metodą momentów i metodą powierzchni — wnioski z badań numerycznych	301
Cz. S t e f a ń s k i: O obliczaniu wrażliwości z wykorzystaniem symbolicznych funkcji układowych o postaci zagnieżdżonej	313
L. T o m a w s k i, M. M a ń k a, J. D ą b r o w s k i: Nieliniowa pojemność ujemna w układzie oscylatorowym	321
A. S a d o w s k i: Zależne od klucza permutacje bitów w sieciach podstawieniowo-permutacyjnych	333
M. W ę g r z y n, M. A d a m s k i: Wykorzystanie standardowych kompilatorów FPLD do syntezy sterowników logicznych	353
S. K u t a, G. K r a j e w s k i, J. J a s i e l s k i: Kształtowanie charakterystyk częstotliwościowych wzmacniaczy z prądowym sprzężeniem zwrotnym	373
W. P a w e l s k i: Inteligentne moduły mocy z tranzystorami IGBT	397
J. K i s i l e w i c z: Korekcja interferencji metodą przewidywania nadchodzących danych ...	407
B. W o ł c z a k: Ocena szerokości pasm w analizie widmowej i mikrometrii	421

Nr 4

M.S. M a k h o u l: Efekty zniekształcenia złożonego sygnału w pasmie częstotliwości SCM AM/VSB CATV CD	443
A. S a d o w s k i: Projektowanie permutacji bitów sieci podstawieniowo-permutacyjnych ...	473
J. S i u z d a k: System segmentacji obrazu i śledzenia obiektów w czasie rzeczywistym	489
J. S i u z d a k, T. C z a r n e c k i: Nowy rodzaj filtru dla detekcji krawędzi na obrazie ...	509
H. G r u c h a ł a: Projektowanie szerokopasmowych zwrotnic mikrofalowych z zastosowaniem tożsamości Kurody	517
A. S a d o w s k i: Proste sprzętowe implementacje modyfikowalnych permutacji bitów oparte na macierzowym opisie permutacji	533
M. G ł o w a c k i, J. S t a n c l i k, M. P i e r z c h a ł a: Dynamiczne zniekształcenia intermodulacyjne we wzmacniaczach mocy m.cz.	551
Lista recenzentów	569

CONTENTS

Nr 1

S. K o z i e ł: Evolutionary algorithms in constrained numerical optimization problems on convex spaces	5
S. K o z i e ł: Multiobjective optimization of electronic circuits using evolutionary method ...	19
A. B o r y s: An approximation of nonlinear digital systems	37
A. B o r y s: Complementary transformation and intermodulation distortion	57
Z. W r ó b e l: Nullor analysis of networks containing current conveyor	71
P. A n t o s z c z y s z y n, E. K o r n a t o w s k i, W. L i p i ń s k i: Reliable recovery of characteristic features motion in video sequences	83
T.M. A d a m o w i c z, M. K o c i k, J. M i z e r a c z y k, J. M e n t e l: Infrared laser properties of sputtered He—Cu mixtures excited by radio-frequency and hollow-cathode discharges	99
W. P o g r i b n y, Z. D r z y c i m s k i, P. K i e d r o w s k i: The data compression and extreme analysis using linear delta modulation	109
K. W i a t r: Specialised hardware processors architecture for a real-time fast image data pre-processing	120
S. K u t a, R. W y d m a ń s k i: Accurate current conveyor CCH [±] macromodel	149
A. Z. A b d a l l a: Design considerations on the 3 dB hybrid-ring based power dividers	167

Nr 2

S.K. Pandey: A self tuning controller for unstable models	181
J. Siuzdak, Z. Kwaśnik: The comparison of the methods of edge detection and processing speed	197
D. Czekaj, Z. Surowiak, V.A. Alyoshin, S.V. Biryukov, V.P. Dudkevich, I.M. Sem, E.V. Svirydov: Intrinsic polarization of the pyro- and piezoelectric thin film sensors	221
J. Dudek, D. Czekaj, Z. Surowiak, A.V. Gorish, Y.N. Koptev, A.A. Kuprienko, A.E. Panich: Temperature independent multilayer piezoelectric transducers	233
M.S. Makhouli: Using M-ary digital modulation for coherent optical subcarrier multiplexing SCM system	239
A. Cysewska-Sobusiak: Metrological aspects of pulse oximetry	257
J. Halawa: The obtaining of the time constant and the methods of setting the PI controllers	279

Nr 3

H. Kurnert: Identification of the two methods of momentum and surface — summary of numeric examination	301
Cz. Stefański: On sensitivity calculations based on symbolic network functions	313
L. Tomawski, M. Mańka, J. Dąbrowski: Nonlinear negative capacitance in oscillator circuit	321
A. Sadowski: Key-controlled bit permutations in substitution-permutation networks	333
M. Węgrzyn, M. Adamski: Synthesis of logic controllers using standard FPLD compilers	353
S. Kuta, G. Krajewski, J. Jasielski: CMOS current feedback operational amplifiers frequency characteristics forming	373
W. Pawelski: The intelligent power modules with IGBT switches	397
J. Kisilewicz: Intersymbol interference equalization using further decision prediction ...	407
B. Wołczak: Evaluation of bandwidth in spectral analysis and micrometry	421

Nr 4

M.S. Makhouli: Intermodulation distortion effects on the bandwidth frequency of the SCM AM/VSB CATV CD system	443
A. Sadowski: Design of bit permutations for substitution-permutation networks	473
J. Siuzdak: Real-time image segmentation system	498
J. Siuzdak, T. Czarnecki: A novel filter for edge detection	509
H. Gruchała: Designing of broadband microwave multiplexers with using of Kurosada's identities	517
A. Sadowski: Simple hardware implementations of modifiable bit permutations based on matrix description of permutation	533
M. Głowacki, J. Stanclik, M. Pierzchała: Transient intermodulation distortion in audio power amplifiers	551
List of the reviewers	569