POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

TOM 52 — ZESZYT 2

WARSZAWA 2006

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

Dyżury Redakcji: środa i piątki, godz. 14–16
tel. (022) 660 77 37

Telefony domowe: Redaktora Naczelnego: 812 17 65
Zast. Red. Naczelnego: 826 83 41
Sekretarza Odpowiedzialnego: 633 92 52

Ark. wyd. 12,0 Ark. druk. 10,0	Podpisano do druku w lutym 2006 r.
Papier offset. kl. III 80 g. B-1	Druk ukończono w marcu 2006 r.

Skład, druk i oprawa: Warszawska Drukarnia Naukowa PAN
00-656 Warszawa, ul. Śniadeckich 8
Tel./fax: 628-87-77

Szanowni Autorzy

„Kwartalnik Elektroniki i Telekomunikacji" — Electronics and Telecommunications Quartery jest kontynuatorem tradycji powstałego 49 lat temu kwartalnika p.t. „Rozprawy Elektrotechniczne".

Kwartalnik jest czasopismem Komitetu Elektroniki i Telekomunikacji Polskij Akademii Nauk. Wydawany jest przez Wydawnictwo Naukowe PWN SA w Warszawie. Kwartalnik jest czasopismem naukowym, na którego łamach są publikowane artykuły i komunikaty prezentujące wyniki oryginalnych prac teoretycznych i doświadczalnych, a także przeglądowych. Związane są one z szeroko rozumianymi dziedzinami współczesnej elektroniki, telekomunikacji, mikroelektroniki, oproelektroniki, radiotechniki i elektroniki medycznej.

Autorami publikacji są wybitni naukowcy, znani specjaliści o wieloletnim doświadczeniu, a także młodzi badacze — głównie doktoranci.

Artykuły charakteryzują się oryginalnym ujęciem zagadnienia, interesującymi wynikami badań, krytyczną oceną teorii lub metod, omówieniem aktualnego stanu, lub postępu danej gałęzi techniki oraz omówieniem perspektyw rozwojowych. Sposób pisania matematycznej części artykułów zgodny jest z wytycznymi IEC (International Electronics Commision) oraz ISO (International Organization of Standarization).

Wszystkie publikowane w Kwartalniku artykuły są recenzowane przez znanych krajowych specjalistów, co zapewnia że publikacje te są uznawane jako autorski dorobek naukowy. Opublikowane w kwartalniku wyników prac naukowych zrealizowanych w ramach „GRANTów" Komitetu Badań Naukowych spełnia więc jeden z wymogów stawianych tym pracom.

Czasopismo dociera do wszystkich zajmujących się elektroniką i telekomunikacją krajowych ośrodków naukowych oraz technicznych, a także szeregu instytucji zagranicznych. Jest ponadto prenumerowane przez liczne grono specjalistów i biblioteki.

Każdy Autor otrzymuje bezpłatnie 20 egzemplarzy nadbitek swojego artykułu, co ułatwia przesłanie go do indywidualnych wybranych przez Autora osób i instytucji w kraju lub za granicą. Ułatwia to dodatkowo fakt, że w Kwartalniku są publikowane artykuły w języku angielskim.

Nadesłane do redakcji artykuły są publikowane w termini około pół roku, w przypadku sprawnj współpracy Autora z Redakcją. Wytyczne dla Autorów dotycząc formy publikacji są zamieszczone w zeszytach Kwartalnika, można je także otrzymać w siedzibie Redakcji.

Artykuły można dostarczać osobiście, lub pocztą pod adresem zamieszczanym na stronie redakcyjnej w każdym zeszycie.

Redakcja

SPIS TREŚCI

CONTENTS

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2006, 52, z. 2, ss. 141–152

Zastosowanie algorytmu Rate Monotonic Scheduling do szeregowania zadań w architekturach wieloprocesorowych o topologii hipersześcianu

MIROSŁAW GAJER

Katedra Automatyki, Akademia Górniczo-Hutnicza
al. Mickiewicza 30, 30-059 Kraków
mgajer@ia.agh.edu.pl

Otrzymano 2005.01.04
Autoryzowano 2005.03.21

Artykuł stanowi propozycję rozszerzenia obszaru stosowalności popularnego algorytmu szeregujące zadania Rate Monotonic Scheduling (RMS) na przypadek zadań wieloprocesorowych. Dotychczas algorytm RMS wykorzystywany był do szeregowania zbioru niezależnych, wywłaszczalnych i periodycznych zadań przeznaczonych tylko dla jednego procesora. Rosnąca coraz bardziej popularność rozwiązań wieloprocesorowych wymusza dokonanie takiej adaptacji algorytmu RMS, aby algorytm ten nadawał się również do szeregowania zadań wieloprocesorowych. Autor skupił swoją uwagę na architekturach wieloprocesorowych o topologii hipersześcianu. Topologia ta charakteryzuje się bardzo korzystnym stosunkiem liczby połączeń komunikacyjnych pomiędzy poszczególnymi jednostkami obliczeniowymi do maksymalnej długości drogi przesyłu komunikatu, dzięki czemu stanowi jedno z bardziej popularnych rozwiązań wieloprocesorowych. Zaproponowane przez autora zastosowanie klasycznego algorytmu RMS do szeregowania zbioru zadań realizowanych w systemie równoległym o topologii hipersześcianu, polega na transformacji wartości okresów niektórych z zadań. W wyniku rozważanej transformacji wybrane zadania uzyskują identyczne wartości swoich okresów, dzięki czemu mogą zostać połączone w jedno większe tzw. superzadanie, do realizacji którego wymagana jest jednoczesna dostępność wszystkich jednostek obliczeniowych występujących w systemie. Następnie zbiór superzadań może zostać potraktowany tak, jak zbiór zadań jednoprocesorowych, do realizacji których wymagana jest jednostka obliczeniowa specjalnego typu, tzn. taka, która stanowi klaster zbudowany z odpowiedniej liczby procesorów. Jednak z punktu widzenia programu szeregującego superzadania wewnętrzna budowa jednostki obliczeniowej nie jest istotna, a szeregowane superzadania można potraktować w taki sam sposób, w jaki traktuje się zadania jednoprocesorowe, czyli można już bezpośrednio zastosować algorytm RMS. Zaprezentowana w artykule propozycja modyfikacji algorytmu RMS została zilustrowana na wybranym przykładzie szeregowania zadań dla systemu wieloprocesorowego o topologii hipersześcianu czterowymiarowego.

Słowa kluczowe: szeregowanie zadań, Rate Monotonic Scheduling, systemy wieloprocesorowe, hipersześcian

1. WPROWADZENIE

Podstawową różnicą, jaka występuje pomiędzy systemami komputerowymi ogólnego przeznaczenia a systemami czasu rzeczywistego o ostrych ograniczeniach czasowych (ang. hard real-time), jest, w przypadku tych drugich, fakt posiadania przez każde z wykonywanych przez jednostkę obliczeniową zadań ograniczenia czasowego, którego naruszenie jest niedopuszczalne [1]. W systemie czasu rzeczywistego o ostrych ograniczeniach czasowych przekroczenie przez dowolne z zadań jego ograniczenia czasowego prowadzić może do całkowitej utraty kontroli nad sterowanym obiektem, a tym samym do powstania katastrofy, wielkich strat finansowych, utraty życia ludzkiego itp.

Z tego powodu systemy czasu rzeczywistego o ostrych ograniczeniach czasowych muszą być niezwykle starannie projektowane i nie można sobie w ich przypadku pozwolić na żadne eksperymenty, ponieważ metoda prób i błędów może w tym przypadku po prostu zbyt wiele kosztować [2]. Takie postawienie sprawy wymusiło intensywny rozwój metod formalnych służących do specyfikacji i modelowania systemów czasu rzeczywistego o ostrych ograniczeniach czasowych. Bowiem w przypadku tych systemów już na etapie ich projektowania należy zagwarantować, że każde z zadań, w wszelkich możliwych warunkach pracy systemu, zawsze dochowa swoje ograniczenie czasowe. Takich gwarancji może udzielić jedynie teoria szeregowania zadań [3]. W ramach tej teorii opracowywane są różnego typu algorytmy szeregujące zadania, które precyzyjnie określają momenty rozpoczęcia i zakończenia wykonywania poszczególnych zadań, a ponadto, w przypadku spełnienia pewnych warunków, dają gwarancję nienaruszalności ograniczeń czasowych szeregowanych zadań [4].

Jednym z najbardziej popularnych algorytmów szeregowania zadań w systemach czasu rzeczywistego o ostrych ograniczeniach czasowych jest algorytm RMS (ang. Rate Monotonic Scheduling), który bazuje na systemie przydziału priorytetów do zadań i w przypadku spełnienia odpowiedniego warunku odnośnie stopnia wykorzystania jednostki obliczeniowej przez szeregowany zbiór zadań, algorytm ten jest w stanie zagwarantować dotrzymanie ograniczenia czasowego przez każde z zadań [5].

2. PODSTAWOWE WŁAŚCIWOŚCI ALGORYTMU RMS

Algorytm RMS przeznaczony jest do szeregowania zbioru zadań okresowych, tzn. takich, których poszczególne instancje są aktywowane w regularnych odstępach czasu, przy czym ograniczenie czasowe zadania równe jest jego okresowi, czyli poprzednia instancja zadania musi zostać zawsze zakończona przed aktywowaniem kolejnej [6]. Jeśli w szeregowanym zbiorze zadań występują również pewne zadania aperiodyczne, można je sprowadzić do zadań okresowych stosując technikę tzw. periodycznego

serwera zdań sporadycznych [3]. Ponadto szeregowane zadania muszą posiadać cechę wywłaszczalności, polegającą na tym, że wykonywanie dowolnego zadania może zostać w każdej chwili przerwane (kontekst zadania jest oczywiście zawsze zapisywany na stosie procesora), a jednostka obliczeniowa może zostać oddana do dyspozycji innego zadania, po którego zakończeniu przerwane zadanie może zostać wznowione. Ponadto szeregowane zadania powinny być zadaniami niezależnymi, czyli wyniki obliczeń uzyskane w wyniku realizacji dowolnego z zadań nie powinny stanowić danych wejściowych dla żadnego innego zadania. W konsekwencji tego kolejność wykonywania poszczególnych zadań jest dowolna.

Algorytm RMS bazuje na systemie priorytetów [12]. Zasady przydziału priorytetów do zadań są następujące. Im mniejsza wartość okresu zadania, tzn. im zadanie jest aktywowane częściej, tym wyższy jest jego priorytet. W danym momencie spośród wszystkich zadań będących w stanie gotowości wykonywane jest to, które posiada najwyższy priorytet. Jeśli w stan gotowości wejdzie jakieś inne zadanie o wyższym priorytecie, wówczas aktualnie wykonywane zadanie zostanie wywłaszczone, a procesor zostanie przydzielony do nowoprzybyłego zadania o wyższej wartości priorytetu. Realizacja wywłaszczonego zadania może zostać wznowiona tylko w sytuacji, w której w stanie gotowości nie znajduje się już żadne inne zadanie o priorytecie wyższym od niego [7].

Każde z szeregowanych zadań z_i charakteryzowane jest przez wartość swojego okresu aktywacji T_i oraz czas wykonania C_i. Wartość stosunku C_i/T_i określa, w jakim stopniu zadanie z_i wykorzystuje czas pracy procesora [4]. W 1973 roku Liu i Layland udowodnili bardzo ważne twierdzenie, które mówi, że jeżeli spełniona jest nierówność $\sum_{i=1}^{N} \frac{C_i}{T_i} \leqslant N\left(2^{\frac{1}{N}} - 1\right)$, wówczas dany zbiór N niezależnych, okresowych i wywłaszczalnych zadań jest na pewno szeregowalny, tzn. nigdy nie zdarzy się tak, aby któreś z zadań naruszyło swoje ograniczenie czasowe [3].

Prawa strona rozważanej nierówności zmierza do wartości równej ln2, czyli około 0,69. Oznacza to, że jeżeli szeregowany zbiór zadań wykorzystuje czas pracy procesora w nie więcej niż 69%, to wówczas zbiór ten jest na pewno szeregowalny. W sytuacji, gdy szeregowany zbiór zadań wykorzystuje czas pracy procesora w więcej niż 69% (ale mniej niż 100%), dany zbiór zadań może już nie być szeregowalny. Aby się o tym przekonać, należy rozpatrzyć najgorszy z możliwych przypadków, tzn. taki, w którym wszystkie zadania zostają aktywowane równocześnie — wówczas zadanie o najniższym priorytecie zostanie wykonane jako ostanie. Dla rozpatrywanego przypadku należy dla każdego z zadań policzyć czas zakończenia jego realizacji. Jeżeli okaże się, że wszystkie zadania zostaną zakończone przed upływem swoich ograniczeń czasowych, to dany zbiór zadań jest szeregowalny. W przypadku przeciwnym rozważany zbiór zadań szeregowalny nie jest.

Czas zakończenia realizacji zadania, w sytuacji, gdy równocześnie z nim zostało aktywowanych N-1 zadań o wyższej wartości priorytetów, może zostać policzony za pomocą następującej procedury iteracyjnej. Jako pierwsze przybliżenie czasu zakoń-

czenia zadania z_N wyznacza się sumę czasów realizacji tego zadania i wszystkich zadań o wyższych priorytetach, ponieważ zanim rozpocznie się wykonywanie zadania z_N każde z pozostałych zadań musi zostać wykonane przynajmniej jeden raz. Zatem $t_0 = \sum_{i=1}^{N} C_i$. Kolejne przybliżenia czasu zakończenia realizacji zadania z_N wyznaczane są ze wzoru $t_{k+1} = \sum_{i=1}^{N} C_i \cdot \left\lceil \frac{t_k}{T_i} \right\rceil$, do momentu, aż spełniona jest równość $t_k = t_{k+1}$. Uzyskana wartość czasu t_k jest czasem zakończenia realizacji zadania z_N.

3. ZASTOSOWANIE ALGORYTMU RMS W SYSTEMACH WIELOPROCESOROWYCH

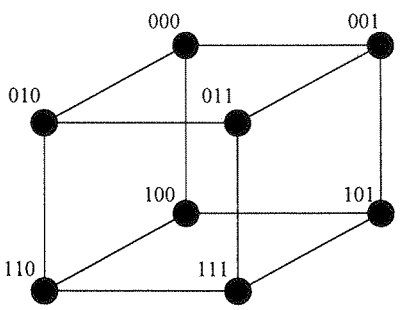
W swej klasycznej postaci algorytm RMS przewidziany jest jedynie do szeregowania zadań przeznaczonych do realizacji na pojedynczym procesorze. Jednak obecnie wciąż wzrastające zapotrzebowanie na moc obliczeniową wymusza coraz powszechniejsze stosowanie architektur wieloprocesorowych typu MIMD (ang. Multiple Instruction Multiple Data) [13]. Architektury takie bazują na przesyłaniu komunikatów pomiędzy poszczególnymi jednostkami obliczeniowymi realizującymi dany algorytm równoległy [8], [9].

W idealnym przypadku każda jednostka obliczeniowa powinna być połączona z każdą inną dedykowanym łączem komunikacyjnym, umożliwiającym bezpośrednią wymianę danych. Jednak w takim przypadku dla N jednostek obliczeniowych należałoby utworzyć aż $N(N-1)$ łączy komunikacyjnych, co dla dużych wartości N nie jest realizowalne w praktyce. Zatem w systemach wieloprocesorowych należy pogodzić się z sytuacją, w której nie wszystkie jednostki obliczeniowe mają bezpośrednie połączenia komunikacyjne [14]. Przesłanie komunikatu pomiędzy takimi jednostkami wymusza konieczność skorzystania z pewnej liczby procesorów pośredniczących, co dodatkowo wydłuża czas wędrówki komunikatu [15]. Konieczny jest zatem pewien rozsądny kompromis pomiędzy liczbą koniecznych do utworzenia połączeń komunikacyjnych, a maksymalna długością drogi wędrówki komunikatu w systemie wieloprocesorowym. Taki kompromis, charakteryzujący się korzystną wartością wymienionych parametrów, stanowi architektura równoległa o topologii hipersześcianu.

Hipersześcian, zwany również hiperkostką (ang. hypercube), jest to sześcian występujący w przestrzeni posiadającej D wymiarów. W przypadku przestrzeni trójwymiarowej (D = 3) hipersześcian jest zwykłym sześcianem, w którego narożach umieszczone są poszczególne jednostki obliczeniowe, natomiast krawędzie tego sześcianu odpowiadają łączom komunikacyjnym, jakie istnieją pomiędzy wybranymi jednostkami obliczeniowymi. Poszczególnym jednostkom obliczeniowym przypisane zostają numery binarne, w ten sposób, że bezpośrednie połączenia komunikacyjne istnieją tylko pomiędzy tymi jednostkami obliczeniowymi, których numery binarne różnią się

dokładnie wartością jednego bitu. Na rys. 1, na którym przedstawiono hipersześcian trójwymiarowy, można zobaczyć, że jednostka obliczeniowa o binarnym numerze 000 posiada bezpośrednie połączenia komunikacyjne tylko i wyłącznie z trzema innymi jednostkami obliczeniowymi o numerach 001, 010 i 100. W takim wypadku droga przesyłu komunikatu wynosi 1 jednostkę. Z kolei przesłanie komunikatu do dowolnej innej jednostki obliczeniowej wymaga skorzystania z pomocy co najmniej jednej jednostki obliczeniowej pośredniczącej w transferze komunikatu. Zatem, aby można było przesłać komunikat z jednostki obliczeniowej o numerze 000 do jednostki o numerze 111, należy skorzystać z pośrednictwa dwóch innych jednostek, np. 001 i 011. W tym przypadku droga wędrówki komunikatu wynosi 3 jednostki.

W ogólnym przypadku hipersześcianu o D wymiarach każda z jednostek obliczeniowych posiada dokładnie D sąsiadów, z którymi jest bezpośrednio połączona dedykowanymi łączami komunikacyjnymi [7]. Również maksymalna droga wędrówki komunikatu wynosi D jednostek. Hipersześcian D-wymiarowy zbudowany jest z 2^D jednostek obliczeniowych, przy czym bezpośrednie połączenia komunikacyjne istnieją tylko pomiędzy tymi jednostkami obliczeniowymi, których numery binarne różnią się pomiędzy sobą wartością dokładnie jednego bitu.



Rys. 1. Hipersześcian trójwymiarowy

Fig. 1. Three-dimensional hypercube

Każdy hipersześcian D-wymiarowy może zostać potraktowany jako złożenie dwóch hipersześcianów o D-1 wymiarach, z których każdy jest zbudowany z 2^{D-1} jednostek obliczeniowych.

Istota zaproponowanej przez autora adaptacji algorytmu RMS dla potrzeb szeregowania zbioru periodycznych, niezależnych i wywłaszczalnych zadań jedno- i wieloprocesorowych, realizowanych przez system równoległy o architekturze hipersześcianu, sprowadza się do transformacji wartości okresów wybranych zadań i utworzeniu z nich tzw. superzadania (ang. supertask), do realizacji którego potrzebna jest jednoczesna dostępność wszystkich jednostek obliczeniowych.

Rozważana transformacja okresów zadań polega na zmniejszeniu ich wartości. Zabieg taki jest często stosowany w systemach czasu rzeczywistego i jest warunkiem koniecznym do przeprowadzenia tzw. binaryzacji okresów zadań, podczas której zadania uzyskują nowe, mniejsze wartości okresów, spełniające warunek $T_i = r \cdot 2^i$. Dzięki temu szeregowaniu podlegają zadania o okresach będących wielokrotnością pewnego okresu podstawowego r, przy czym każda następna wartość okresu jest dwukrotnie większa od poprzedniej [10]. Dzięki binaryzacji zmniejszeniu ulega tzw. horyzont czasowy systemu, czyli przedział czasu, po upływie którego cały scenariusz realizacji wszystkich zadań powtarza się na nowo. Jest to zatem swego rodzaju superokres, po którym stan systemu powraca do punktu początkowego.

Niestety ujemną stroną binaryzacji okresów zadań jest fakt, iż powoduje ona zwiększenie obciążenia jednostki obliczeniowej przez szeregowany zbiór zadań, ponieważ zmniejszając wartość okresu zadania T_i, jednoczesnemu zwiększeniu ulega stosunek C_i/T_i. Natomiast dodatkową zaletą binaryzacji okresów zadań jest powiększenie się marginesu stabilności sterowanego obiektu, ponieważ dane zbierane są ze sterowanego obiektu z większą częstotliwością, w związku z czym trajektoria sterowanego obiektu jest korygowana częściej [11].

W systemach komputerowych o architekturze hipersześcianu D-wymiarowego występują zadania przewidziane do realizacji na wszystkich 2^D procesorach, jak również zadania, do których realizacji potrzebne są jedynie procesory wchodzące w skład hipersześcianów o niższych wymiarach. Sposób transformacji okresów wybranych zadań zostanie zilustrowany na przykładzie szeregowania zadań w systemie równoległym o architekturze hipersześcianu czterowymiarowego.

4. PRZYKŁAD SZEREGOWANIA ZBIORU ZADAŃ

Niech dany będzie następujący zbiór zadań niezależnych, wywłaszczalnych i okresowych, których charakterystyki zostały zebrane w tab. 1. Rozważane zadania przeznaczone są do realizacji w systemie wieloprocesorowym o architekturze hipersześcianu czterowymiarowego. W kolejnych kolumnach tab. 1 podano odpowiednio numer zadania, czas jego realizacji, wartość okresu zadania oraz liczbę procesorów, których jednoczesna dostępność jest wymagana do realizacji danego zadania.

Zamieszczone w tab. 1 zadania zostały zgrupowane w pięć większych superzadań w następujący sposób.

Pierwsze superzadanie Z_1 powstaje poprzez zgrupowanie zadań t_1, t_2, t_3 i t_4 w jedno większe zadanie, do którego realizacji wymagana jest dostępność wszystkich szesnastu procesorów. Okresy zadań t_2, t_3 i t_4 ulegają transformacji i otrzymują wartość taką, jak w przypadku zadania t_1, tzn. 245 ms. W ten sposób wszystkie zadania t_1, t_2, t_3 i t_4 otrzymują identyczną wartość okresu, dzięki czemu można je połączyć w jedno większe superzadanie. Zadanie t_1 zostaje przydzielone do hipersześcianu trójwymiarowego, a zatem w jego realizacji uczestniczą procesory o numerach binarnych: 0000, 0001, 0010, 0011, 0100, 0101, 0110 i 0111. Z kolei zadanie t_2 zostaje przydzielone do

procesorów o numerach binarnych: 1000, 1001, 1010 i 1011, tworzących hipersześcian dwówymiarowy. Analogicznie zadanie t_3 zostaje przydzielone do procesorów o numerach binarnych 1100 i 1101, a zadanie t_4 do procesorów 1110 i 1111. Czas realizacji superzadania Z_1 równy jest czasowi zakończenia tego zadania wchodzącego w jego skład, które ma najdłuższy czas realizacji. W rozważanym przypadku zadaniem takim jest zadanie t_4, a zatem czas realizacji superzadania Z_1 wynosi 14 ms.

Tabela 1

Charakterystyki zadań podlegających szeregowaniu algorytmem RMS

Characteristics of the tasks scheduled by the medium of RMS algorithm

Zadanie	Czas realizacji zadania [ms]	Okres zadania [ms]	Liczba procesorów wymaganych do realizacji zadania
t_1	12	245	8
t_2	13	257	4
t_3	9	289	2
t_4	14	307	2
t_5	25	325	16
t_6	29	336	8
t_7	13	365	4
t_8	18	374	4
t_9	20	387	2
t_{10}	16	399	2
t_{11}	12	469	1
t_{12}	16	478	1
t_{13}	28	495	16
t_{14}	19	520	8
t_{15}	15	536	2
t_{16}	25	549	2
t_{17}	27	585	4
t_{18}	17	604	1
t_{19}	24	648	1
t_{20}	26	672	1

Kolejne superzadanie Z_2 utworzone zostaje tylko z zadania t_5, które jest zadaniem szesnastoprocesorowym, wymagającym do swej realizacji jednoczesnej dostępności wszystkich procesorów systemu równoległego. Z tego powodu zadanie t_5 nie jest łą-

czone z żadnym innym zadanie. W rozważanym wypadku również nie ulega zmianie wartość okresu tego zadania, która w dalszym ciągu wynosi 325 ms.

Z kolei superzadanie Z_3 zostaje utworzone z siedmiu zadań t_6, t_7, t_8, t_9, t_{10}, t_{11} i t_{12}, przy czym okresy zadań t_7, t_8, t_9, t_{10}, t_{11} i t_{12} ulegają transformacji i uzyskują wartość okresu zadania t_6, czyli 336 ms. Zadanie t_6 zostaje przydzielone do procesorów o numerach binarnych 0000, 0001, 0010, 0011, 0100, 0101, 0110 i 0111. Z kolei zadanie t_7 zostaje przydzielone do procesorów o numerach binarnych: 1000, 1001, 1010 i 1011. Natomiast zadanie t_8 zostaje przydzielone do procesorów o numerach binarnych: 1100, 1101, 1110 i 1111. Z kolei zadanie dwuprocesorowe t_9 zostaje przydzielone do procesorów 1000 i 1001. Analogicznie zadanie t_{10} zostaje przydzielone do procesorów 1010 i 1011. Zadania jednoprocesorowe t_{11} i t_{12} zostają przydzielone do procesorów o numerach binarnych 1100 i 1101. Czas realizacji superzadania Z_3 może zostać wyznaczony następująco. Najpierw na procesorach 1100, 1101, 1110 i 1111 wykonywane jest zadanie t_8, którego realizacja trwa 18 ms. Dopiero po zakończeniu realizacji zadania t_8 zwolniony zostaje procesor 1101, na którym realizację rozpoczyna zadanie t_{12}, co zajmuje kolejne 16 ms. Analizując czasy wykonywania poszczególnych zadań wchodzących w skład superzadania Z_3 można się łatwo przekonać, że zadanie t_{12} zostanie zakończone jako ostatnie. Zatem czas realizacji superzadania Z_3 równy jest czasowi zakończenia wykonywania zadania t_{12} i wynosi 34 ms.

W skład superzadania Z_4 wchodzi tylko jedno zadanie t_{13}, ponieważ jest to zadanie, do realizacji którego wymagana jest jednoczesna dostępność wszystkich 16 procesorów. Czas realizacji superzadania Z_4 wynosi 28 ms, a jego okres 495 ms.

Ostatnie z superzadań Z_5 zostaje utworzone z zadań t_{14}, t_{15}, t_{16}, t_{17}, t_{18}, t_{19} i t_{20}. Okres superzadania Z_5 wynosi 520 ms. Zadanie t_{14} zostaje przydzielone do procesorów 0000, 0001, 0010, 0011, 0100, 0101, 0110 i 0111. Z kolei zadanie t_{15} zostaje przydzielone do procesorów 1000 i 1001, zadanie t_{16} do procesorów 1010 i 1011, a zadanie t_{17} do procesorów 1100, 1101, 1110 i 1111. Zadania jednoprocesorowe zostają przydzielone do poszczególnych procesorów w następujący sposób. Zadanie t_{18} do procesora 1010, zadanie t_{19} do procesora 1000, a zadanie t_{20} do procesora 1001. Czas realizacji superzadania Z_5 wynosi 42 ms. Sposób przydziału zadań wchodzących w skład superzadania Z_5 do poszczególnych procesorów pokazano na rys. 2. Natomiast w tab. 2 zebrano charakterystyki wszystkich superzadań podlegających szeregowaniu.

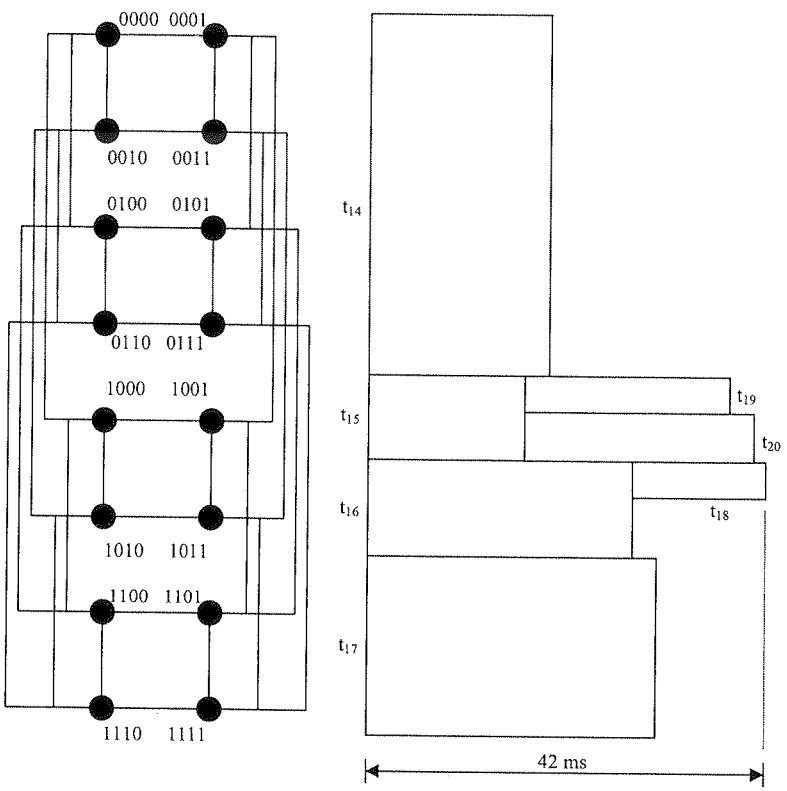
Sumaryczne obciążenie jednostki obliczeniowej przez szeregowany zbiór superzadań wynosi 0,373, a zatem zbiór ten jest szeregowalny na mocy twierdzenia udowodnionego przez Liu i Laylanda [3].

Tabela 2

Charakterystyki superzadań podlegających szeregowaniu algorytmem RMS

Characteristics of the supertasks scheduled by the medium of RMS algorithm

Superzadanie	Czas realizacji superzadania [ms]	Okres superzadania [ms]	Wykorzystanie jednostki obliczeniowej
Z_1	14	245	0,057
Z_2	25	325	0,077
Z_3	34	336	0,101
Z_4	28	495	0,057
Z_5	42	520	0,081



Rys. 2. Grupowanie zadań w jedno superzadanie Z_5

Fig. 2. Grouping of tasks into one supertask Z_5

5. ZAKOŃCZENIE

W artykule zamieszczono propozycję rozszerzenia obszaru stosowalności algorytmu szeregującego Rate Monotonic Scheduling dla przypadku szeregowania zadań jedno- i wieloprocesorowych w równoległych architekturach o topologii hipersześcianu. Szeregowany zbiór zadań musi posiadać kilka istotnych właściwości. Mianowicie, zadania podlegające szeregowaniu muszą być okresowe, niezależne i wywłaszczalne.

Istota zaproponowanej adaptacji algorytmu RMS polega na dokonaniu transformacji wartości okresów wybranych zadań, w taki sposób, aby pewna liczba zadań uzyskała taką samą wartość okresu aktywacji [13]. Wówczas rozważane zadania można połączyć w jedno większe tzw. superzadanie, które do swego wykonania potrzebuje jednoczesnej dostępności wszystkich jednostek obliczeniowych występujących w systemie równoległym. W przypadku zgrupowania wszystkich zadań w pewną liczbę superzadań, dalszemu szeregowaniu podlegają tylko superzadania. Ponieważ każde z superzadań wymaga do swej realizacji całego systemu równoległego, system ten może zostać potraktowany jako jedna wielka jednostka obliczeniowa, w przypadku której realizowane zadania mogą być już bezpośrednio szeregowane algorytmem RMS. Bowiem z punktu widzenia algorytmu RMS wewnętrzna budowa jednostki obliczeniowej, tzn. fakt, że składa się ona z pewnej liczby mniejszych podjednostek, jest nieistotny, gdyż szeregowaniu podlegają tylko superzadania, które są traktowane jak zadania jednoprocesorowe, z tym tylko zastrzeżeniem, że do ich realizacji wymagana jest jednostka obliczeniowa specjalnego typu.

Pewnego komentarza wymaga jeszcze sposób wyboru zadań, których wartości okresów podlegają transformacji, w taki sposób, aby wiele różnych zadań uzyskało taką samą wartość okresu aktywacji i następnie można było utworzyć z nich jedno superzadanie. Niestety, w rozważanej sytuacji istnieje kombinatoryczna liczba możliwych sposobów, na jakie można dokonać transformacji wartości okresów poszczególnych zadań oraz sposobów utworzenia różnych superzadań. W przypadku, gdy szeregowaniu podlega niewielka liczba zadań można dokonać, metodą przeglądu zupełnego, optymalnego sposobu transformacji wartości okresów zadań. Jednak w przypadku większej liczby zadań koszt obliczeniowy przeglądu zupełnego może okazać się zbyt wielki, aby można było zastosować tę metodę w praktyce. W takim przypadku pozostaje jedynie zastosowanie pewnej metody heurystycznej, np. algorytmu genetycznego, który metodą poszukiwań losowych starał się będzie znaleźć jak najlepsze rozwiązanie, tzn. takie, w przypadku którego zbiór superzadań jest szeregowalny oraz jednostka obliczeniowa wykonująca superzadania jest obciążona w jak najmniejszym stopniu.

6. BIBLIOGRAFIA

1. T. Szmuc: *Modele i metody inżynierii oprogramowania systemów czasu rzeczywistego*, Uczelniane Wydawnictwa Naukowo-Dydaktyczne AGH, Kraków, 2001.
2. K.G. Shin, P. Ramanathan: *Real-time computing: A new discipline of computer science and engineering*. Proceedings of the IEEE, vol. 82, no. 1, January 1994, pp. 6-24.

3. L. Liu, J.W. Layland: *Scheduling algorithms for multiprogramming in a hard real time environment*. Journal on Assoc. Comput. Mach., vol. 20, no. 1, 1973, pp. 46-61.
4. T. Stoyenko, P. Baker: *Real-time schedulability-analyzable mechanizms in Ada9X*. Proceedings of the IEEE, vol. 82, no. 1, January 1994, pp. 95-107.
5. K. Ramanathan, J.A. Stankovic: *Scheduling algorithms and operating systems support for real-time systems*. Proceedings of the IEEE, vol. 82, no. 1, January 1994, pp. 55-67.
6. A. Zalewski: *What every engineer needs to know about rate monotonic scheduling: A tutorial*. Odessa, TX, 1995, pp. 321-335.
7. A.S. Tanenbaum: *Rozproszone systemy operacyjne*, PWN, Warszawa, 1997.
8. J. Varriet: *Scheduling inverted-ordered task with non-uniform deadlines subject tonon-zerocommunication delay*, Parallel Computing 25 (1999), pp. 3-21.
9. O. Kwon, K.Y. Chwa: *Approximation algorithms for general parallel task scheduling*, Information Processing Letters 81 (2002), pp. 143-150.
10. A. Czajka, J. Nawrocki: *Szeregowanie zadań o okresach binarnych w systemach silnie uwarunkowanych czasowo*, I Krajowa Konferencja Metody i systemy komputerowe w badaniach naukowych i projektowaniu inżynierskim, Kraków, 1997, ss. 669-676.
11. J. Nawrocki, A. Czajka: *Binaryzacja okresów zadań cyklicznych*, VII Konferencja Systemy Czasu Rzeczywistego, Kraków, 2000, ss. 41-51.
12. J. Werewka: *Zagadnienia alokacji zadań w rozproszonych systemach czasu rzeczywistego*, Elektrotechnika, Tom 14, Zeszyt 4, 1995, ss. 479-488.
13. C. Hsueh, K.J. Lin: *Scheduling real-time systems with end-to-end timing constraints using the distributed pinwheel model*, IEEE Transaction on Computers, vol. 50, no. 1, 2001, pp. 51-68.
14. O.H. Kwon, K.Y. Chwa: *Approximation algorithms for general parallel task scheduling*, Information Processing Letters 81, 2002, pp. 143-150.
15. D. Ye, G. Zhang: *On-line scheduling with extendable working time on a small number of machines*, Information Processing Letters 85, 2003, pp. 171-177.
16. J.H. Kim, K.Y. Chwa: *Online deadline scheduling on faster machines*, Information Processing Letters 85, 2003, pp. 31-37.

M. GAJER

THE IMPLEMENTATION OF RATE MONOTONIC SCHEDULING IN MULTIPROCESSOR ARCHITECTURES OF TOPOLOGY OF HYPERCUBE

Summary

The real-time systems are getting more and more popular. In fact most of contemporary industrail and communication systems could not do without them. The popularity of real-time systems with hard real-time constraints forced the extensive development of task scheduling theory. In the case of real-time systems with hard real-time constraints it does not suffice that the task produces logically correct results but these results must be delivered within their time constraints.

In such systems even logically correct results that are delivered with the violation of their time constraints are totally useless. Moreover, the consequences of violation of time constraints can very often be quite severe and can cause the great economic losses and even losses of human lives, e.g. in the case of control systems of nuclear reactors, space ships etc.

The main goal of the task scheduling theory is to prove at the stage of the system project that the time constraints for all tasks will always be met under any possible circumstances.

In the case of the real-time systems with hard real-time constraints there is very often a necessity of scheduling a set of independent, pre-emptive and periodic tasks. The most popular algorithm for scheduling such set of independent, pre-emptive and periodic tasks is the Rate Monotonic Scheduling algorithm.

In the case of Rate Monotonic Scheduling each task is assigned a priority. There are several rules basing on which the priorities are assigned to the tasks and then the tasks are being scheduled. First of all, the shorter the period of task is the higher priority it is assigned. Then, in a given moment, among all the tasks actually in a ready state the one is being executed that has the highest priority. If some task with higher priority enters into the ready state the task being executed is automatically pre-empted and the task with higher priority begins its execution. The pre-empted task can restart its execution only in the case if there is actually no other task with higher priority in the ready state.

The Rate Monotonic Scheduling is adequate for scheduling uniprocessor tasks. This author has proposed a new method of adaptation of Rate Monotonic Scheduling theory also for the purpose of scheduling multiprocessor tasks in the multiprocessor architectures of the topology of hypercube.

The clue of the method proposed by this author is concatenation of many uniprocessor and multiprocessor tasks that should form one so called supertask. In order to achieve this the periods of some tasks must be transformed, i.e. they must be shortened in such a way that several subsets of tasks of the same value of period should be made. Then each subset of tasks is treated as a uniprocesor supertask and for the set of such supertasks the Rate Monotonic Scheduling algorithm can be used directly.

The method developed by this author was illustrated on the example of scheduling set of tasks for the multiprocessor system of the topology of hypercube of dimension 4. The method can be easily extended for the higher dimensions of hypercube multiprocessor architectures.

Keywords: Task Scheduling Theory, Rate Monotonic Scheduling, Multiprocessor Systems, Hypercube Architecture

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2006, 52, z. 2, ss. 153–183

Wejściowe pobudzenie optyczne a parametry transmisyjne światłowodów wielomodowych

JERZY SIUZDAK, ROMAN NOWAK, GRZEGORZ STĘPNIAK

Instytut Telekomunikacji Politechniki Warszawskiej
ul. Nowowiejska 15/19, 00-665 Warszawa
siuzdak@tele.pw.edu.pl
nowak@tele.pw.edu.pl

Otrzymano 2005.04.06
Autoryzowano 2005.06.27

W pracy zaprezentowano model propagacji światła w światłowodach wielomodowych. Na podstawie tego modelu określono zależności teoretyczne podstawowych parametrów transmisyjnych światłowodu (pasmo i tłumienność) od sposobu jego pobudzenia. Zależności te zostały zweryfikowane doświadczalnie przy pomocy opracowanego w tym celu układu pomiarowego składającego się z przestrajanego generatora, analizatora widma oraz konwerterów E/O i O/E. Badania eksperymentalne zostały przeprowadzone zarówno dla światłowodów kwarcowo-polimerowych, jak również dla światłowodów kwarcowych. W pracy wykazano, że wpływ typu pobudzenia światłowodu na jego parametry transmisyjne jest tym istotniejszy im mniejszy jest współczynnik sprzężenia modów. Wskazano również na możliwość selektywnego pobudzania na wejściu różnych grup modów za pomocą przesunięć offsetowych.

Słowa kluczowe: transmisja danych, sieci lokalne, światłowody wielomodowe, pasmo, tłumienność, dyspersja modowa, laser półprzewodnikowy, LED

1. WSTĘP

Światłowody wielomodowe, zarówno ze szkła kwarcowego, jak i polimerowe (np. PMMA, CYTOP) znajdują obecnie zastosowanie do transmisji danych o stosunkowo dużych szybkościach (do 10 Gbit/s włącznie) na niewielkie odległości (typowo do kilkuset metrów) w sieciach lokalnych, takich jak np. Fibre Channel, Gigabit Ethernet (IEEE 802.3z), 10 Gbit Ethernet i in. Istotnym, niedostatecznie do tej pory rozwiązanym problemem badawczym w projektowaniu takich linii jest określenie ich podstawowych parametrów, a mianowicie przede wszystkim pasma wyznaczonego przez

dyspersję modową, jak również (w mniejszym stopniu) tłumienności. Ze względu na szybkość działania, linie te muszą współpracować z nadajnikami laserowymi, takimi jak np. VCSEL, a nie z diodami elektroluminescencyjnymi. Stwarza to istotne trudności z określeniem parametrów linii, gdyż np. tłumienność linii mierzona zgodnie z dotychczasową metodyką pomiaru (np. TIA/EIA 526-14-A [19]) przy użyciu diod LED różni się od tłumienności mierzonej z laserem jako nadajnikiem.

Najogólniej rzecz biorąc problemy pomiarowe są związane z tym, że przy sprzężeniu światłowodu wielomodowego ze źródłem światła mody różnych rzędów są pobudzane w różnym stopniu (dioda LED pobudza więcej modów niż laser). W trakcie propagacji mody wymieniają się swymi energiami aż do osiągnięcia stanu równowagi (tzw. równowaga modowa), w której wartości oczekiwane mocy transmitowanych przez poszczególne mody nie ulegają dalszym zmianom. Parametry światłowodu będącego w stanie równowagi modowej można stosunkowo łatwo określić. Niestety odcinek światłowodu, na którym stan równowagi modowej nie został jeszcze osiągnięty (tzw. odcinek niestabilności modowej) jest dosyć znaczny (kilkaset metrów i więcej) i może obejmować całą linię, co bardzo utrudnia obliczenie jej parametrów. Dotyczy to zarówno wspomnianej już tłumienności, jak i przede wszystkim wypadkowego pasma modowego.

Czynniki określające warunki pobudzenia światłowodu i mieszania się modów, a co za tym idzie również wyznaczające podstawowe parametry linii wykorzystującej światłowód wielomodowy, to przede wszystkim profil współczynnika załamania oraz jego niedokładności, typ źródła światła, jego charakterystyka kierunkowa i warunki sprzężenia ze światłowodem. Pewne znaczenie ma również rodzaj i ilość wtrąceń w światłowodzie oraz rodzaj pokrycia pierwotnego. W mniejszym stopniu na te parametry mają wpływ warunki ułożenia światłowodu, jego naprężenia, promienie zgięcia i falowanie.

Problemem omówionym w pracy jest wpływ warunków pobudzenia światłowodu wielomodowego oraz mieszania modów na jego podstawowe parametry transmisyjne tj. pasmo i tłumienność.

2. TEORIA

W tzw. przybliżeniu słabego prowadzenia rozwiązaniem równania falowego w światłowodzie są mody liniowo spolaryzowane, oznaczane przez LP_{lp}, gdzie liczby l i p określają rząd (rodzaj) modu. Przy analizie światłowodów wielomodowych stosuje się opis za pomocą tzw. modów złożonych (ang. compound mode; inna nazwa: grupy modowe). Mianowicie każdy z modów LP_{lp} należy do grupy modów o indeksie m, wyrażonym zależnością [1, 7, 30]

$$m = 2p + l - 1, \quad m = 1, 2, 3 \ldots m_c \tag{1}$$

Indeks m_c odpowiada najwyższej grupie modów prowadzonej jeszcze poniżej odcięcia.

W przypadku regularnego profilu współczynnika załamania światłowodu mody należące do danej grupy modowej charakteryzują się jednakowymi lub bardzo zbliżonymi stałymi propagacji γ_m (podobną tłumiennością i opóźnością jednostkową) [7]. Liczba modów w danej grupie modowej jest w przybliżeniu proporcjonalna do jej indeksu m, ponadto możliwe są dwie polaryzacje ortogonalne.

Przy analizie sprzężenia między modami operuje się zazwyczaj wielkością m znormalizowaną do m_c i oznaczaną przez x:

$$x = \frac{m}{m_c}, 0 < x < 1, \tag{2}$$

która traktowana jest jako zmienna ciągła (tzw. założenie continuum) [7, 8]. Oznaczając przez P = P(x,z,t) średnią moc modu w grupie modowej określonej przez x, w odległości z od początku światłowodu (od źródła światła) i dla czasu t, można napisać równanie określające zależność mocy w grupach modowych w funkcji czasu t i odległości z (tzw. równanie dyfuzji) [7]:

$$\frac{\partial P}{\partial z} + \tau(x)\frac{\partial P}{\partial t} + \alpha(x)P = \frac{1}{M}\frac{1}{x}\frac{\partial}{\partial x}\left[xd(x)\frac{\partial P}{\partial x}\right] \tag{3}$$

Tutaj M jest liczbą wszystkich modów prowadzonych. Ponadto w powyższym równaniu wprowadzono następujące oznaczenia:

τ(x), [s/km], jest jednostkową opóźnością grupową modu w grupie modowej określonej przez x,

α(x), [Np/km], jest tłumiennością jednostkową modu w grupie modowej określonej przez x,

d(x), [1/km], jest współczynnikiem sprzężenia modów w grupie modowej określonej przez x i charakteryzuje on szybkość przechodzenia energii między modami w sąsiednich grupach modów. Omówimy teraz kolejno powyższe parametry.

Dla regularnego profilu współczynnika załamania **jednostkowa opóźność grupowa** wyraża się przybliżoną zależnością [1, 6]:

$$\tau(x) = \frac{N_1}{c}\left(1 + \frac{g-2-\varepsilon}{g+2}\cdot\Delta\cdot x^{\frac{2g}{g+2}} + \frac{1}{2}\cdot\frac{3g-2-2\varepsilon}{g+2}\cdot\Delta^2\cdot x^{\frac{4g}{g+2}}\right) \tag{4}$$

Tutaj c jest prędkością światła, g jest parametrem określającym kształt profilu rdzenia (g = 2 dla profilu parabolicznego), N_1 jest (grupowym) współczynnikiem załamania (w centrum) rdzenia, Δ jest względną różnica współczynników załamania rdzenia i płaszcza, zaś ε jest (bezwymiarowym) parametrem związanym z tzw. dyspersją profilową [1, 2].

Tłumienność jednostkowa, α(x), jest funkcją grupy modowej opisanej przez x i określa jej tłumienie na jednostkę długości. Na wartość tłumienności jednostkowej

mają wpływ zjawiska wspólne dla wszystkich modów (rozpraszanie Rayleigha, absorpcja), efekty związane z nieregularnością styku miedzy rdzeniem i płaszczem oraz zjawiska zachodzące w płaszczu. Te dwie ostatnie grupy zjawisk są przyczyną istotnego wzrostu tłumienia modów wyższych rzędów. Dla światłowodów kwarcowych ten wzrost tłumienności jest rzędu kilku do kilkunastu dB/km dla x = 1 w porównaniu z wartością dla x = 0 [2, 3, 4]. Dla światłowodów polimerowych, w związku z ich znacznie większą tłumiennością, ten wzrost jest znacznie większy i może wynosić nawet 150 dB/km [5].

Ostatecznie w niniejszej pracy przyjęto następującą zależność na tłumienność modową:

$$\alpha(x) = \alpha_0 + \Delta\alpha \cdot x^C \tag{5}$$

Tutaj C jest wykładnikiem potęgowym określającym przebieg wzrostu tłumienności (dalej dla światłowodów kwarcowych przyjęto C = 6), $\Delta\alpha$ jest wzrostem tłumienności, zaś α_0 jest wspólną dla wszystkich modów tłumiennością ograniczoną od dołu przez rozpraszanie Rayleigha.

Współczynnik sprzężenia modów, d(x), najczęściej jest wyrażany jako [1, 8, 9]

$$d(x) = d_0 x^{-2q} \tag{6}$$

gdzie parametr d_0, wyrażony w [1/km], określa wielkość sprzężenia między modami należącymi do sąsiednich grup modowych, zaś wykładnik q opisuje jego postać funkcjonalną i jest zależny zarówno od typu światłowodu, jak i rodzaju perturbacji wywołujących sprzężenie między modami (wahania wymiarów rdzenia, mikrozgięcia itp.).

W zastosowanym dalej modelu przyjęto dla światłowodu o profilu skokowym q = 1, zaś dla światłowodu o profilu parabolicznym q = 0 oraz q = -0.5. Używane w obliczeniach numerycznych wartości współczynnika sprzężenia znajdują się w zakresie do 30 [1/km] dla światłowodów gradientowych oraz w zakresie do 20...100 [1/km] dla światłowodów skokowych. Należy przy tym zaznaczyć, że bardzo małe wartości współczynnika sprzężenia praktycznie świadczą o braku wymiany energii między modami, co jest czasem przyjmowane jako punkt odniesienia do dalszej analizy [20].

Równanie (3) jest wygodniej rozwiązywać operując transformatą Fouriera wielkości P(x,z,t)

$$F(x,z,f) = \int_{-\infty}^{+\infty} P(x,z,t) \cdot \exp(-j2\pi f t)dt \tag{7}$$

Dokonując transformacji Fouriera obydwu stron równania (3) otrzymujemy

$$\frac{\partial F}{\partial z} + [\alpha(x) - j2\pi\tau(x)]F = \frac{1}{M}\frac{1}{x}\frac{\partial}{\partial x}\left[xd(x)\frac{\partial F}{\partial x}\right], \tag{8}$$

przy czym warunki brzegowe są następujące [1, 7, 10, 14, 15]

$$F(x,z,f)_{x=1} = 0 \tag{9}$$

$$\left[d(x)\frac{\partial F}{\partial x}\right]_{x=0} = 0 \tag{10}$$

$$F(x,0,f) = p(x) \tag{11}$$

Przyjęto pobudzenie impulsowe na wejściu w postaci:

$$P(x,0,t) = p(x)\cdot\delta(t) \tag{12}$$

gdzie p(x) jest wielkością określającą moc modu w grupie modowej x, zaś przez δ oznaczono deltę Diraca.

Jak już wspomniano poprzednio, liczba modów w danej grupie modowej jest w przybliżeniu proporcjonalna do x, a ponadto możliwe są dwie polaryzacje ortogonalne. Zatem moc przenoszona przez grupę modową x wynosi 2x·P(x,z,t), a moce całkowite na wejściu i wyjściu światłowodu są określone przez

$$P_{we}(t) = 2\delta(t)\int_0^1 x\cdot p(x)dx \tag{13}$$

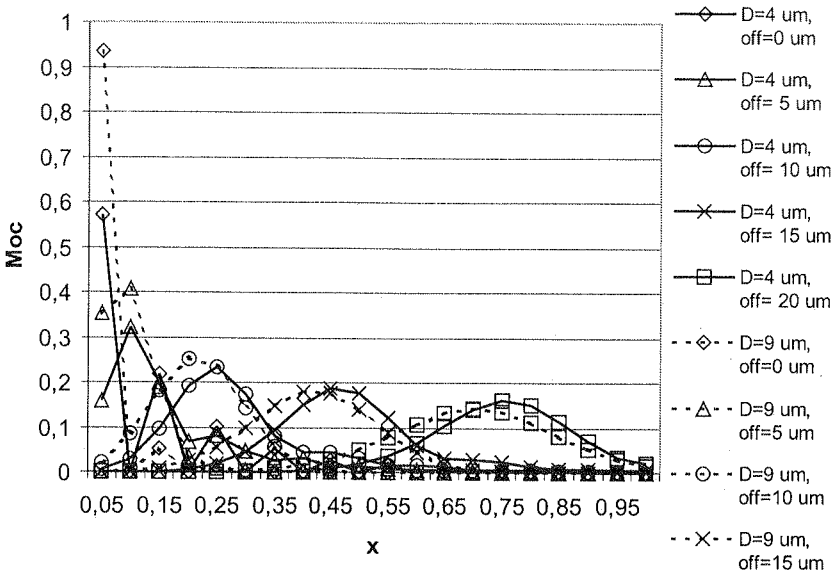
$$P_{wy}(t) = 2\int_0^1 x\cdot H(x)\cdot P(x,z,t)dx \tag{14}$$

W ostatniej zależności przyjęto możliwość filtracji grup modowych na wyjściu światłowodu. Opisywana jest ona przez funkcję H(x) (H(x) ≤ 1); przy braku takiej filtracji H(x) = 1.

Dokonując transformacji Fouriera zależności (13) i (14) otrzymujemy ostatecznie optyczną transmitancję częstotliwościową światłowodu T(f,z):

$$T(f,z) = \frac{\int_0^1 x\cdot H(x)\cdot F(x,z,f)dx}{\int_0^1 x\cdot p(x)dx} \tag{15}$$

Tutaj funkcja F(x,z,f) jest rozwiązaniem równania (8) przy warunkach brzegowych (9–12). Ze względu na trudności w rozwiązaniu analitycznym powszechnie poszukuje się rozwiązań równań (3) i (8) na drodze obliczeń numerycznych [2, 10, 14] i takie podejście zostanie dalej przyjęte w niniejszej pracy.



Rys. 1. Moce w grupach modowych, określonych przez x, obliczone numerycznie dla światłowodu wielomodowego (D' = 50 μm, λ = 0.78 μm, NA = 0.2, g = 2 – parametr określający profil współczynnika załamania) przy pobudzeniu offsetowym modem LP_{01} ze światłowodów skokowych o dwóch średnicach rdzenia (D = 4 μm, częstotliwość znormalizowana ν = 1.9; D = 9 μm, ν = 3.7). Widać, że węższa wiązka (D = 4 μm) pobudza nieco wyższe grupy modowe niż wiązka szersza (D = 9 μm)

Fig. 1. Power of the mode group x in a MM GI optical fiber calculated numerically for an offset LP_{01} mode launch

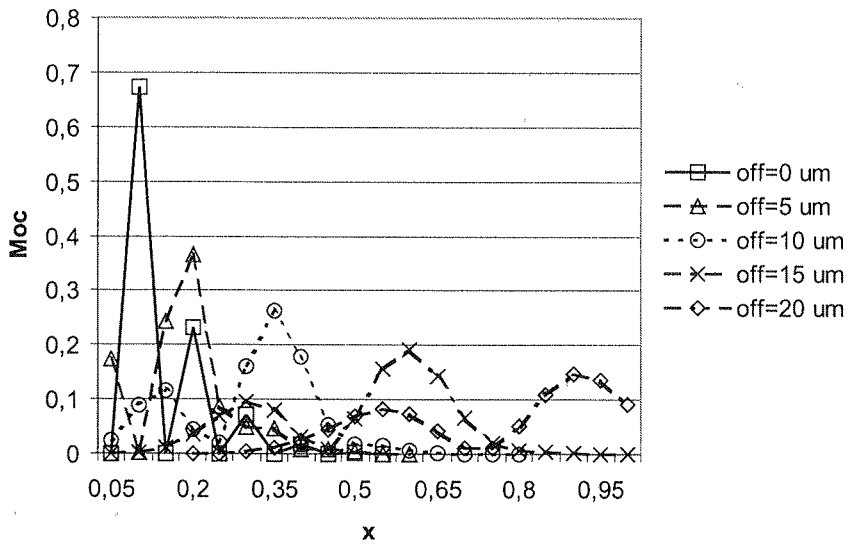
Osobną kwestią jest określenie pobudzenia p(x) na wejściu światłowodu. Zastosowane w niniejszej pracy podejście polega na rozkładzie pola sygnału pobudzającego na wejściu światłowodu na pola modów w badanym światłowodzie, które tworzą układ ortogonalny. Ta metoda, [4, 18], wymaga w etapie wstępnym obliczenia (numerycznego) rozkładów pól wszystkich modów rozchodzących się w światłowodzie. W niniejszej pracy rozkłady te były obliczane numerycznie zgodnie z metoda elementów skończonych [17]. Dla uzyskania wyników konieczna jest znajomość parametrów geometrycznych światłowodu, profilu współczynnika załamania oraz długość fali świetlnej. Następnie moc sprzężona do każdego modu jest obliczana jako korelacja elektrycznego pola wiązki wejściowej z polem danego modu [4, 18]. W dalszej części tego punktu zaprezentujemy wyniki uzyskane numerycznie przy różnych typach pobudzeń. Współczynniki pobudzenia (sprzężenia) modu LP_{lp} można określić z zależności [22, 30]:

$$c_{lp} = \iint \Phi_{lp} \Psi_{in}^{*} dS \tag{16}$$

gdzie Φ_{lp}, Ψ_{in} są odpowiednio rozkładami pól modu lp i światła pobudzającego, przy czym całkowanie rozciąga się po płaszczyźnie wejściowej. Przy założeniu jednostkowej mocy wejściowej i unormowaniu rozkładów pól modalnych, moc w modzie określonym przez indeks lp, P_{lp}, określona jest zależnością

$$P_{lp} = |c_{lp}|^2 \tag{17}$$

Aby obliczyć moc danej grupy modowej, należy zsumować moce wszystkich modów należących do tej grupy, zaś dla wyliczenia mocy modu należącego do tej grupy należy moc w grupie podzielić przez liczbę modów należących do tej grupy.



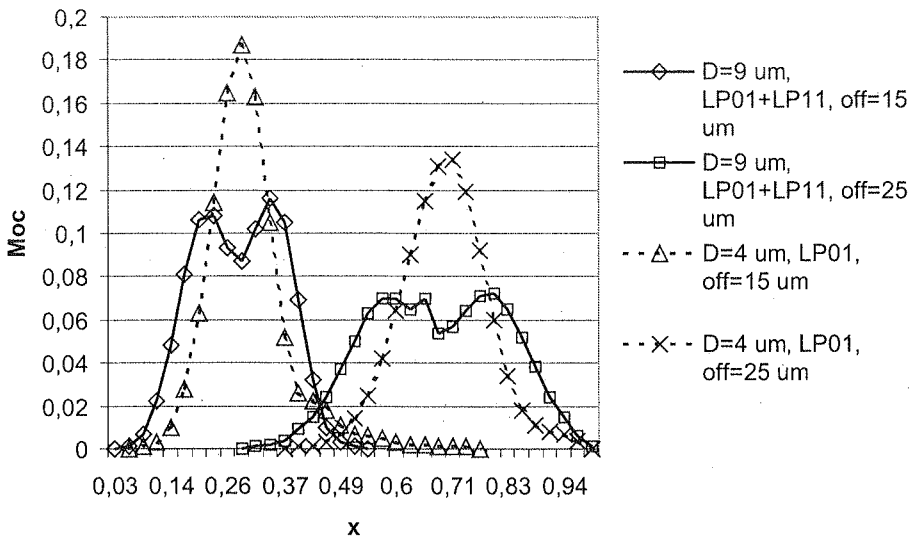
Rys. 2. Moce w różnych grupach modowych obliczone numerycznie dla światłowodu wielomodowego (D' = 50 μm, λ = 0.78 μm, NA = 0.2, g = 2) przy pobudzeniu offsetowym modem LP_{11} ze światłowodu skokowego (D = 9 μm, v = 3.7)

Fig. 2. Power of the mode group x in a MM GI optical fiber calculated numerically for an offset LP_{11} mode launch

Wartości pobudzeń, określone jako suma mocy wszystkich modów w grupie modów x, zostały porównane dla dwóch średnic rdzenia światłowodu pobudzającego. Rezultaty porównania pokazano na rys. 1. To porównanie pozwala wyciągnąć następujące wnioski:

1. Wiązka o mniejszej średnicy pobudza nieco wyższe grupy modów niż wiązka o większej średnicy.

2. W miarę wzrostu przesunięcia offsetowego zostają pobudzone coraz to wyższe grupy modów.



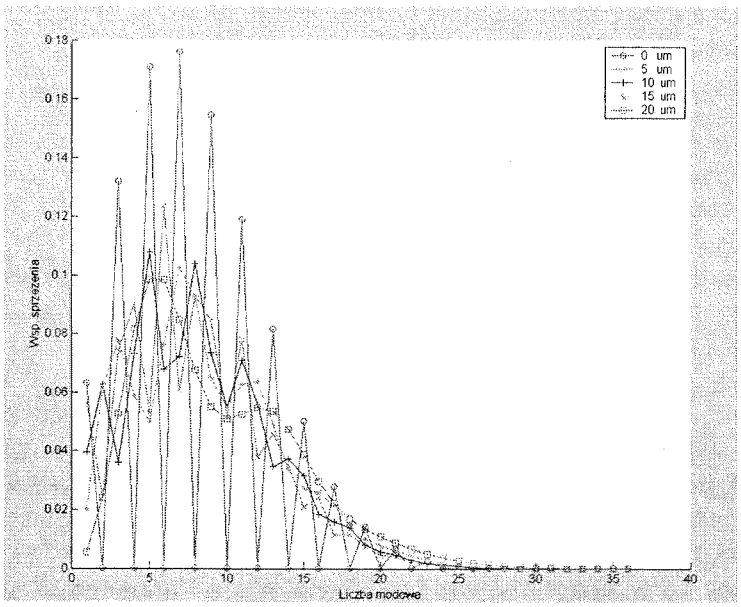
Rys. 3. Moce w różnych grupach modowych obliczone numerycznie dla światłowodu wielomodowego (D' = 62.5 μm, λ = 0.78 μm, NA = 0.275, g = 2) przy pobudzeniu offsetowym: modem LP_{01} ze światłowodu skokowego (D = 4 μm, ν = 1.9) – linie przerywane oraz kombinacją modów LP_{01} i LP_{11} o równych mocach (D = 9 μm, ν = 3.7) – linie ciągłe

Fig. 3. Power of the mode group x in a MM GI optical fiber calculated numerically for an offset LP_{01} mode launch, and for combined LP_{01} + LP_{11} modes launch

W niektórych przypadkach istotne mogą być pobudzenia modami innych rzędów niż zbliżony do gaussowskiego mod podstawowy LP_{01} oraz pobudzenia kombinacjami modów. Rezultaty obliczeń numerycznych dotyczących pobudzeń offsetowych modem LP_{11} zaprezentowano na rys. 2, zaś na rys. 3. pokazano pobudzenie poszczególnych grup modowych kombinacją modów LP_{01} i LP_{11} o równych mocach. Wszystkie wymienione tutaj przypadki mają istotne znaczenie, jeśli światłowód wielomodowy jest pobudzany w pierwszym oknie transmisyjnym światłowodem jednomodowym zgodnym z jednym z zaleceń G.652...G.655. W takim światłowodzie w pierwszym oknie transmisyjnym oprócz modu LP_{01} rozchodzi się jeszcze co najmniej mod LP_{11}. Jak widać z rezultatów obliczeń pokazanych na rys. 2, mod LP_{11} pobudza dwie odrębne grupy modów oddzielone od siebie przerwą (modami nie pobudzanymi). Taki rodzaj pobudzenia nie jest korzystny, gdyż te dwie grupy modowe istotnie różnią się opóźnościami grupowymi, co prowadzi do obniżenia pasma światłowodu.

Również pobudzenie kombinacją modów LP_{01} i LP_{11} powoduje rozchodzenie się znacznie większej liczby grup modowych, aniżeli dzieje się to w przypadku pobudzenia jedynie modem podstawowym LP_{01}, co pokazano na rys. 3. Prowadzi to do obniżenia pasma, co zostało potwierdzone eksperymentalnie.

Podobna sytuacja istnieje również przy pobudzaniu światłowodu skokowego; generalnie mod LP_{11} pobudza wyższe grupy modów niż mod LP_{01}. W odróżnieniu od światłowodu gradientowego (parabolicznego), gdzie zwiększenie offsetu prowadziło do pobudzenia coraz to wyższych grup modowych, dla światłowodu skokowego offset praktycznie nie zmienia pobudzonych grup modowych, co zaprezentowano na rys. 4. Prowadzi to do wniosku, iż w przypadku światłowodów skokowych zmiana offsetu nie prowadzi do zmian pasma światłowodu. Potwierdzono to eksperymentalnie (patrz część pomiarowa).



Rys. 4. Pobudzenie różnych grup modowych (określone przy pomocy współczynnika sprzężenia) w światłowodzie skokowym (D = 50 μm, NA = 0.2, λ = 0.78 μm) przy pobudzeniu offsetowym wiązką gaussowską o średnicy 4 μm

Fig. 4. Power of the mode group x in a MM SI optical fiber calculated numerically for an offset gaussian beam launch

Pobudzenie offsetowe światłowodem jednomodowym jest przykładem tzw. pobudzenia ograniczonego (ang. restricted launch – RL) [21, 29], w którym pobudzona zostaje jedynie ograniczona liczba modów. Cel stosowania pobudzenia ograniczonego jest dosyć prosty. Otóż, przy takim pobudzeniu w przypadku słabego mieszania modowego i niezbyt długiej linii, większość energii pozostanie w modach, którymi pobudzono światłowód. Wówczas, niektóre mody nie zostaną w ogóle pobudzone, co znakomicie zmniejsza dyspersję (między)modową i zwiększa pasmo modowe światłowodu. Zazwyczaj pozwala to na kilkukrotny wzrost pasma w stosunku do danych katalogowych uzyskanych przy opisanym niżej pobudzeniu OFL.

Poza pobudzeniami światłowodem jednomodowym i wiązką równoległą (czyli pobudzeń ograniczonych) dla celów porównawczych badano również wyniki uzyskiwane przy pobudzeniach standardowych EMD i OFL, które stanowią dobre modele pobudzenia światłowodu za pomocą diody elektroluminescencyjnej. Pierwsze z tych pobudzeń (EMD – equilibrium mode distribution) odpowiada pobudzeniu światłowodu będącego w stanie równowagi modowej i jest wykorzystywane przy pomiarach tłumienności światłowodu [16, 25]. Natomiast drugie pobudzenie (OFL – overfilled launch) odpowiada pobudzeniu uzyskiwanemu ze skramblera modowego typu SGS (połączone odpowiednio odcinki światłowodów skokowego, gradientowego i skokowego) i jest wykorzystywane przy pomiarze pasma modowego światłowodów [16, 25]. Zależności określające obydwa typy pobudzeń są następujące [23, 29].
Dla pobudzenia EMD

$$p(x) = 1 - x^{0.75}, \tag{18}$$

zaś dla pobudzenia OFL

$$p(x) = 1 - 0.5x. \tag{19}$$

3. WYNIKI OBLICZEŃ

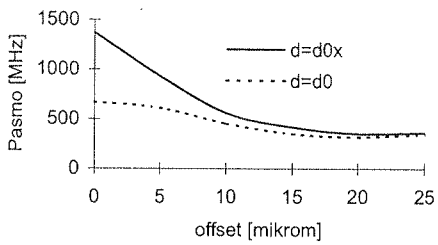
Jak już wspomniano, równanie (8) w ogólnym przypadku daje się rozwiązać jedynie numerycznie. Do tego celu w niniejszej pracy użyto metody różnic skończonych [24] w dwóch wariantach: metody explicite oraz metody Du Fort'a i Frankel'a [24]. W tym miejscu trzeba tylko zaznaczyć, iż każdy rozpatrywany przypadek rozwiązywano niezależnie dwiema metodami, aż do uzyskania zgodności rozwiązań.

Jeśli nie określono tego inaczej, prezentowane dalej rezultaty uzyskano dla następujących parametrów światłowodu: g = 1.8, ε = 0.05, λ = 0.85 μm, D = 62.5 μm, NA = 0.275, n_1 = 1.48, L = 1 km, co odpowiada danym katalogowym jednego ze światłowodów użytych w części pomiarowej. Przy obliczeniach pasma przyjęto także α_0 = 1.53 dB/km, $\Delta\alpha$ = 10 dB, jednak te dwa ostatnie parametry miały mały wpływ na wyniki obliczeń pasma modowego (zmiana $\Delta\alpha$ o 5 dB powodowała zmianę pasma o mniej niż 10 MHz w zbadanych przypadkach).

Stosowane dalej, przy omówieniu wyników obliczeń numerycznych pojęcie pasma (modowego) dotyczy pasma określonego przez spadek mocy optycznej o 3 dB (pasmo optyczne) i jest zgodne z odpowiednimi zaleceniami ITU [16]. Odpowiada ono 6 dB pasmu elektrycznemu.

Na pasmo określone przez dyspersję modową mają wpływ następujące czynniki:
- charakterystyka opóźności grupowej τ(x)
- wielkość sprzężenia między modami, określona przez typ zależności d(x) oraz przez stałą d_0

- długość światłowodu
- rodzaj pobudzenia.



Rys. 5. Porównanie pasma modowego uzyskanego przy pobudzeniu offsetowym plamką gaussowską o średnicy 9 μm dla dwóch modeli zależności współczynnika sprzężenia (d_0 = 10), profil bez zniekształceń

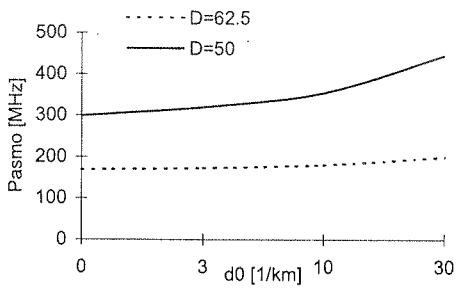
Fig. 5. Comparison of the modal bandwidth for an offset launch, and for two coupling models

W przypadku pobudzenia ograniczonego (RL) słabsze mieszanie modów (mniejsze d0 oraz zależność typu d(x)~ x) prowadzi do zwiększenia pasma, o ile światłowód nie zawiera błędów profilu współczynnika załamania. Jest to związane z tym, że przy mniejszym sprzężeniu nie zostaną pobudzone wszystkie mody, a zatem maksymalna różnica opóźności grupowych będzie mniejsza i pasmo będzie większe. Zależność pasma od charakteru sprzężenia przy pobudzeniu offsetowym plamką gaussowską o średnicy 9 μm pokazano na rys. 5 dla dwóch różnych zależności funkcjonalnych współczynnika sprzężenia: $d(x) = d_0$ oraz $d(x) = d_0 x$. Pasma modowe są istotnie różne dla pobudzenia centralnego, gdzie praktycznie przy braku sprzężenia dla zależności d(x)~ x pasmo jest 2-krotnie większe niż dla zależności $d(x) = d_0$, zaś zbiegają do siebie dla pobudzeń peryferyjnych, gdyż współczynniki sprzężenia modów są wtedy porównywalne.

W przypadku ustalonej odległości i pobudzenia OFL (19) ze wzrostem d_0 następuje wzrost pasma modowego; pokazano to na rys. 6. Można to wyjaśnić silniejszym mieszaniem modów, a co za tym idzie zależnością pasma od długości światłowodu typu $1/\sqrt{L}$, a nie 1/L.

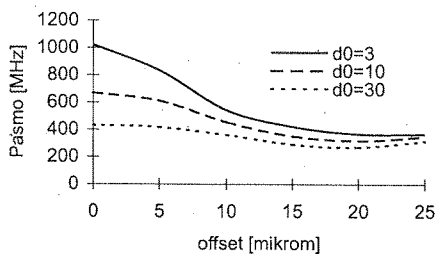
Dla pobudzenia offsetowego, pasmo światłowodu w dużym stopniu zależy od przesunięcia plamki (światłowodu pobudzającego) od osi światłowodu pobudzanego (offsetu). Dla regularnego profilu współczynnika załamania pasmo światłowodu istotnie zależy również od średnicy plamki. W przypadku stosunkowo dużych średnic (np. 9 μm) pobudzona zostaje ograniczona liczba modów i wówczas pasmo maleje wraz z przesunięciem offsetowym, gdyż przy pobudzeniu centralnym pobudzona zostają niewielka liczba modów najniższych rzędów o mało różniących się opóźnościach grupowych (patrz rys. 1). W przypadku większych wartości przesunięć offsetowych liczba pobudzonych modów jest większa, bardziej różnią się one opóźnościami, a zatem i pasmo jest mniejsze. Z przyczyn omówionych poprzednio wzrost wartości współczynnika sprzężenia powoduje spadek pasma. Powyższe zależności pokazano na rys. 7. Z tych

samych przyczyn (generalnie mniejsze wartości współczynników sprzężenia) pasma dla zależności typu $d(x) = d_0x$ są większe.



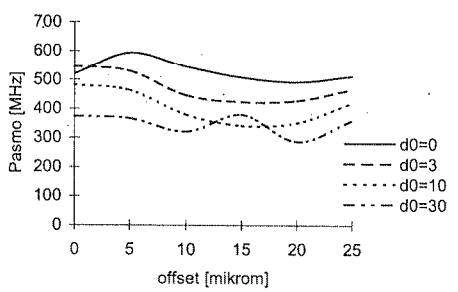
Rys. 6. Zależność pasma modowego przy pobudzeniu OFL od współczynnika sprzężenia między modami ($d(x) = d_0$) dla dwóch typów światłowodów: D = 62.5 μm, NA = 0.275 oraz D = 50 μm, NA = 0.2. Profil bez zniekształceń

Fig. 6. Modal bandwidth for OFL versus the coupling coefficient



Rys. 7. Zależność pasma modowego od przesunięcia offsetowego: pobudzenie plamką gaussowską o średnicy 9 μm, $d(x) = d_0$). Profil bez zniekształceń

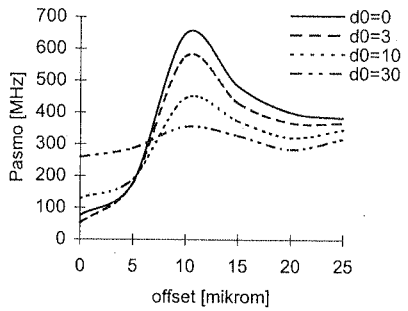
Fig. 7. Modal bandwidth for gaussian 9 μm beam launch versus the offset



Rys. 8. Zależność pasma modowego od przesunięcia offsetowego: pobudzenie plamką gaussowską, średnica plamki 5 μm, $d(x) = d_0$). Profil bez zniekształceń

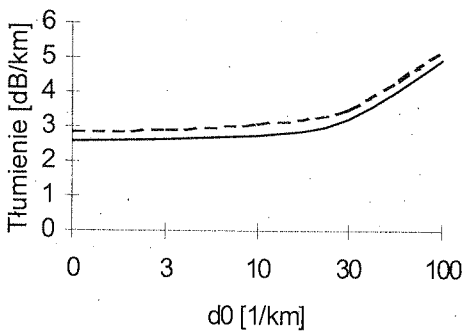
Fig. 8. Modal bandwidth for gaussian 5 μm beam launch versus the offset

Zagadnienie jest bardziej złożone dla mniejszych średnic plamki pobudzającej (np. 5 μm). Generalnie takie małe plamki pobudzają większą liczbę modów i wyższe grupy modowe, w związku z tym omówione powyżej zależności nie są tak wyraźne. Pasmo zależy wtedy znacznie słabiej od wielkości przesunięcia offsetowego. Pokazano to na rys. 8. Taką zależność zaobserwowano również eksperymentalnie.



Rys. 9. Zależność pasma modowego od przesunięcia offsetowego: plamka gaussowska o średnicy 9 μm, d(x) = d_0). Profil zniekształcony: h = 0.5 i x_0 = 0.05 – patrz zależność (20)

Fig. 9. Modal bandwidth for gaussian 9 μm beam launch versus the offset. Distorted n(r) profile



Rys. 10. Zależność tłumienności jednostkowej od wielkości sprzężenia dla pobudzenia EMD, d(x) = const: linia ciągła: α_0 = 2 dB/km, $\Delta\alpha$ = 10 dB, linią przerywana: α_0 = 2.5 dB/km, $\Delta\alpha$ = 6 dB

Fig. 10. Fiber attenuation for EMD launch versus the coupling coefficient

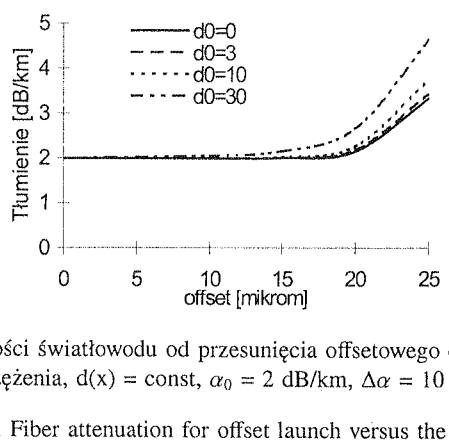
Zupełnie inna sytuacja ma miejsce w przypadku, kiedy profil współczynnika załamania jest zniekształcony. Dla celów niniejszej pracy wpływ zaburzeń profilu współczynnika załamania był modelowany w obliczeniach poprzez uzupełnienie zależności (4) o zaburzenie o kształcie funkcji trójkątnej [27, 28]:

$$\tau_1(x) = \tau(x) - \frac{N_1}{c}\Delta \cdot h \cdot \left(1 - \frac{x}{x_0}\right), \quad x \leqslant x_0$$
$$\tau_1(x) = \tau(x), \quad x > x_0 \tag{20}$$

Tutaj h jest względną głębokością wgłębienia (ogólnie zaburzenia); typowo $0 < h < 1$, zaś x_0 określa grupę modową o największym numerze jeszcze dotkniętą tym zaburzeniem. Pasmo modowe dla profilu zniekształconego i opóźności $\tau(x)$ zgodnej z zależnością (20) dla $h = 0.5$ i $x_0 = 0.05$ pokazano na rys. 9. Widać wyraźny spadek pasma (nawet znacznie poniżej wartości katalogowej mierzonej dla OFL) dla pobudzenia centralnego związany z tym, że zniekształcenie profilu ma największy wpływ na mody najniższych rzędów. Wspomniany spadek pasma jest łagodzony przez intensywne mieszanie modów (wzrost d_0). Przy bardzo silnych sprzężeniach pasmo słabo zależy od przesunięcia offsetowego. Przedstawione powyżej wykresy potwierdzają słuszność stosowania pobudzenia offsetowego [27, 31] oraz nakładania ograniczeń na tzw. strumień otoczony (ang. encircled flux) [13, 26], które pozwalają zredukować wpływ częstych zniekształceń profilu w centralnej i zewnętrznych częściach światłowodu.

Trzeba zaznaczyć, że chociaż większość wyników obliczeń pasma modowego jest prezentowana dla światłowodu o średnicy rdzenia 62.5 μm, to podobne zależności uzyskano również dla światłowodów o innej średnicy (np. 50 μm). Należy przy tym pamiętać o większym paśmie, jakim charakteryzują się takie światłowody ze względu na mniejszą aperturę numeryczną (mniejsze NA).

Tłumienie światłowodu wielomodowego zależy nieco słabiej od warunków pobudzenia aniżeli jego pasmo. Ponadto tłumienie nie jest funkcją ewentualnych zniekształceń profilu współczynnika załamania. Słabo także zależy od postaci funkcjonalnej $d(x)$, a w większym stopniu od wielkości sprzężenia, określonej przez d_0. Na rys. 10. pokazano zależność tłumienności jednostkowej dla pobudzenia EMD zbliżonego do pobudzenia diodą elektroluminescencyjną.



Rys. 11. Zależność tłumienności światłowodu od przesunięcia offsetowego dla różnych współczynników sprzężenia, $d(x) = \text{const}$, $\alpha_0 = 2$ dB/km, $\Delta\alpha = 10$ dB

Fig. 11. Fiber attenuation for offset launch versus the offset

Wzrost siły sprzężenia modów powoduje wzrost tłumienia wskutek stałej dyfuzji mocy do modów wyższych rzędów i w efekcie do modów płaszczowych. Widać to również na rys. 11, gdzie pokazano zależność tłumienia od przesunięcia offsetowego. Obserwowany jest wzrost tłumienia dla dużych przesunięć offsetowych: może ono być większe niż tłumienie przy EMD, podczas gdy dla małych przesunięć offsetowych

tłumienie jest mniejsze niż tłumienie przy EMD. Taką zależność tłumaczy się znowu większym tłumieniem modów wyższych rzędów. Zależność pokazana na rys. 11 jest dobrze udokumentowana eksperymentalnie (patrz część pomiarowa).

4. POMIARY

Wyniki pomiarów zaprezentowane dalej mogą służyć jedynie do jakościowej weryfikacji rezultatów obliczeń teoretycznych. Autorzy nie mieli bowiem możliwości ilościowego porównania wyników teoretycznych z rezultatami pomiarów. Wynika to stąd, iż dla określenia parametrów światłowodu wykorzystywanych w modelach teoretycznych konieczne są:

- bardzo precyzyjna znajomość profilu współczynnika załamania światłowodu n(r) (przeprowadzone pomiary nie były dostatecznie dokładne),
- dokładna znajomość dyspersji materiałowej, a zwłaszcza dyspersji profilowej badanego światłowodu, co wymaga albo dokładnej znajomości składu materiałowego światłowodu, co jest nierealne w przypadku światłowodów komercyjnych, albo pomiarów sprzętem, którym nie dysponowaliśmy,
- wreszcie znajomości zależności funkcjonalnej współczynnika sprzężenia modów oraz jego wartości liczbowej, czego znowu nie da się przeprowadzić bez złożonych pomiarów; w dodatku ten współczynnik jest funkcją pokrycia światłowodu i jego ułożenia.

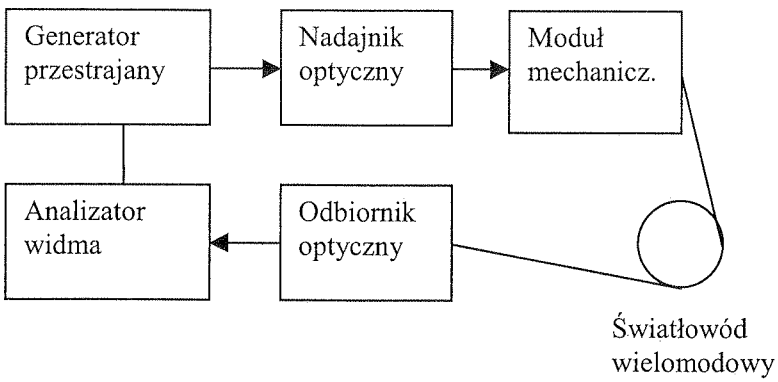
4.1. UKŁAD POMIAROWY

Dla przeprowadzenia pomiarów wykonano specjalizowany zestaw pomiarowy, którego podstawowymi blokami są wobulowany generator sterujący nadajnikiem optycznym oraz odbiornik optyczny połączony z analizatorem widma elektrycznego. Nadajnik optyczny poprzez użycie wymiennych głowic pozwala na pracę na długościach fal 650 nm i 780 nm. Schemat blokowy układu pomiarowego pokazano na rys. 12, zdjęcie samego zestawu pomiarowego – na rys. 13. Sygnał o zmiennej częstotliwości z wobulowanego generatora steruje nadajnikiem optycznym, w którym jest dokonywana konwersja elektryczno-optyczna. Zastosowany w urządzeniu układ optyczno-mechaniczny umożliwia precyzyjne wprowadzenie do badanego światłowodu wielomodowego:

a) skolimowanej wiązki optycznej pod zmiennym kątem,
b) plamki optycznej pochodzącej z pobudzającego światłowodu jednomodowego w regulowanym punkcie rdzenia badanego światlowodu wielomodowego (pobudzenie offsetowe).

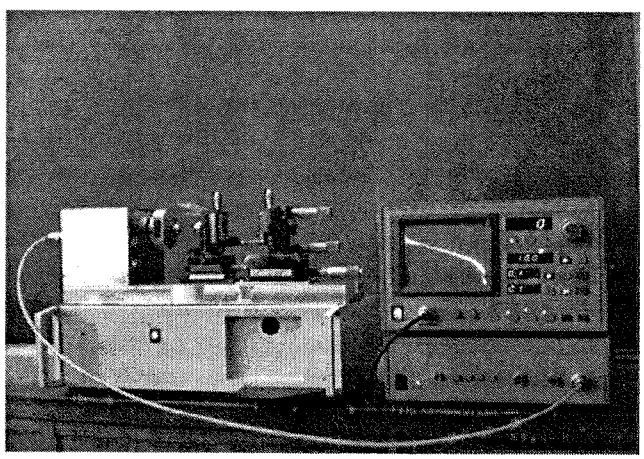
Po propagacji w światłowodzie wielomodowym o zadanym typie i długości w odbiorniku zostaje dokonana konwersja optyczno-elektryczna i sygnał zostaje przekazany do analizatora widma. Pozwala to na pomiar zależności pasma światłowodu od warunków pobudzenia, przy czym pasmo optyczne samego układu pomiarowego wynosiło około 700 MHz. Dołączenie po stronie odbiorczej zamiast konwertera O/E i wobulo-

skopu miernika mocy optycznej pozwala dokonać analogicznych pomiarów odnośnie zależności mocy optycznej na wyjściu światłowodu od warunków pobudzenia. Za pomocą opisanego zestawu pomiarowego dokonano kilku serii pomiarów różnych typów światłowodów wielomodowych.



Rys. 12. Schemat blokowy układu pomiarowego

Fig. 12. Block scheme of the measurement setup



Rys. 13. Widok zestawu pomiarowego

Fig. 13. Photo of the measurement setup

Mierzono dwa rodzaje typowych gradientowych światłowodów kwarcowych o długościach 500 m i 1000 m oraz odcinek światłowodu skokowego o rdzeniu kwarcowym i płaszczu polimerowym o długości 750 m. Dane katalogowe światłowodów kwarcowych podano w tabl. 1, zaś parametry światłowodu kwarcowo-polimerowego są następujące: apertura numeryczna NA = 0.375, współczynniki załamania rdzenia

i płaszcza odpowiednio $n_1 = 1.458$ i $n_2 = 1.409$ (dla $\lambda = 589$ nm), wreszcie średnica rdzenia wynosi 200 μm.

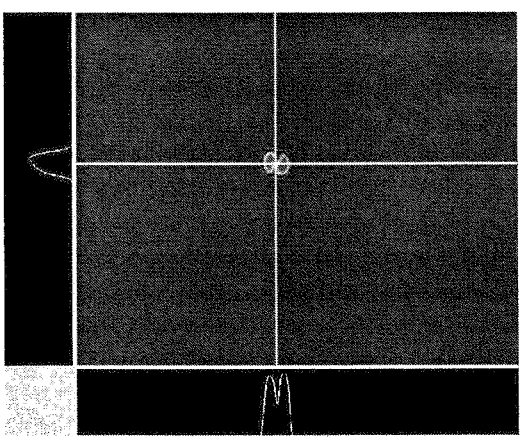
Tabela 1

Podstawowe dane katalogowe badanych kwarcowych światłowodów gradientowych [11]

Basic data of the tested silica GI optical fibers

Parametr/średnica rdzenia	50 μm	62.5 μm
Tłumienność jednostkowa [0.85 μm]	3.0 dB/km	3.5 dB/km
Tłumienność jednostkowa [1.31 μm]	1.0 dB/km	1.0 dB/km
Pasmo modowe [0.85 μm]	$\geq$ 200 MHz-km	$\geq$ 160 MHz-km
Pasmo modowe [1.3 μm]	$\geq$ 500 MHz-km	$\geq$ 500 MHz-km
Średnica rdzenia	50 $\pm$ 2 μm	62.5 $\pm$ 2 μm
Średnica płaszcza	125 $\pm$ 2 μm	125 $\pm$ 2 μm
Aperatura numeryczna	0.200 $\pm$ 0.015	0.275 $\pm$ 0.015
Błąd koncentryczności rdzenia względem płaszcza	$\leq$ 3 μm	$\leq$ 3 μm

Oprócz zbadania zależności pasma i tłumienia światłowodu od warunków pobudzenia, dokonano również pomiarów rozkładów natężenia pola optycznego wiązek pobudzających, jak również wyjściowych. Do tego celu wykorzystano, opracowany w Instytucie Telekomunikacji PW, monitor luminancji [12]. Przyrząd ten umożliwia również przybliżony pomiar rozkładu współczynnika załamania światłowodu.

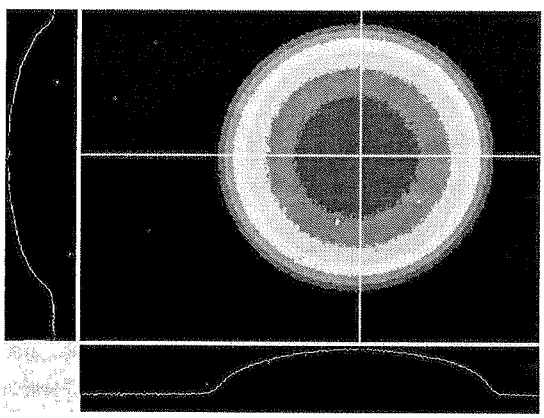


Rys. 14. Pomierzony monitorem luminancji rozkład poprzeczny wiązki świetlnej wychodzącej ze światłowodu jednomodowego o średnicy rdzenia 9 μm. Widać, że oprócz modu LP_{01} pobudzony został również mod LP_{11}

Fig. 14. Power distribution of the light at the output of single mode fiber

4.2. WYNIKI POMIARÓW

Na wstępie zajmiemy się omówieniem wyników pomiarów rozkładów natężeń pól. Na rys. 14 przedstawiono wyniki pomiaru rozkładu pola pobudzającego, przy pobudzeniu światłowodem jednomodowym o średnicy rdzenia 9 μm. Z pomiaru widać, że oprócz modu LP_{01} istnieje również mod LP_{11}, co (jak wykazano w części teoretycznej) pobudza więcej grup modowych i (jak wykażą zaprezentowane dalej pomiary pasma) obniża pasmo modowe badanego światłowodu.



Rys. 15. Wynik pomiaru współczynnika załamania dla światłowodu 50 μm

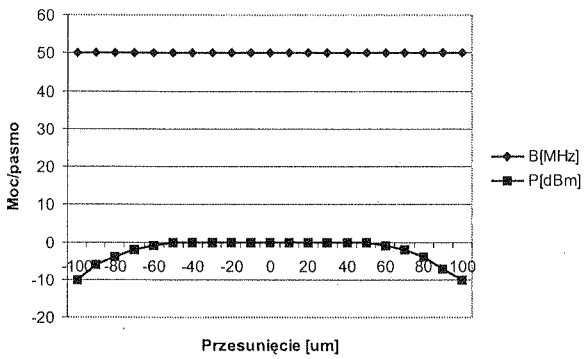
Fig. 15. Index of refraction profile of 50 μm fiber

Wreszcie na rys. 15 pokazano przykładowy wynik pomiaru profilu współczynnika załamania badanego światłowodu 50 μm. Dokładność tych pomiarów była jednak niewystarczająca, aby na ich podstawie można było wnioskować o parametrach częstotliwościowych światłowodów.

Na rys. 16–18 zaprezentowano wyniki pomiarów pasma i tłumienności dla światłowodu skokowego kwarcowo-polimerowego o średnicy rdzenia 200 μm. Z przedstawionych rezultatów można wyciągnąć następujące wnioski:

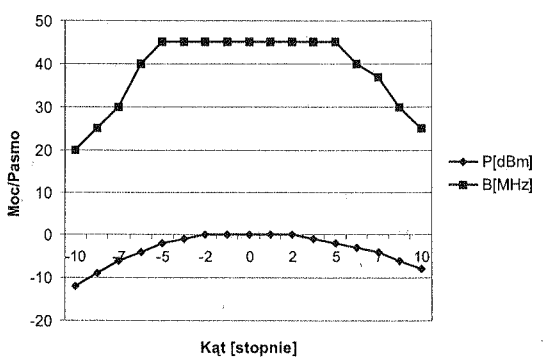
1. Pasmo modowe światłowodu skokowego jest w szerokich granicach niezależne zarówno od przesunięcia offsetowego (przy pobudzeniach światłowodami jednomodowy mi i wiązką skolimowaną), jak również od kąta, jaki wiązka skolimowana tworzy z osią światłowodu (przy pobudzeniu wiązką skolimowaną rozchodzącą się pod kątem do osi światłowodu). Podobnie zachowuje się tłumienie światłowodu. Wyjątkiem jest przypadek pobudzania wiązką równoległą skierowaną pod kątem do osi światłowodu, jeśli ten kąt przekracza 5°. W tym przypadku pasmo światłowodu zaczyna maleć, a tłumienie rosnąć – patrz rys. 17. Należy zaznaczyć, że omawiane kąty znajdują się znacznie poniżej kąta akceptacji światłowodu, który w omawianym przypadku przekracza 20°. Wynika stąd, iż wzrost tłumienia jest spowodowany większym tłumieniem modów wyższych rzędów.

2. Pobudzanie wiązką równoległą do osi daje największe pasmo modowe, nieco mniejsze pasmo osiągane jest przy pobudzeniu modem LP_{01} (światłowód o średnicy rdzenia 4 μm), zaś najmniejsze pasmo jest uzyskiwane dla pobudzeń kombinacją modów LP_{01} i LP_{11} (światłowód o średnicy rdzenia 9 μm). Takie rezultaty są spowodowane tym, że wiązka skolimowana pobudza najmniej grup modowych, zaś kombinacja modów LP_{01} i LP_{11} – najwięcej.



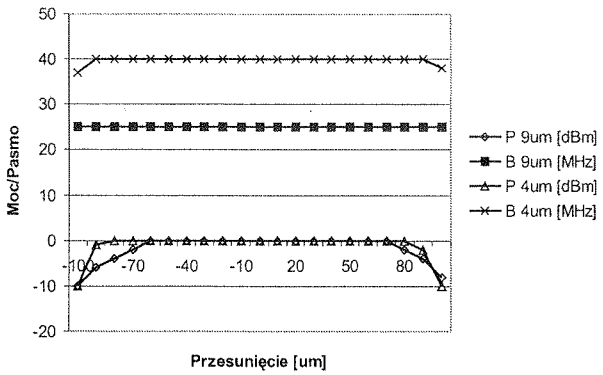
Rys. 16. Zależność mocy wyjściowej (dBm) i pasma (MHz) dla światłowodu skokowego o rdzeniu kwarcowym (Φ = 200 μm) i płaszczu polimerowym. Długość światłowodu L = 750 m. Pobudzenie wiązką równoległą do osi światłowodu o zmiennym przesunięciu względem środka rdzenia. Długość fali λ = 650 nm

Fig. 16. Output power and bandwidth for SI (silica core, plastic cladding) fiber excited with an offset beam



Rys. 17. Zależność względnej mocy wyjściowej (dBm) i pasma (MHz) dla światłowodu skokowego o rdzeniu kwarcowym (Φ = 200 μm) i płaszczu polimerowym. Długość światłowodu L = 750 m. Pobudzenie wiązką równoległą położoną centralnie względem rdzenia i skierowaną pod zmiennym kątem do osi światłowodu. Długość fali λ = 650 nm

Fig. 17. Output power and bandwidth for SI (silica core, plastic cladding) fiber excited with an angled offset

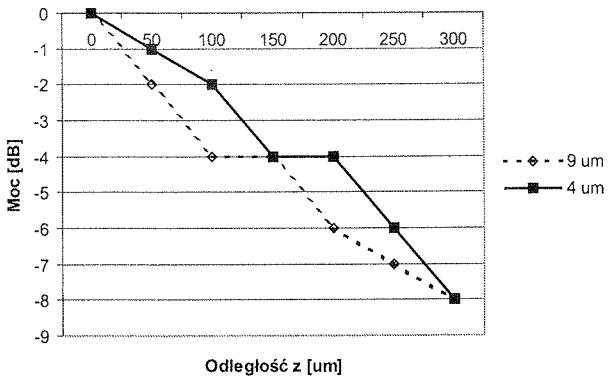


Rys. 18. Zależność względnej mocy wyjściowej P (dBm) i pasma B (MHz) dla światłowodu skokowego o rdzeniu kwarcowym (Φ = 200 μm) i płaszczu polimerowym. Długość światłowodu L = 750 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm stykającymi się czołowo z badanym światłowodem przy zmiennym przesunięciu (offset) względem środka jego rdzenia. Długość fali λ = 780 nm

Fig. 18. Output power and bandwidth for SI (silica core, plastic cladding) fiber excited with an offset single mode fiber

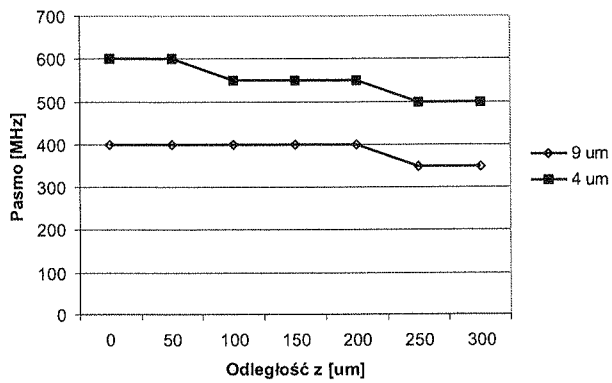
W przypadku światłowodów gradientowych stosowano wyłącznie pobudzenia światłowodami jednomodowymi o średnicach rdzenia 4 μm i 9 μm. Zbadano wpływ odległości z między czołami światłowodów pobudzającego i pobudzanego dla zerowego offsetu (pobudzenie centralne) na pasmo i tłumienie światłowodu oraz wpływ przesunięcia offsetowego między środkami rdzeni obydwu światłowodów przy z = 0 na te same parametry. Rezultaty pomiarów dla zmiennej odległości z między czołami światłowodów pokazano na rys. 19–25 dla różnych średnic rdzenia i długości mierzonego światłowodu. Pomiarów pasma nie udało się przeprowadzić dla światłowodu o Φ = 50 μm i długości 500 m, gdyż jego pasmo przekraczało pasmo układu pomiarowego. Zaobserwowano następujące prawidłowości:

1. Moc odbierana na końcu światłowodu maleje wraz ze wzrostem z, co jest oczywiste, gdyż spada wtedy również moc wprowadzana do światłowodu
2. W przypadku pobudzenia światłowodem o średnicy rdzenia Φ = 4 μm w miarę wzrostu z maleje również pasmo światłowodu badanego. Pasmo przy pobudzaniu światłowodem o Φ = 9 μm słabo zależy od wartości z, przy czym to pasmo jest wyraźnie mniejsze od pasma przy pobudzeniu światłowodem o mniejszej średnicy. Przy wzroście odległości z wiązka Φ = 4 μm obejmuje coraz większy obszar rdzenia pobudzanego światłowodu wielomodowego i pobudza coraz więcej grup modowych, co powoduje spadek pasma. Ten efekt nie jest widoczny w przypadku pobudzenia światłowodem o Φ = 9 μm, gdyż odpowiada ono pobudzeniu kombinacją modów LP_{01} i LP_{11}, która pobudza wiele grup modowych (patrz część teoretyczna) i oddalanie czoła światłowodu niewiele tu zmienia.



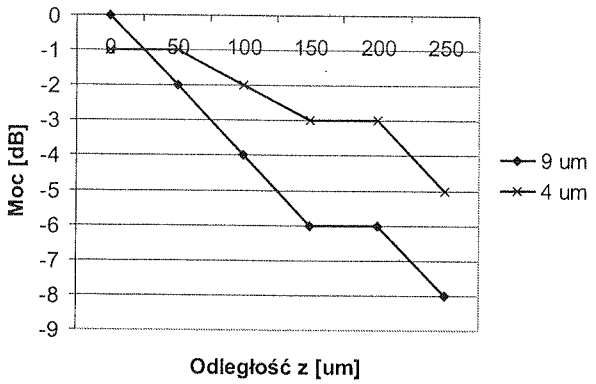
Rys. 19. Zależność mocy odbieranej dla gradientowego światłowodu kwarcowego (Φ = 62.5 μm) i długości L = 500 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm położonymi centralnie względem badanego światłowodu przy zmiennej odległości z między czołami światłowodów. Długość fali λ = 780 nm

Fig. 19. Received power for GI silica fiber versus distance z to the exciting single mode fiber



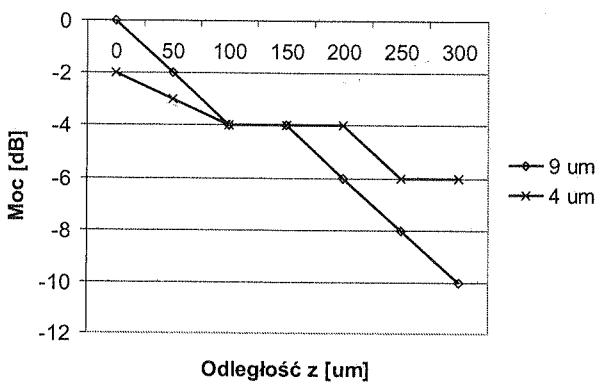
Rys. 20. Zależność pasma (MHz) dla gradientowego światłowodu kwarcowego (Φ = 62.5 μm) i długości L = 500 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm położonymi centralnie względem badanego światłowodu przy zmiennej odległości z między czołami światłowodów. Długość fali λ = 780 nm

Fig. 20. Modal bandwidth of GI silica fiber versus distance z to the exciting single mode fiber



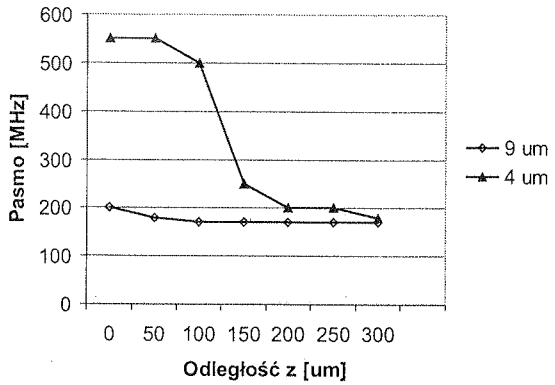
Rys. 21. Zależność mocy odbieranej dla gradientowego światłowodu kwarcowego (Φ = 50 μm) i długości L = 500 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm położonymi centralnie względem badanego światłowodu przy zmiennej odległości z między czołami światłowodów. Długość fali λ = 780 nm

Fig. 21. Received power for GI silica fiber versus distance z to the exciting single mode fiber



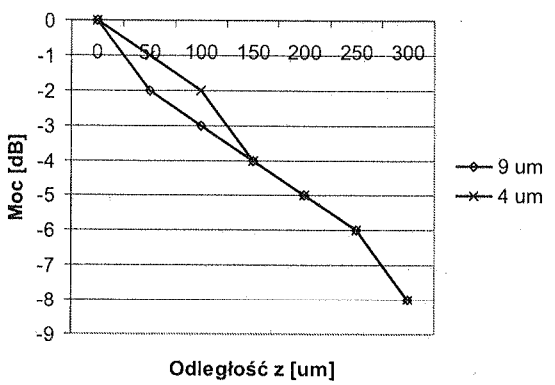
Rys. 22. Zależność mocy odbieranej dla gradientowego światłowodu kwarcowego (Φ = 50 μm) i długości L = 1000 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm położonymi centralnie względem badanego światłowodu przy zmiennej odległości z między czołami światłowodów. Długość fali λ = 780 nm

Fig. 22. Received power for GI silica fiber versus distance z to the exciting single mode fiber



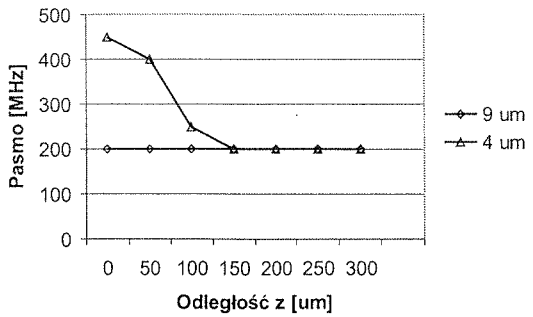
Rys. 23. Zależność pasma (MHz) dla gradientowego światłowodu kwarcowego (Φ = 50 μm) i długości L = 1000 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm położonymi centralnie względem badanego światłowodu przy zmiennej odległości z między czołami światłowodów. Długość fali λ = 780 nm

Fig. 23. Modal bandwidth of GI silica fiber versus distance z to the exciting single mode fiber



Rys. 24. Zależność mocy odbieranej dla gradientowego światłowodu kwarcowego (Φ = 62.5 μm) i długości L = 1000 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm położonymi centralnie względem badanego światłowodu przy zmiennej odległości z między czołami światłowodów. Długość fali λ = 780 nm

Fig. 24. Received power for GI silica fiber versus distance z to the exciting single mode fiber



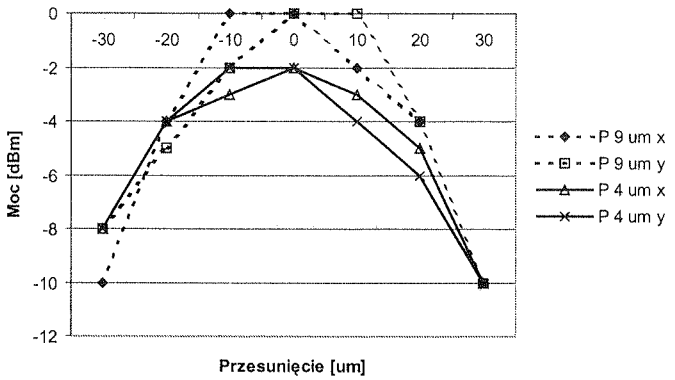
Rys. 25. Zależność pasma (MHz) dla gradientowego światłowodu kwarcowego (Φ = 62.5 μm) i długości L = 1000 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm położonymi centralnie względem badanego światłowodu przy zmiennej odległości z między czołami światłowodów. Długość fali λ = 780 nm

Fig. 25. Modal bandwidth of GI silica fiber versus distance z to the exciting single mode fiber

Z kolei rezultaty pomiarów dla zmiennych przesunięć offsetowych pokazano na rys. 26–32. przy czym ponownie nie udało się przeprowadzić pomiarów pasma dla światłowodu o Φ = 50 mm i długości 500 m z powodu niedostatecznego pasma układu pomiarowego. Rezultaty pomiarów można podsumować następująco:

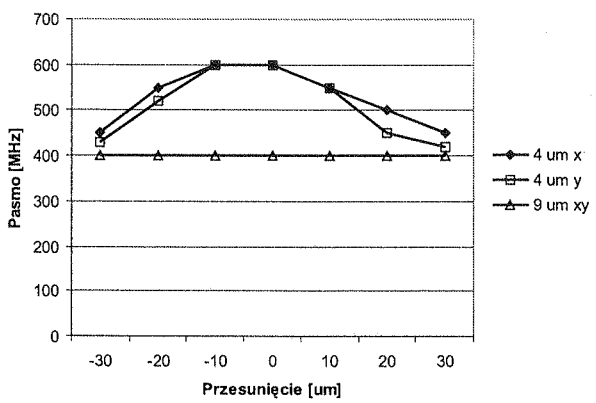
1. Przebiegi zależności tłumienia światłowodu od przesunięcia offsetowego są dosyć podobne i charakteryzują się minimum tłumienia dla zerowych przesunięć offsetowych i wzrostem tłumienia rzędu 3 dB dla przesunięć offsetowych rzędu 15–20 μm i rzędu 10 dB dla przesunięć rzędu 25–30 μm. Należy tutaj zaznaczyć, że ten wzrost jest spowodowany różnicą tłumienia modów różnych rzędów, a nie zmiana wejściowego współczynnika sprzężenia do modów prowadzonych; współczynik ten jest bliski 100% nawet dla dużych wartości przesunięć offsetowych. Przy tym, co jest oczywiste, dla danego offsetu tłumienie rośnie wraz ze wzrostem długości światłowodu i maleniem jego średnicy. Zaobserwowano również mniejszy wzrost tłumienności przy jednakowym offsecie dla światłowodu 4 μm niż dla światłowodu 9 μm. Można to wyjaśnić następująco: wzrost tłumienia przy wzroście offsetu jest spowodowany pobudzaniem coraz wyższych grup modowych, które mają większą tłumienność. Jednakże światłowód 4 μm pobudza niższe grupy modowe niż światłowód 9 μm (kombinacja modów LP_{01} i LP_{11}), co powoduje jego mniejszą tłumienność.
2. Mierzone pasma modowe są największe przy pobudzeniu centralnym (offset = 0) i maleją przy wzroście offsetu w przypadku światłowodu 4 μm, zaś słabo zależą od wielkości offsetu dla pobudzenia światłowodem 9 μm. Ponadto pasma dla światłowodu 4 μm są zawsze większe niż pasma dla światłowodu 9 μm. Ten pierwszy fakt jest dobrze udokumentowany teoretycznie: przy regularnych profilach pobudzenia centralne dają większe pasma, gdyż pobudzają mniejszą liczbę modów (patrz część teoretyczna). Z kolei kombinacja modów LP_{01} i LP_{11}, charakterystyczna dla światłowodu o większej średnicy rdzenia, pobudza wiele grup modowych nieza-

leżnie od przesunięcia offsetowego. Powoduje to słabą zależność pasma od offsetu, a jednocześnie pasmo to jest mniejsze niż dla światłowodu prawdziwie jednomodowego (4 μm).



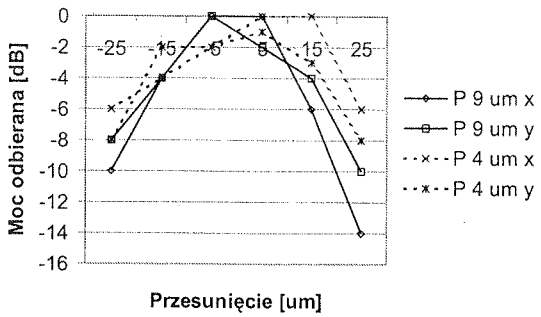
Rys. 26. Zależność mocy odbieranej dla gradientowego światłowodu kwarcowego (Φ = 62.5 μm) i długości L = 500 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm stykającymi się czołowo z badanym światłowodem przy zmiennym przesunięciu względem środka jego rdzenia. Długość fali λ = 780 nm

Fig. 26. Received power for GI silica fiber versus offset of the exciting single mode fiber



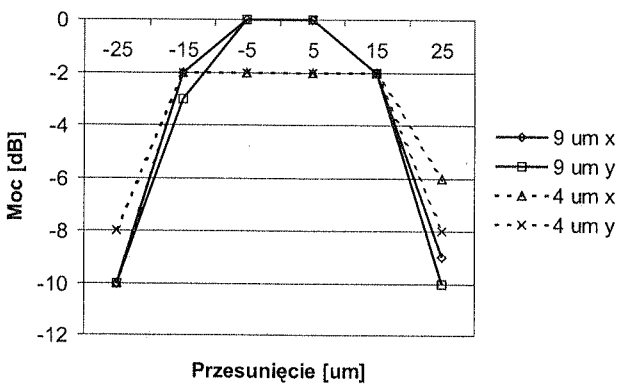
Rys. 27. Zależność pasma (MHz) dla gradientowego światłowodu kwarcowego (Φ = 62.5 μm) i długości L = 500 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm stykającymi się czołowo z badanym światłowodem przy zmiennym przesunięciu względem środka jego rdzenia. Długość fali λ = 780 nm

Fig. 27. Modal bandwidth of GI silica fiber versus offset of the exciting single mode fiber



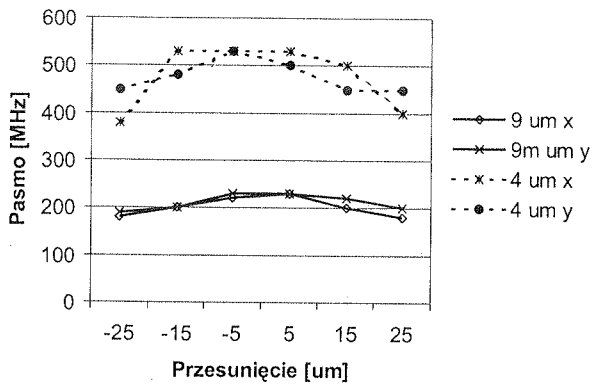
Rys. 28. Zależność mocy odbieranej dla gradientowego światłowodu kwarcowego (Φ = 50 μm) i długości L = 500 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm stykającymi się czołowo z badanym światłowodem przy zmiennym przesunięciu względem środka jego rdzenia. Długość fali λ = 780 nm

Fig. 28. Received power for GI silica fiber versus offset of the exciting single mode fiber



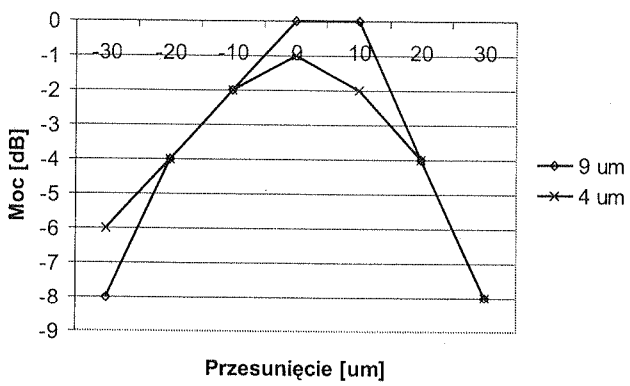
Rys. 29. Zależność mocy odbieranej dla gradientowego światłowodu kwarcowego (Φ = 50 μm) i długości L = 1000 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm stykającymi się czołowo z badanym światłowodem przy zmiennym przesunięciu względem środka jego rdzenia. Długość fali λ = 780 nm

Fig. 29. Received power for GI silica fiber versus offset of the exciting single mode fiber



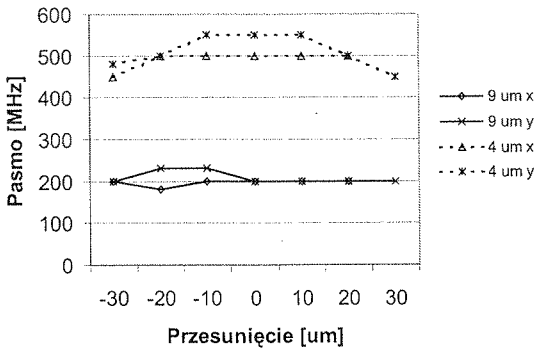
Rys. 30. Zależność pasma (MHz) dla gradientowego światłowodu kwarcowego (Φ = 50 μm) i długości L = 1000 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm stykającymi się czołowo z badanym światłowodem przy zmiennym przesunięciu względem środka jego rdzenia. Długość fali λ = 780 nm

Fig. 30. Modal bandwidth of GI silica fiber versus offset of the exciting single mode fiber



Rys. 31. Zależność mocy odbieranej dla gradientowego światłowodu kwarcowego (Φ = 62.5 μm) i długości L = 1000 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm stykającymi się czołowo z badanym światłowodem przy zmiennym przesunięciu względem środka jego rdzenia. Długość fali λ = 780 nm

Fig. 31. Received power for GI silica fiber versus offset of the exciting single mode fiber



Rys. 32. Zależność pasma (MHz) dla gradientowego światłowodu kwarcowego (Φ = 62.5 μm) i długości L = 1000 m. Pobudzenie światłowodami jednomodowymi o średnicach rdzenia 4 i 9 μm stykającymi się czołowo z badanym światłowodem przy zmiennym przesunięciu względem środka jego rdzenia. Długość fali λ = 780 nm

Fig. 32. Modal bandwidth of GI silica fiber versus offset of the exciting single mode fiber

5. PODSUMOWANIE

W pracy zaprezentowano model propagacji światła w światłowodach wielomodowych. Na podstawie tego modelu określono zależności teoretyczne podstawowych parametrów transmisyjnych światłowodu (pasmo i tłumienność) od sposobu jego pobudzenia. Zależności te zostały zweryfikowane doświadczalnie przy pomocy opracowanego w tym celu układu pomiarowego (patrz część pomiarowa niniejszej pracy). Badania eksperymentalne zostały przeprowadzone zarówno dla światłowodów kwarcowo-polimerowych (światłowodów z płaszczem polimerowym), jak również dla światłowodów kwarcowych.

Najważniejsze wnioski płynące z pracy są następujące:

1. Obliczenia numeryczne wykazują, iż wiązanie opóźności grupowej jedynie z numerem grupy modowej (modu złożonego) jest słuszne jedynie w przypadku regularnych profili współczynnika załamania o wykładniku zbliżonym do 2 (np. 1<g<3). Profile zaburzone oraz np. profil skokowy charakteryzują się sporą rozbieżnością opóźności grupowych modów należących do jednej grupy modowej.
2. Wpływ typu pobudzenia światłowodu na jego parametry transmisyjne jest tym istotniejszy im mniejszy jest współczynnik sprzężenia modów. Dla współczesnych światłowodów współczynniki sprzężenia modów są stosunkowo niewielkie, stąd wpływ warunków pobudzenia na parametry transmisyjne jest niezwykle istotny zwłaszcza dla krótkich światłowodowych linii transmisyjnych opartych na światłowodach wielomodowych (kilkadziesiąt... kilkaset metrów) spotykanych we współczesnych sieciach LAN (Gbit Ethernet, 10 Gbit Ethernet).

3. Pobudzanie światłowodów wielomodowych w pierwszym oknie transmisyjnym za pomocą światłowodu jednomodowego zgodnego z którymkolwiek z zaleceń G.652...G.655, których jednomodowość jest gwarantowana co najwyżej w drugim oknie transmisyjnym prowadzi do pobudzenia światłowodu kombinacją modów $LP_{01}+LP_{11}+$ ewnt. mody wyższych rzędów. Taki rodzaj sprzężenia prowadzi do pobudzenia wielu grup modowych różniących się opóźnościami grupowymi, a co za tym idzie zmniejszenia pasma modowego i zwiększenia tłumienności w porównaniu z innymi typami pobudzeń, co jest wyraźne zwłaszcza w przypadku słabego mieszania modów (małe d_0).

4. Niniejsza praca wskazuje na możliwość selektywnego pobudzania na wejściu różnych grup modów za pomocą przesunięć offsetowych (patrz rys. 1). Przy niewielkich wartościach współczynnika sprzężenia modów i niezbyt długich liniach te grupy modów praktycznie nie wymieniają się mocami. Stwarza to możliwość zastosowania zwielokrotnienia modowego, czyli transmisji różnych informacji przez różne grupy modów. Z rys. 1 wynika, że przez jednoczesne pobudzenie światłowodu wielomodowego kilkoma światłowodami jednomodowymi różniącymi się przesunięciem offsetowym istnieje możliwość pobudzenia kilku niezależnych zbiorów grup modowych niosących odrębne informacje. Na wyjściu linii światłowodowej te zbiory grup modowych muszą zostać rozdzielone do odrębnych fotoodbiorników na drodze filtracji przestrzennej. Taki system realizuje ideę zwielokrotnienia modowego. Szczegóły realizacji należy dopracować, ale sama idea wydaje się możliwa do wprowadzenia do praktyki.

Praca niniejsza została zrealizowana ze środków KBN w ramach grantu nr 4T11D01124.

6. BIBLIOGRAFIA

1. H.G. Unger: *Planar optical waveguides and fibres*, Oxford University Press, Oxford, 1977.
2. G. Yabre: *Comprehensive theory of dispersion in graded index optical fibres*, Journal of Lightwave Technology, vol. 18, no. 2, February 2000, pp. 166-177.
3. D. Marcuse: *Principles of optical fiber measurements*, Academic Press, New York, 1981.
4. L. Raddatz, I.H. White, D.G. Cunningham, M.C. Nowell: *An experimental and theoretical study of the offset launch technique for the enhancement of the bandwidth of multimode fiber links*, Journal of Lightwave Technology, vol. 16, no. 3, March 1998, pp. 324-331.
5. T. Ishigure, M. Kano, Y. Koike: *Which is a more serious factor to the bandwidth of GI POF: differential mode attenuation or mode coupling?*, Journal of Lightwave Technology, vol. 18, no. 7, July 2000, pp. 959-965.
6. R. Olshansky, D.B. Keck: *Pulse broadening in graded-index optical fibers*, Applied Optics, vol. 15, no. 2, February 1976, pp. 483-491.
7. R. Olshansky: *Mode coupling effects in graded index optical fibers*, Applied Optics, vol. 14, no. 4, April 1975, pp. 935-945.
8. D. Gloge: *Impulse response of clad optical multimode fibers*, Bell System Technical Journal, vol. 52, no. 6, 1973, pp. 801-816.
9. J.M. Senior: *Optical fiber communications. Principles and Practice.* Prentice Hall, New York, 1992.

10. M. Rousseau, L. Jeunhomme: *Numerical solutions of the coupled power equation in step-index optical fibers*, IEEE Transactions on Microwave Theory and Techniques, vol. 25, no. 7, July 1977, pp. 577-585.
11. Tele-fonika Kable SA: Fibre optic cables, www.tele-fonika.com.pl
12. K. Holejko, R. Nowak, R. Holejko: *Laser beam diagnostics with CCD cameras*, Instytut Telekomunikacji PW.
13. P. Pepljugorski, M.J. Hackert, J.S. Abott, S.E. Swanson, S.E. Golowich, A.J. Ritger, P. Kolesar, Y.C. Chen, P. Pleunis: *Development of system specification for laser-optimized 50-μm multimode fiber for multigigabit short-wavelength LANs*, Journal of Lightwave Technology, vol. 21, no. 5, May 2003, pp. 1256-1274.
14. T.P. Tanaka, S. Yamada: *Numerical solution of power flow in multimode W-type optical fibers*, Applied Optics, vol. 19, no. 10, 15 May 1980, pp. 1647-1652.
15. J.N. Kutz, J.A. Cox, D. Smith: *Mode mixing and power diffusion in multimode optical fibers*, Journal of Lightwave Technology, vol. 16, no. 7, July 1998, pp. 1195-1202.
16. ITU G.651: Characteristics of a 50/125 μm multimode graded index optical fibre cable, 1998.
17. T.A. Lenahan: *Calculation of modes in an optical fiber using the finite element method and EISPACK*, Bell System Technical Journal, vol. 62, no. 9, November 1983, pp. 2663-2694.
18. M. Webster, L. Raddatz, I.H. White, D.G. Cunningham: *A statistical analysis of conditioned launch for Gigabit Ethernet links using multimode fiber*, Journal of Lightwave Technology, vol. 17, no. 9, September 1999, pp. 1532-1541.
19. TIA-526-14-A: OFSTP-14 – Optical Power Loss Measurement of Installed Multimode Fiber Cable Plant (1998) (r2003).
20. P. Pepeljugoski, S.E. Golowich, A.J. Ritger, P. Kolesar, A. Risteski: *Modeling and simulation of next generation multimode fiber links*, Journal of Lightwave Technology, vol. 21, no. 5, May 2003, pp. 1242-1255.
21. A.K. Agarwal, G. Evers, U. Unrau: *New and simple method for selective mode group excitation in graded-index optical fibres*, Electronics Letters, vol. 19, no. 17, 18 August 1983, pp. 694-695.
22. N. Guan, K. Takenaga, S. Matsuo, K. Himeno: *Multimode fibers for compensating intermodal dispersion of graded-index multimode fibers*, Journal of Lightwave Technology, vol. 22, no. 7, July 2004, pp. 1714-1719.
23. K. Yamashita, Y. Koyamada, Y. Hatano: *Launching condition dependence of bandwidth in graded-index multimode fibers fabricated by MCVD or VAD method*, Journal of Lightwave Technology, vol. 3, no. 3, June 1985, pp. 601-607.
24. G.D. Smith: *Numerical solution of partial differential equations. Finite difference methods*, Oxford University Press, Oxford 1989.
25. M.J. Hackert: *Explanation of launch condition choice for GRIN multimode fiber attenuation and bandwidth measurements*, Journal of Lightwave Technology, vol. 10, no. 2, February 1992, pp. 125-129.
26. J.B. Schlager, M.J. Hacket, P. Peplјugorski, J. Gwinn: *Measurement for enhanced bandwidth performance over 62.5 μm multimode fiber in short- wavelength local area networks*, Journal of Lightwave Technology, vol. 21, no. 5, May 2003, pp. 1276-1284.
27. D. Harshbarger, P. Pondillo: *Multimode fiber for use with laser sources*, Corning Inc., WP4119, December 1999.
28. M.J. Hackert: *Characterizing multimode fiber bandwidth for Gigabit Ethernet applications*, Corning Inc., WP4062, August 2001.
29. Y. Koyamada, K. Yamashita: *Launching condition dependence of graded-index multimode fiber loss and bandwidth*, Journal of Lightwave Technology, vol. 6, no. 12, December 1988, pp. 1866-1871.

30. S.E. Golowich, W.A. Reed, A.J. Ritger: *A new modal power distribution measurement for high-speed short reach optical systems*, Journal of Lightwave Technology, vol. 22, no. 2, February 2004, pp. 457-468.
31. L. Raddatz, I.H. White, D.G. Cunningham, M.C. Nowell: *Influence of restricted mode excitation on bandwidth of multimode fiber links*, IEEE Photonics Technology Letters, vol. 10, no. 4, April 1998, pp. 534-536.

J. SIUZDAK, R. NOWAK, G. STĘPNIAK

INPUT OPTICAL EXCITATION AND TRANSMISSION PARAMETERS OF MULTIMODE OPTICAL FIBERS

Summary

The multimode (MM) optical fibers both silica and plastic are used nowadays in LANs for high bit rate data transmission (up to 10 Gbit/s) and distances of a few hundred meters. In this paper, the influence of the launching conditions on the MM fiber attenuation and bandwidth are investigated both theoretically and experimentally for various launch types. The partial differential equation of mode group power diffusion was solved numerically. Also the initial excitation of mode groups was found. The theory indicated that the influence of initial distribution of modes is the greatest when the MM fiber modes are not coupled. Restricted launches such as offset launches usually lead to the bandwidth increase for regular n(r) profiles as compared to the overfilled launches. The reason is that they excite fewer modes than the overfilled launch. The situation is reversed for the profiles with flaws (such as a central dip). Then some launch types may even lead to the bandwidth reduction below the values guaranteed by the manufacturer. In order to verify the theoretical results a series of measurements was taken. The measurement set up consisted of a tuned generator connected to an optical transmitter (with switched wavelengths 650 nm and 780 nm), and an optical receiver coupled to an electrical spectrum analyzer. An opto-mechanical mechanism made it possible to vary the launch offset by moving a position of the end of a single mode patchcord with regard to the MM fiber center. Two SM patchcords were used with diameters of 4-μm (truly monomode) and 9 μm (bimodal- LP_{01} and LP_{11} modes were observed experimentally). Two typical GI MM silica fibers with core diameters 50 μm and 62.5 μm, and numerical apertures of 0.2 and 0.275, respectively, were examined, as well as a silica core plastic cladding SI 200 μm fiber. For the 4 μm exciter the modal bandwidth is the greatest for central launches and decreases for offsets, whereas for the 9 μm exciter the bandwidth only slightly depends on offset. Furthermore, no matter the offset, the bandwidth for the 4 μm exciter is always greater than for the 9 μm one.

Keywords: data transmission, LAN, multimode optical fiber, bandwidth, attenuation, modal dispersion, semiconductor laser, LED

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2006, 52, z. 2, ss. 185–192

Photoelectric position transducer with the optimum compensation of the dynamic constant component of its signals

ZBIGNIEW SZCZEŚNIAK, Ph.D.

Kielce University of Technology, Faculty of Electrical Engineering, Automatic Control and Computer Science
Al. Tysiąclecia PP7, 25-314 Kielce, Poland
Zbigniew.Szczesniak@poczta.fm

Otrzymano 2005.09.05
Autoryzowano 2005.12.28

The development of specialised electronic systems makes it possible to process measurement signals that change very fast. This option affects the design of photoelectric position transducers. A tendency is observed to apply photoelectric position transducers of simpler design and thus of lower accuracy. The accuracy enhancement is obtained by electronic means. In order to process transducer measurement signals with high accuracy, it is necessary to eliminate signal DC component. The paper presents the method of dynamic compensation of the signal constant component that is generated while optic signals are converted into electric ones in the displacement measurement process. The method makes use of appropriately shaped measurement signals from three photoelectric systems. Signal shaping is possible owing to a proper design of the scanning distribution grating of optic signals in relation to the index grid of the measurement bar. Their defined mutual relations and provided diagrams of reading fields of the measurement gauge make it possible to obtain voltage signals that change in the sinusoidal manner and have appropriate phase shifts. The shaped signals give two sinusoidal voltage signals, symmetrical with respect to zero and shifted by 1/4 of the period in relation to each other.

The DC component compensation method presented in the paper accounts for the impact of environmental conditions on the component change. Such compensation ensures stable measurement step for systems further processing measurement signals. Generated signals are used in the systems multiplying signal frequencies when compared with the transducer output signals and then for the sake of measurement gauge motion direction discrimination.

Keywords: signal shaping in photoelectric position transducers, compensation of signal constant component in photoelectric position transducer, motion direction discrimination, photoelectric position transducer accuracy enhancement

1. INTRODUCTION

Incremental position transducers using photoelectric methods of measurement signal processing have been frequently used for position measurements.

At the output, transducers produce two signals whose period equals the period of the measurement gauge. The signals are shifted with respect to each another by 1/4 of the measurement gauge period. The measure of measurement accuracy becomes the measurement gauge period. The reduction in displacement quantization interval can be achieved by means of a more precise design of the measurement gauge or by appropriate processing of photoelectric transducer signals. The electronic method of signal processing is generally preferred due to the fact that transducer manufacturing technology is simpler. In electronic processing systems, it is important for the measurement signal not to contain a DC component. The application of dynamic methods of signal DC component compensation based on signal shaping in four photoelectric systems is presented in the literature [8], [9], [10].

The paper presents the method of shaping measurement signals with their DC component compensation, where the number of photoelectric systems of the measurement transducer has been reduced to three.

2. PRINCIPLE OF OPERATION OF PHOTOELECTRIC QUANTIZATION TRANSDUCERS

Photoelectric transducers are the most frequently used at present. They have a form of a disc (or a bar for linear displacement measurements) whose circumference is made of a sequence of transparent and opaque fields (Fig. 1.) or fields reflecting and absorbing light [2], [4], [7]. Linear transducers for length measurement contain either a bar, the length of which equals the length of position measurement, or a rotary disc for angle measurement. The disc (or the rule) modulates the light source beam incident on the photoelement in such a way that the number of pulses produced at its output is proportional to linear or angular displacement.

A scanning unit consists of a light source, condensing lenses, which help obtain a parallel light beam scanning the distribution grating with index grating, and also siliceous photovoltaic cells. When the scale is shifted with respect to the scanning unit, the scale lines coincide alternately with the lines or spaces in the index grating. Resultant periodic light intensity fluctuations are converted into electric signals by photovoltaic cells.

In linear transducers operating on reflected light, the measurement reference takes the form of a steel scale with a grid composed of highly reflective gold lines and light diffusing or absorbing spaces. When the scale is shifted, photovoltaic cells of the scanning unit produce periodic signals, similar to those generated by the glass scale.

Some linear transducers make use of light wave diffraction and interference [7]. The measurement reference is provided by phase grating with the pitch approx. 0.2μm

and the scanning unit comes in the form of corresponding diffraction grating on transparent glass. The transducers achieve high accuracy and account for very accurate measurement steps.

Carriers on which scale is placed, commonly in the form of a bar grid, play the most important role in measurement instruments [2], [7]. The scales, precisely manufactured in accordance with proper technologies, decide about the accuracy of measurement instruments. The scales placed on glass or glass ceramics (Fig. 1) are made by depositing an extremely thin chromium film, yielding the accuracy in micrometer range or even more precise.

The scales placed on steel carriers are made of highly reflective gold bars and mat spaces between them. They are characterised by high accuracy of bar location with respect to one another as well as considerable edge sharpness and are resistant to the action of mechanical or chemical agents. They are, however, sensitive to loads resulting from vibrations or impact. They also demonstrate precisely defined thermal behaviour. Changes in air pressure and humidity do not affect accuracy at all. The scale length – the measurement range for the scale on steel carriers (steel strips) amounts to 30m, whereas for the scale on glass to 3m.

For incremental measurement transducers, it is impractical to shift an object by a long distance to establish a base point of the object under consideration again.

The problem is solved by introducing length-coded reference marks to linear and circular scales [7]. The scale consists of both a grid and a parallel path of reference marks. The absolute position of each reference mark is coded by the distance between neighbouring reference marks, which changes in accordance with a defined formula.

3. PRINCIPLE OF PHOTOELECTRIC SIGNAL GENERATION IN QUANTIZATION TRANSDUCER

On the scanning distribution grating 2, three rectangular fields with diffraction measurement gratings 1 are placed. They are assigned to individual photovoltaic cells.

The parallel light beam produced by a light source and the lens is incident through the above mentioned grids on the scale of the glass gauge and from there on the photovoltaic cell. The grids have identical widths of light transparent and opaque fields (Fig. 1), which is a condition for photovoltaic cells to generate sinusoidal voltage signals [2], [7]. In order to obtain sufficient signal value, it is advantageous that the measurement field should consist of N periods T of the displacement division. Then, for a specified field lit with a radiation beam of intensity E_o, at a given efficiency of the optical system, we will obtain an effective beam φ getting to the photooptic system U_a, U_b, U_c.

We can form the beam incident on photoelements by placing scanning distribution grating with appropriate relative displacement in relation to the gauge index grid 2 of the transducer. The relative displacement for the photoelectric system U_a amounts to 90°. The displacement of the photoelectric system U_b is 1/3 of the period T in relation

to the photoelectric system U_a. Its position is defined by the dependence $N_1T + 1/3T$. On the other hand, the system U_c is shifted by 2/3 of the period T with respect to the system U_a. The position of this photoelectric system is given by the dependence $N_2T + 2/3T$. Exemplary positions of the reading fields are presented in Fig. 2. in which N, N1, N2 denote the number of periods T.

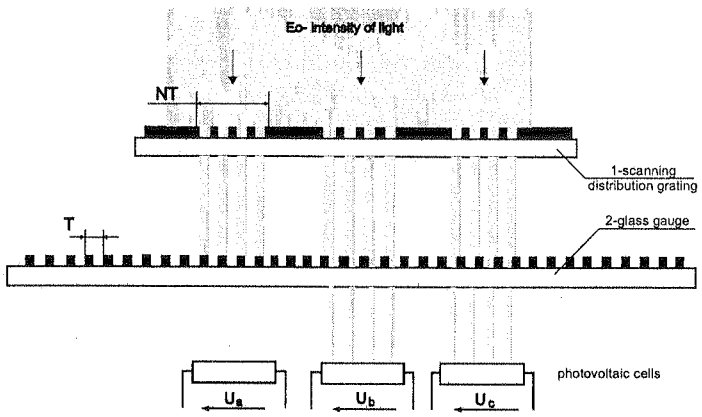


Fig. 1. Method of signal generation in photoelectric position transducer for three photoelectric systems E_o – intensity of light after going through condensing lenses, 1 – scanning distribution grating, 2 – glass gauge, U_a, U_b, U_c – photovoltaic cells

Rys. 1. Metoda wytwarzania sygnałów optoelektronicznego przetwornika położenia dla trzech układów optoelektroniki E_o – natężenie światła po przejściu przez soczewki skupiające, 1 – skanująca siatka rozdzielcza, 2 – liniał szklany, U_a, U_b, U_c – ogniwa fotoelektryczne

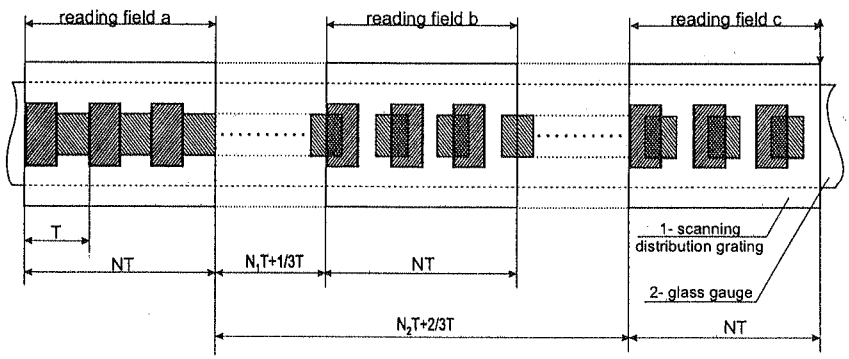


Fig. 2. Diagram of the reading fields of the measurement gauge for three photoelectric systems 1– scanning distribution grating, 2 – glass gauge

Rys. 2. Schemat pól odczytowych liniału pomiarowego dla trzech układów optoelektroniki 1– skanująca siatka rozdzielcza, 2 – liniał szklany

When the scanning distribution grating is in motion, photovoltaic cells generate sinusoidal voltage signals:

$$\begin{aligned} U_a &= A_1 \cos\varphi + A_o, \\ U_b &= A_1 \cos(\varphi + 120°) + A_o, \\ U_c &= A_1 \cos(\varphi + 240°) + A_o, \end{aligned} \tag{1}$$

where: $\varphi = 2\pi y/T$ in addition: A_1 – signal amplitude;
A_o – signal constant component;
y – displacement;
T – signal period.

Subtracting sinusoidal signals U_a and U_b, we get rid of the DC component and receive sinusoidal voltage signal that is symmetrical in relation to zero

$$U_1 = A\sin\varphi \tag{2}$$

The other signal

$$U_2 = A\sin\varphi \tag{3}$$

is obtained on the basis of the dependence

$$U_2 = k_1 U_a - k_2 (U_a + U_c) \tag{4}$$

where: signal amplitude $A = \sqrt{3}\ A_1$
coefficient $k_1 = 2\sqrt{3} / 3$
coefficient $k_2 = \sqrt{3} / 3$

The principle of signal processing is presented in Fig. 3, Fig. 4. The signal period equals the grid period of the scale T. The method presented in the paper for the compensation of the signal DC component takes into account its dynamic changes when the transducer operates under changeable environmental conditions (temperature, humidity, dustiness, etc.).

Generated signals sin, cos are used to determine motion direction. They can also be processed further in order to increase the accuracy of position measurement, Fig. 4.

Signals that do not contain a DC component make it possible to generate, in the system UZF, phase signals and then multiply the frequency of the signals in comparison with input signals [4], [5], [6], [8], [11]. The generated signals are rectangular and have pulse-duty factor 0.5. Due to the lack of constant component in the signal U1, U2 it is possible to directly receive from them rectangular signals of pulse-duty factor 0.5. Constant and stable pulse-duty factor ensures obtaining uniform measurement step. The latter was defined as a distance between two subsequent slopes of rectangular signals [1], [3], [4], [5].

The number of pulses generated from the slopes of the signals and counted in the system UZ in the course of measurement gauge shift is the measure of position.

Rectangular signals shifted in relation to each another by 1/4 of the period and of pulse-duty factor 0.5 in the system UZ also provide basis for the determination of the photoelectric transducer motion direction [2], [6].

The development of specialised analogue and digital systems makes it possible to process signals that change very fast. Therefore, measurement electronics almost does not set limits on signal frequency, mechanical limits are much more significant in this respect due to the dynamics of the measured object motion and its range.

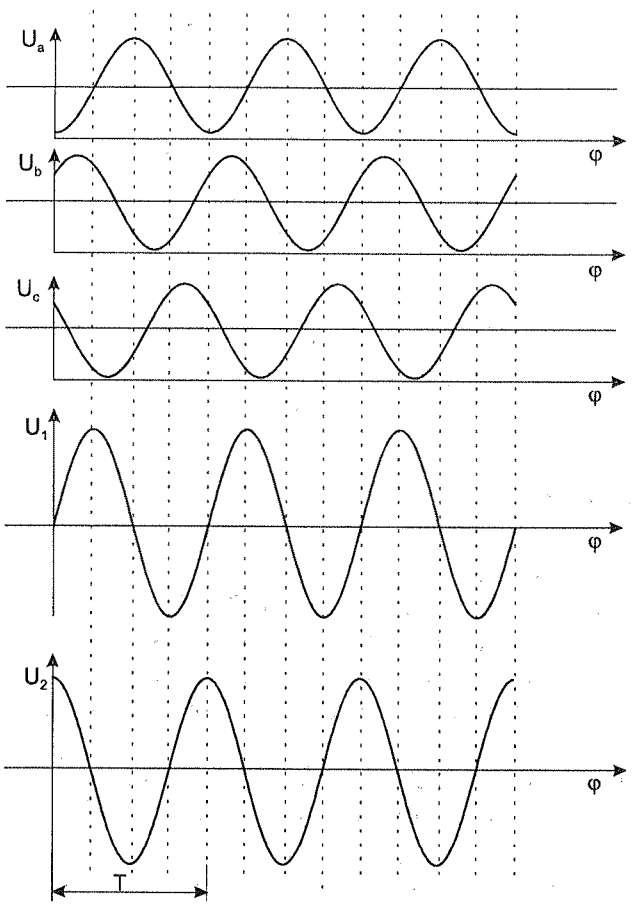


Fig. 3. Principle of signal generation in photoelectric quantization transducer for position measurement in three photoelectric systems

Rys. 3. Zasada wytwarzania sygnałów optoelektronicznego przetwornika kwantującego do pomiaru położenia dla trzech układów optoelektroniki

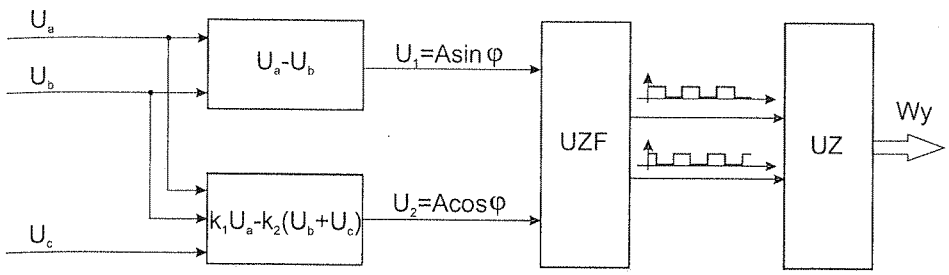


Fig. 4. Block diagram of signal processing for three photoelectric systems.
UZF – frequency multipling system, UZ – counting system

Rys. 4. Schemat blokowy przetwarzania sygnałów dla trzech układów optoelektroniki.
UZF – układ zwielokrotnienia częstotliwości, UZ – układ zliczania

4. CONCLUSIONS

1. The production of precise photoelectric position transducers requires advanced technologies, which involves high manufacture costs. There has been a general tendency towards applying photoelectric quantization transducers of simpler design and lower accuracy which could be enhanced by electronic means.

2. The development of specialised analogue and digital systems makes it possible to process signals that change very fast. The DC component should be eliminated in order to process those signals with high accuracy. The goal can be achieved with the method presented in the paper, which accounts for dynamic compensation of the signal DC component.

3. The algorithm presented in the paper for the measurement signals processing makes it possible to reduce the number of the transducer photoelectric subsystems thus contributing to its design optimisation.

5. REFERENCES

1. D. Hanselman: *Resolver signal requirements for high accuracy resolver-to-digital conversion.* IEEE Trans. on Idustrial Elekctr., Vol. 37, No. 6, 1990, pp. 556-561.
2. K. Holejko: *Precise Electronic Measurements of Distance and Angles* (in Polish). Warsaw, WNT 1981.
3. C. Christiansen, R. Battaiotto, D. Fandez, E. Tacconi: *Digital measurement of angular velocity for speed control.* IEEE Trans. on Idustrial Elekctr., Vol. 36, No. 1, 1989, pp. 79-83.
4. Zb. Szcześniak, K. Sikora, A. Pizoń, I. Smolewski: *Prototype construction of the electronic system for forging height measurement for the 30 MN press at Warsaw Steel Works and author supervising on construction and practical application of the system* (in Polish). Elaborated at Cracow University of Technology for Warsaw Steel Works. Stage II, 1989.
5. Zb. Szcześniak: *Method of Interpolation of Electric Signals of Photoelectric Transducers in Position Measurements* (in Polish). Electronics no 1. Warsaw 2005. pp. 74-76.

6. Zb. Szcześniak: *Electronic System for the Enhancement of Photoelectric Position Transducer Accuracy and its Motion Direction Discrimination* (in Polish). Kielce University of Technology Scientific Papers, Electrical Engineering no 42 Kielce 2005 pp. 333-342.
7. Zb. Szcześniak: *Precise Photoelectric Position Transducers* (in Polish) Kielce University of Technology Scientific Papers, Electrical Engineering no 42 Kielce 2005 pp. 343-351.
8. Zb. Szcześniak: *Electronic Module for Tenfold Accuracy Enhancement of Signal Processing in Photoelectric Position Transducer* (in Polish). Polish Academy of Sciences. Electronics and Telecommunications vol. 51 no 3/2005 pp. 439-447.
9. Zb. Szcześniak: *Method of Compensation of Measurement Signal Constant Component in Photoelectric Position Transducer* (in Polish). Measurements Automatics Control No 6 pp. 12-13 Warsaw 2005.
10. Zb. Szcześniak: *Method of Shaping Sinusoidaly Changeable Measurement Signals in Photoelectric Position Transducer* (in Polish). Kielce University of Technology Scientific Papers, Electrical Engineering no 43 pp. 193-198 Kielce 2005.
11. Эб. Шесьняк, М. Дорожовець: Метод помноження частоти вимірювального сигналу фотоелектричного перетворювача положення. Вісник НУ "Львівська Політехніка", "Комп'ютерні мережі та системи". N523. С.154-157. 2005 р.

ZB. SZCZEŚNIAK

OPTOELEKTRONICZNY PRZEWODNIK POŁOŻENIA Z OPTYMALNĄ KOMPENSACJĄ DYNAMICZNEJ SKŁADOWEJ STAŁEJ JEGO SYGNAŁÓW

Streszczenie

Rozwój specjalizowanych układów elektronicznych umożliwia przetwarzanie bardzo szybko zmieniających się sygnałów pomiarowych. Fakt ten wykorzystano w projektowaniu optoelektronicznych przetworników położenia. Zarysowuje się tendencja do stosowania optoelektronicznych przetworników położenia o prostszej budowie, a tym samym o mniejszej dokładności przetwarzania. Wagę uzyskania zwiększonej dokładności przenosi się na drogę elektroniczną. Aby przetwarzać sygnały pomiarowe przetwornika z dużą dokładnością należy z tych sygnałów wyeliminować składową stałą. W artykule przedstawiono metodę dynamicznej kompensacji składowej stałej sygnału, generowanej w trakcie przetwarzania sygnałów optycznych na elektryczne w procesie pomiaru przemieszczenia. Metoda ta wykorzystuje odpowiednio ukształtowane sygnały pomiarowe z trzech układów optoelektroniki. Kształtowanie sygnałów umożliwia odpowiednia konstrukcja skanującej siatki rozdzielczej sygnałów optycznych w stosunku do siatki indeksowej liniału pomiarowego. Określone ich wzajemne relacje oraz podane schematy pól odczytowych liniału pomiarowego umożliwiają uzyskanie sinusoidalnie zmiennych sygnałów napięciowych o odpowiednich przesunięciach fazowych. Z ukształtowanych sygnałów otrzymano dwa napięciowe sygnały sinusoidalne, symetryczne względem zera i przesunięte względem siebie o 1/4 okresu.

Przedstawiona metoda kompensacji składowej stałej uwzględnia wpływ warunków środowiskowych na zmianę tej składowej. Taka kompensacja zapewnia stały krok pomiarowy dla układów dalszego przetwarzania sygnałów pomiarowych. Wytworzone sygnały wykorzystywane są w układach zwielokrotnienia częstotliwości sygnałów w stosunku do sygnałów wyjściowych przetwornika, a następnie do określenia kierunku ruchu liniału pomiarowego.

Słowa kluczowe: kształtowanie sygnałów optoelektronicznego przetwornika położenia, kompensacja składowej stałej sygnału przetwornika optoelektronicznego, wyróżnienie kierunku ruchu, zwiększenie dokładności przetwornika optoelektronicznego

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2006, 52, z. 2, ss. 193–210

Modelowanie tłumienia propagacyjnego w systemach dostępowych, w warunkach zabudowy miejskiej

RYSZARD J. KATULSKI[1], ANDRZEJ KIEDROWSKI[2]

[1]*Politechnika Gdańska,*
Wydział Elektroniki, Telekomunikacji i Informatyki,
Katedra Systemów i Sieci Radiokomunikacyjnych,
ul. Narutowicza 11/12, 80-952 Gdańsk
rjkat@eti.pg.gda.pl
[2]*NETIA SA,*
ul. Poleczki 13, 02-822 Warszawa
Andrzej_Kiedrowski@netia.pl

Otrzymano 2005.02.10
Autoryzowano 2005.06.25

W pracy opisano nowy sposób wyznaczania tłumienia propagacyjnego w radiowych łączach dostępowych pracujących w terenie zabudowanym. Na wstępie dokonano eksperymentalnej oceny przydatności wybranych modeli stosowanych dotychczas do analiz propagacyjnych radiowych systemów dostępowych. Następnie scharakteryzowano główne czynniki mające wpływ na tłumienie sygnału radiowego w takich warunkach i na podstawie tego sklasyfikowano cztery statystycznie istotne kategorie środowiska propagacyjnego. Opisano wielowariantowy model empiryczny opracowany przy zastosowaniu analizy regresji wielowymiarowej, na podstawie pomiarowych badań propagacyjnych wykonanych w dużych aglomeracjach miejskich. Scharakteryzowano przydatność tego modelu do projektowania radiowych systemów dostępowych w mieście.

Słowa kluczowe: propagacja fal radiowych, modelowanie tłumienia propagacyjnego, radiowe systemy dostępowe

1. WSTĘP

Wyznaczanie tłumienia sygnału radiowego w warunkach zabudowy miejskiej, gdzie mamy do czynienia z dużym zróżnicowaniem infrastruktury środowiska propagacyjnego, stanowi złożony i nie do końca jednoznacznie rozwiązany problem. Co prawda istnieje wiele modeli przeznaczonych do wyznaczania tłumienia propagacyjnego w wa-

runkach miejskich, jednakże zostały one opracowane z przeznaczeniem do analizowania systemów radiokomunikacji ruchomej i z tego powodu nie uwzględniają specyfiki propagacyjnej występującej w stałych łączach dostępowych [1, 2, 3].

Zasadniczym czynnikiem różnicującym pod względem propagacyjnym pracę łącza radiowego w systemie dostępowym – szczególnie w mieście – od warunków pracy takiego łącza w systemie radiokomunikacji ruchomej jest lokalizacja anteny stacji abonenckiej, którą na ogół umieszcza się na dachu budynku, w którym znajduje się użytkownik systemu dostępowego. Natomiast w systemie radiokomunikacji ruchomej antena terminala użytkownika znajduje się zazwyczaj na niewielkiej wysokości nad podłożem trasy propagacji sygnału radiowego, tzn. w warunkach miejskich pomiędzy budynkami nad poziomem ulicy.

Z powyższego wynika odmienny mechanizm propagacyjny w obu wymienionych rodzajach łączy radiokomunikacyjnych. Przykładowo dla przypadku transmisji w „łączu w dół", tzn. od anteny stacji bazowej do anteny stacji użytkownika, o poziomie sygnału radiowego przy antenie terminala użytkownika ruchomego decyduje ta część tego sygnału, która ugnie się na krawędzi dachu i wniknie w obszar międzybudynkowy – zatem im zjawisko tego ugięcia jest większe tym jest lepiej. Natomiast w radiowym łączu dostępowym jest odwrotnie, tzn. tym jest lepiej im mniejsze jest wnikanie sygnału radiowego w przestrzeń około-budynkową, bo wówczas większy jest poziom sygnału dochodzącego do anteny stacji abonenckiej znajdującej się nad poziomem dachu. W podobny sposób można scharakteryzować zróżnicowanie mechanizmu propagacyjnego w obu rodzajach łączy dla przypadku transmisji w „łączu w górę".

Wszystko to sprawia, że zachodzi pilna potrzeba opracowania nowego podejścia do zagadnień propagacyjnych występujących w radiowych systemach dostępowych w mieście. Jest to tym bardziej pilne, że w praktyce mamy do czynienia z rosnącym zapotrzebowaniem na radiowy dostęp do usług telekomunikacyjnych, co rodzi popyt na analizy propagacyjne związane z projektowaniem takich systemów dostępowych. Należy w tym miejscu podkreślić, że ITU-R jak dotąd nie precyzuje zaleceń odnośnie wyznaczania tłumienia propagacyjnego w radiowych systemach dostępowych, pracujących szczególnie w warunkach miejskich. Zatem z konieczności, przy projektowaniu radiowych systemów dostępowych w mieście stosuje się modele propagacyjne przeznaczone do wyznaczania tłumienia sygnału radiowego w systemach radiokomunikacji ruchomej.

W pracy scharakteryzowano przydatność w powyższym względzie wybranych znanych modeli propagacyjnych oraz przedstawiono nowe analityczne ujęcie tłumienia propagacyjnego w radiowych łączach dostępowych pracujących w mieście, opracowane na podstawie obszernych badań pomiarowych tego tłumienia, wykonanych w dużych aglomeracjach miejskich na terenie całego kraju. Punktem wyjścia do tego było sklasyfikowanie środowiska propagacyjnego z punktu widzenia specyfiki pracy systemów dostępowych oraz określenie czynników istotnie wpływających na mechanizm propagacyjny w takich warunkach. Biorąc następnie pod uwagę te czynniki zdefiniowano w sposób wielowariantowy funkcję błędu dla wytypowanych charakterystycznych sytuacji propagacyjnych, przy użyciu której na drodze wielowymiarowej analizy regresji sfor-

mułowano nowe zależności analityczne opisujące tłumienie propagacyjne w badanych warunkach [4].

2. OCENA PRZYDATNOŚCI WYBRANYCH MODELI PROPAGACYJNYCH

Jak to zaznaczono na wstępie, do wyznaczania tłumienia propagacyjnego w warunkach miejskich opracowanych zostało wiele różnych modeli obliczeniowych. Do najbardziej użytecznych w tym względzie można zaliczyć następujące modele:

a) w warunkach LOS (Line-of-Sight)
- COST 231 Walfisha-Ikegamiego [5, 6, 7, 8, 9], oraz
- ITU-R P.1411 [10],

b) w warunkach NLOS (Non LOS)
- Okumury-Haty [11, 12],
- COST 231 Haty [9],
- Egli [13], oraz także
- COST 231 Walfisha-Ikegamiego [14, 15, 16].

Powyżej wymienione modele propagacyjne zostały zweryfikowane pod względem ich przydatności do analizowania tłumienia sygnału w radiowych systemach dostępowych w mieście. Przydatność ta została określona poprzez porównanie teoretycznych wyników obliczeń tłumienia propagacyjnego, wykonanych przy użyciu tych modeli, z wynikami jego pomiarów, które zostały przeprowadzone w rzeczywistych warunkach eksploatacyjnych, w dużych miastach na terenie całego kraju – tzn. w Gdańsku, Gdyni, Krakowie i Poznaniu. Łączna liczba tych pomiarów wyniosła 18924 różnych przypadków propagacyjnych, na co składało się 8556 przypadków tras propagacyjnych o charakterze LOS oraz 10368 przypadków o charakterze NLOS. Dodatkowo w ramach każdej z tych dwóch podgrup dokonano podziału ze względu na charakter zabudowy środowiska propagacyjnego, stosując kryterium w postaci dwóch kategorii tego środowiska. Tzn., kategoria I to śródmieścia miast, natomiast kategoria II to tereny podmiejskie. I tak, w podgrupie przypadków propagacyjnych typu LOS stwierdzono 4709 przypadków środowiska propagacyjnego kategorii I oraz 3847 przypadków kategorii II. Natomiast w podgrupie typu NLOS stwierdzono 7095 przypadków kategorii I oraz 3273 przypadki kategorii II. Pomiary te wykonano w paśmie 2,4 GHz, w którym zlokalizowane są główne zasoby częstotliwościowe przeznaczone do zastosowania w radiowych systemach dostępowych w mieście.

Zastosowaną miarą oceny przydatności były powszechnie używane błędy: średni oraz średni kwadratowy, który nazywany jest także standardowym błędem estymacji. Pierwszy z nich błąd średni *ME* (Mean Error) obliczano przy użyciu poniższej zależności:

$$ME = \frac{1}{N}\sum_{i=1}^{N}\left(L_{pom,i} - L_{obl,i}\right), \tag{1}$$

w której:
$L_{pom,i}$ – pomierzona wartość tłumienia propagacyjnego w *i-tym* przypadku propagacyjnym, w [dB],
$L_{obl,i}$ – obliczona przy użyciu określonego modelu wartość tłumienia propagacyjnego dla *i-tego* przypadku propagacyjnego, w [dB],
N – liczba przypadków pomiarowych, tzn. liczebność badanej próby.

Jak wiadomo, błąd średni służy do oceny stopnia rozproszenia wyników pomiarowych wokół wartości uzyskanych przy użyciu badanego modelu, z uwzględnieniem przy tym kierunku odchylenia. Natomiast standardowy błąd estymacji *SSE* (Standard Error of Estimation) wyznaczano przy użyciu poniższej zależności:

$$SSE = \sqrt{\frac{\sum_{i=1}^{N} \left(L_{pom,i} - L_{obl,i}\right)^2}{N-1}}. \tag{2}$$

Błąd ten oznacza odchylenie standardowe, które jest miarą jakości dopasowania badanego modelu do danych pomiarowych, określając rozrzut wartości pomierzonych wokół wartości teoretycznych.

Zestawienie obliczonych błędów dla wymienionych modeli propagacyjnych oraz – dla porównania – także dla modelu wolnej przestrzeni, łącznie dla wszystkich przypadków pomiarowych oraz z podziałem na wymienione podgrupy tych przypadków, przedstawiono w tabeli 1. Ponadto na rysunkach od 1 do 7 pokazano poglądowo rozrzuty wartości tłumienia sygnału radiowego pomierzone i obliczone przy użyciu tych modeli, w zależności od długości tras propagacyjnych, także dla wszystkich przypadków pomiarowych. Przy czym dla analizowanych łączy dostępowych o takich samych długościach otrzymano różne wyniki tłumienia propagacyjnego, co wynika ze zróżnicowania wartości parametrów technicznych tych łączy, przede wszystkim zaś z różnych wysokości zawieszenia anten.

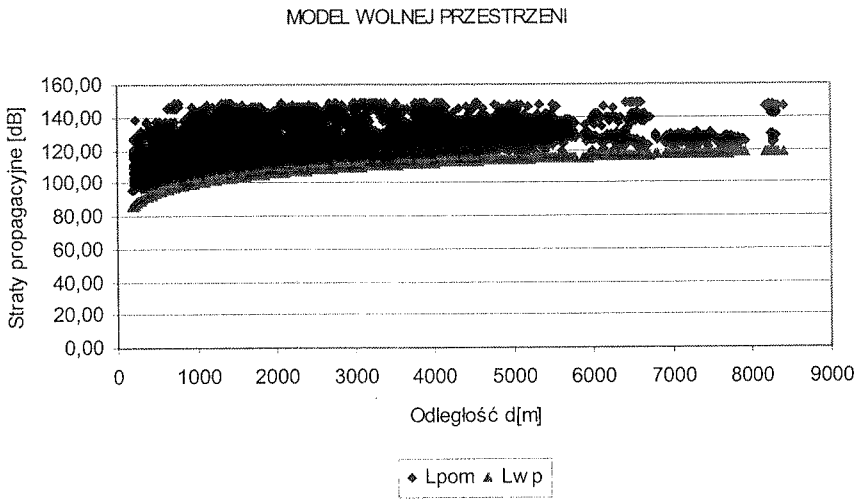
Analizując dane zestawione w tabeli I, można jednoznacznie stwierdzić, że wyniki obliczeń tłumienia propagacyjnego wykonane przy użyciu opisanych modeli propagacyjnych w przytłaczającej większości przypadków pomiarowych nie przystają do wartości tego tłumienia, które zostały uzyskane na drodze pomiarowej. Ma to miejsce zarówno dla poszczególnych przyjętych podgrup przypadków pomiarowych związanych z charakterystycznymi rodzajami środowiska propagacyjnego, jak i również dla przypadku badania przystawalności pełnej populacji pomiarowej do każdego z tych modeli. W odniesieniu do błędów średnich (*ME*), ich wartości wynoszą od conajmniej kilku do przeszło 20 dB. W podobny sposób rzecz się przedstawia w ocenie standardowych błędów estymacji (*SSE*).

Tablica 1

Zestawienie obliczonych błędów – znane modele propagacyjne, ME – błąd średni, SSE – błąd średni kwadratowy (standardowy błąd estymacji),
$L_{LOS,L}$ – graniczna minimalna wartość tłumienia propagacyjnego, wg modelu ITU R P.1411,
$L_{LOS,U}$ – graniczna maksymalna wartość tłumienia propagacyjnego, wg modelu ITU R P.1411

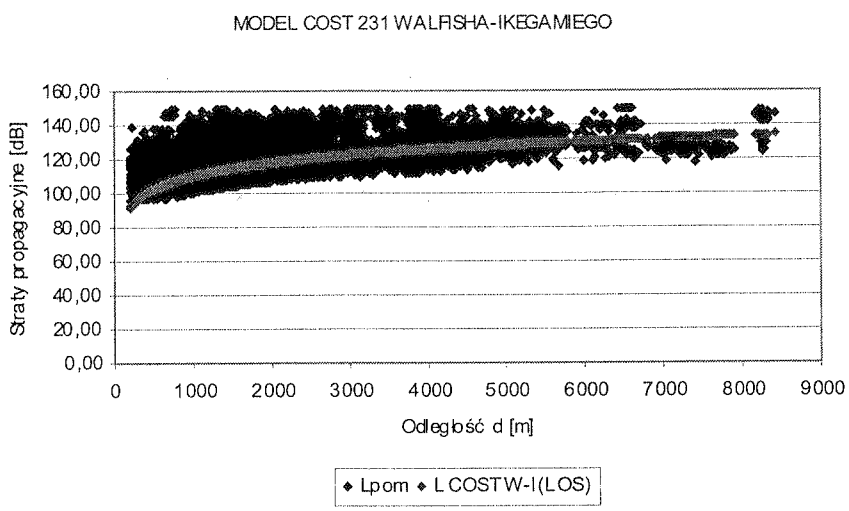
The calculated errors for the known propagation models, ME – mean error, SSE – mean square error (standard error of estimation),
$L_{LOS,L}$ – minimum value of the propagation loss (ITU R P.1411), $L_{LOS,U}$ – maximum value of the propagation loss (ITU R P.1411)

MODEL		WOLNEJ PRZESTRZENI		COST 231 WALFISH-IKEGAMI (LOS)		ITU-R P.1411 (LOS) - $L_{LOS,L}$		ITU-R P.1411 (LOS)- $L_{LOS,U}$		OKUMURA-HATA		COST 231 HATA		EGLI		COST 231 WALFISH-IKEGAMI (NLOS)	
WARUNKI BADAŃ		ME [dB]	SSE [dB]	ME [dB]	SSE [dB]	ME [dB]	SSE [dB]	ME [dB]	SSE [dB]	ME [dB]	SSE [dB]	ME [dB]	SSE [dB]	ME [dB]	SSE [dB]	ME [dB]	SSE [dB]
LOS	Kategoria I N=4709	9,185	9,830	-2,409	4,480	15,164	15,564	0,738	3,576	11,796	16,535	5,883	12,993	8,891	9,691	-3,576	9,161
	Kategoria II N=3847	9,074	9,657	-3,399	4,967	15,053	15,413	-0,074	3,341	4,092	18,936	1,180	18,526	5,449	7,156	-5,049	10,088
NLOS	Kategoria I N=7095	22,283	23,162	10,597	12,645	28,262	28,961	13,259	14,816	18,460	21,725	12,547	16,988	21,004	21,987	4,722	11,572
	Kategoria II N=3273	22,143	23,417	10,412	13,105	28,122	29,137	13,051	15,203	12,637	18,927	9,725	17,119	19,092	21,087	6,822	13,700
WSZYSTKIE POMIARY N=18924		16,330	18,424	4,499	9,990	22,309	23,309	7,412	11,311	12,760	19,242	7,974	16,178	14,514	17,099	1,044	11,147



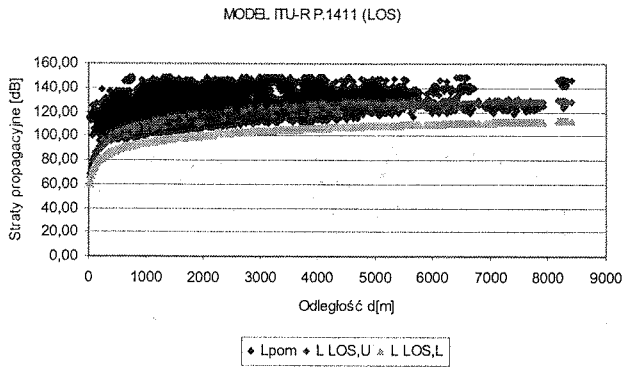
Rys. 1. Rozrzut wartości tłumienia propagacyjnego pomierzonych i obliczonych przy użyciu modelu wolnej przestrzeni - dla wszystkich danych pomiarowych, Lpom – wartości pomierzone, Lwp – wartości obliczone

Fig. 1. The scatter of the propagation loss values measured and calculated using the free space model – for all measurement data, Lpom – measured values, Lwp – calculated values



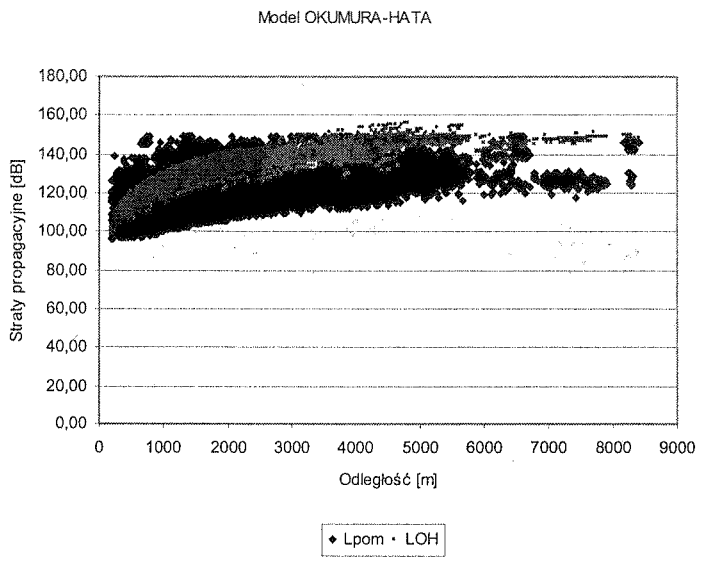
Rys. 2. Rozrzut wartości tłumienia propagacyjnego pomierzonych i obliczonych przy użyciu modelu COST 231 Walfisha-Ikegamiego dla warunków LOS – dla wszystkich danych pomiarowych, Lpom – wartości pomierzone, L COST W-I (LOS) – wartości obliczone

Fig. 2. The scatter of the propagation loss values measured and calculated using the COST 231 Walfish-Ikegami model for LOS conditions – for all measurement data, Lpom – measured values, L COST W-I (LOS) – calculated values



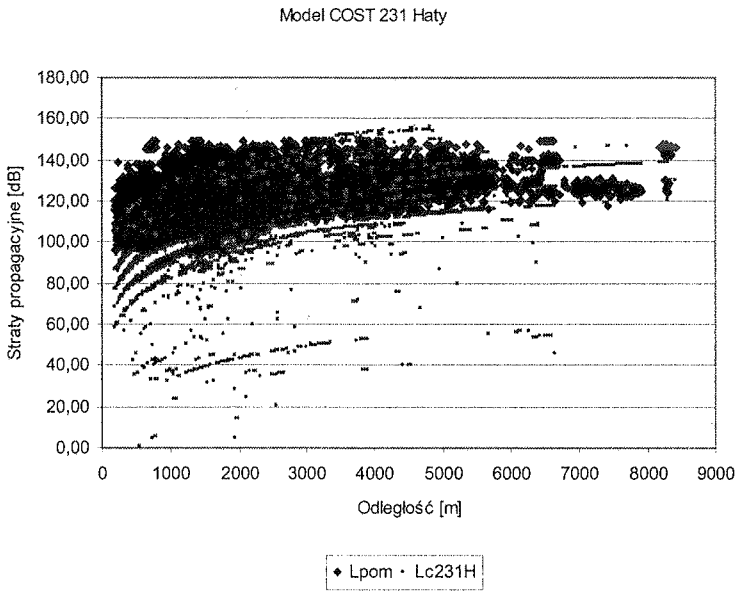
Rys. 3. Rozrzut wartości tłumienia propagacyjnego pomierzonych i obliczonych przy użyciu modelu ITU-R P.1411 (LOS) – dla wszystkich danych pomiarowych, Lpom – wartości pomierzone, L LOS,U – wartości obliczone wg wyrażenia na graniczną maksymalną wartość tłumienia propagacyjnego, L LOS, L – wartości obliczone wg wyrażenia na graniczną minimalną wartość tłumienia propagacyjnego

Fig. 3. The scatter of the propagation loss values measured and calculated using the ITU-R P.1411 (LOS) model – for all measurement data, Lpom – measured values, L LOS,U – calculated values using the formula for maximum propagation loss, L LOS, L – calculated values using the formula for minimum propagation loss



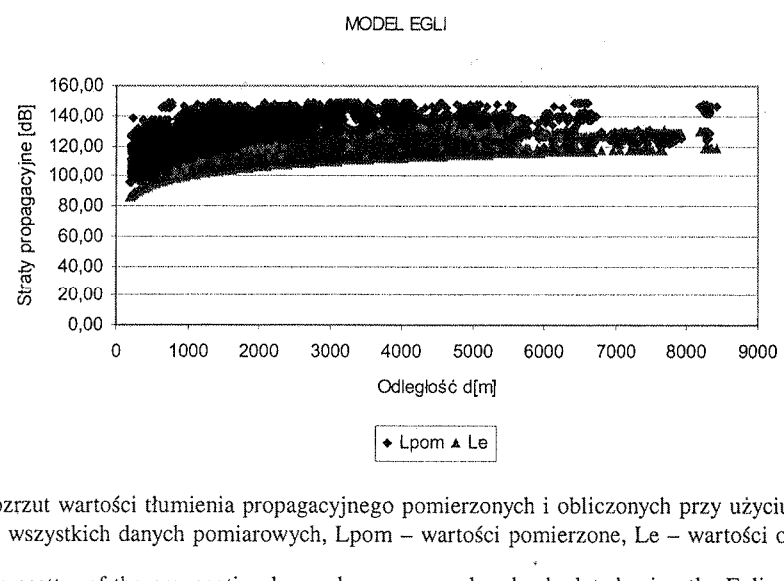
Rys. 4. Rozrzut wartości tłumienia propagacyjnego pomierzonych i obliczonych przy użyciu modelu Okumury-Haty – dla wszystkich danych pomiarowych, Lpom – wartości pomierzone, LOH – wartości obliczone

Fig. 4. The scatter of the propagation loss values measured and calculated using the Okumura-Hata model – for all measurement data, Lpom – measured values, LOH – calculated values



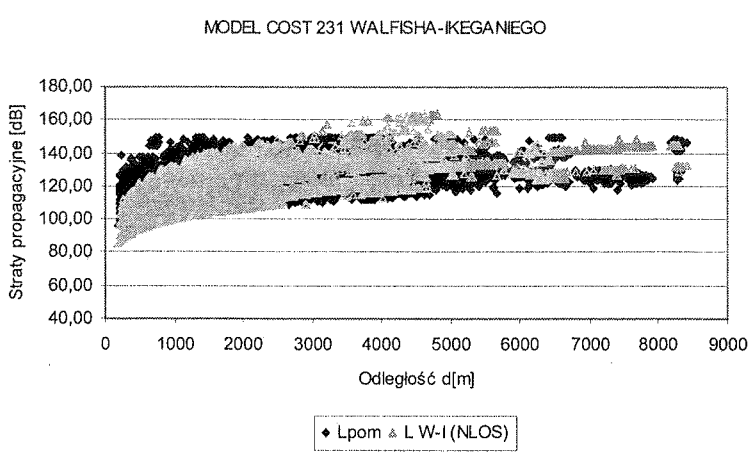
Rys. 5. Rozrzut wartości tłumienia propagacyjnego pomierzonych i obliczonych przy użyciu modelu COST 231 Haty – dla wszystkich danych pomiarowych, Lpom – wartości pomierzone, Lc231H – wartości obliczone

Fig. 5. The scatter of the propagation loss values measured and calculated using the COST 231 Hata model –for all measurement data, Lpom – measured values, Lc231H – calculated values



Rys. 6. Rozrzut wartości tłumienia propagacyjnego pomierzonych i obliczonych przy użyciu modelu Egli – dla wszystkich danych pomiarowych, Lpom – wartości pomierzone, Le – wartości obliczone

Fig. 6. The scatter of the propagation loss values measured and calculated using the Egli model – for all measurement data, Lpom – measured values, Le – calculated values



Rys. 7. Rozrzut wartości tłumienia propagacyjnego pomierzonych i obliczonych przy użyciu modelu COST 231 Walfisha-Ikegamiego dla warunków NLOS – dla wszystkich danych pomiarowych, Lpom – wartości pomierzone, L W-I (NLOS) – wartości obliczone

Fig. 7. The scatter of the propagation loss values measured and calculated using the COST 231 Walfish-Ikegami model for NLOS conditions – for all measurement data, Lpom – measured values, L W-I (NLOS) – calculated values

Na tym tle jedyny wyjątek stanowi porównanie wyników pomiarowych wykonanych w warunkach LOS – dla obu kategorii środowiska propagacyjnego, z wynikami obliczonymi przy użyciu modelu ITU-R P.1411 – według zależności na graniczną maksymalną wartość tłumienia propagacyjnego. Błędy średnie w tych przypadkach są mniejsze od jednego dB. Natomiast w przypadku kategorii II błąd ten wynosi tylko –0,074 dB, przy czym znak minus oznacza, że średnio rzecz oceniając model ten nieznacznie przeszacowuje wartości tłumienia propagacyjnego.

Ponadto inaczej w porównaniu do większości przypadków przedstawiają się wyniki zaprezentowane w tabeli I dla modelu COST 231 Haty – przeznaczonego do analizowania tras propagacyjnych o charakterze NLOS, który został zastosowany do obliczeń tłumienia w warunkach LOS w środowisku propagacyjnym kategorii II. Błąd średni w tym przypadku wynosi 1,18 dB, jednakże standardowy błąd estymacji przekracza 18 dB, co wskazuje na bardzo duży rozrzut wartości pomierzonych wokół wartości teoretycznych.

W podobny sposób jak powyżej można ocenić wyniki badań przystawalności modelu COST 231 Walfisha-Ikegamiego – w wariancie przeznaczonym do wyznaczania tłumienia propagacyjnego w warunkach NLOS, wykonane dla pełnej populacji pomiarowej. Błąd średni wówczas wyniósł 1,044 dB, zaś standardowy błąd estymacji 11,147 dB.

Podsumowując można stwierdzić, że opisane zbadane modele propagacyjne nie nadają się do analizowania tras propagacyjnych w radiowych systemach dostępowych, co dodatkowo wyraźnie pokazują wykresy na rysunkach 1–7. Wyjątek stanowi tutaj

wymieniony wariant modelu ITU-R P.1411, który może być stosowany do wyznaczania tłumienia propagacyjnego w warunkach LOS, szczególnie zaś wtedy, gdy mamy do czynienia z zabudową charakterystyczną dla obszarów podmiejskich.

3. CHARAKTERYSTYKA MIEJSKIEGO ŚRODOWISKA PROPAGACYJNEGO

Tereny miejskie charakteryzują się znacznym stopniem złożoności środowiska propagacyjnego fal radiowych, co w takich warunkach istotnie komplikuje i utrudnia realizację łączy radiokomunikacyjnych. Jednakże pomimo tego, przy projektowaniu radiowych systemów dostępowych w pierwszym podejściu dąży się do zapewnienia dwóch fundamentalnych wymagań propagacyjnych, tzn.:

- zapewnienia bezpośredniej widoczności obu anten radiowego łącza dostępowego, co określa się dobrze znaną skrótową nazwą LOS, oraz
- nieprzysłonięcia pierwszej strefy Fresnela.

Powyższe wymagania opisują stan idealny tzw. wolnej przestrzeni propagacyjnej, kiedy to tłumienie l_{wp} sygnału radiowego w dominującym stopniu zależy wyłącznie od długości d trasy propagacyjnej oraz od częstotliwości f sygnału radiowego, co dla przypomnienia opisuje poniższa zależność:

$$L_{wp}[dB] = 20\lg\left(\frac{4\pi d}{\lambda}\right) = 32{,}4 + 20\lg d[km] + 20\lg f[MHz], \tag{3}$$

w której λ oznacza długość fali wynikającą z częstotliwości sygnału radiowego.

Jednakże w rzeczywistych warunkach miejskich wymagania te są niejednokrotnie trudne do spełnienia, zwłaszcza drugie z nich, kiedy to projektant ze względów praktycznych zmuszony jest akceptować częściowe przysłonięcie pierwszej strefy Fresnela. Wiąże się to z dodatkowym, często znacznym wzrostem tłumienia propagacyjnego. Zatem najogólniej rzecz ujmując, całkowite tłumienie propagacyjne L_{prop} sygnału radiowego w takich warunkach można określić jak sumę tłumienia w wolnej przestrzeni wyrażonego zależnością (3) oraz tłumienia dodatkowego L_{dod}, co formalnie można wyrazić w sposób następujący:

$$L_{prop}[dB] = L_{wp}[dB] + L_{dod}[dB]. \tag{4}$$

Zatem, z punktu widzenia mechanizmu rozchodzenia się fal radiowych w systemach dostępowych w mieście, istotne są następujące okoliczności:

- spełnienie bądź też niespełnienie warunku LOS – bezpośredniej widoczności anten stacji głównej i abonenckiej,
- stopień przysłonięcia pierwszej strefy Fresnela, co przede wszystkim jest zależne od usytuowania anteny stacji abonenckiej względem zabudowy analizowanego obszaru, oraz
- rozproszenie fali radiowej na krawędziach dachów i jej wnikanie w przestrzeń międzybudynkową, przy czym na podstawie wykonanych badań proponuje się charak-

teryzować zabudowę obszaru propagacyjnego poprzez tzw. średnią wysokość tej zabudowy na obszarze objętym zasięgiem oddziaływania stacji głównej.

W rezultacie przeprowadzonej analizy zbadanych pomiarowo sytuacji propagacyjnych przyjęto, że tłumienie sygnału radiowego w radiowych łączach dostępowych w mieście zależy od następujących czynników:

- z oczywistych względów od długości l trasy propagacyjnej ponad dachami budynków – pomiędzy antenami łącza dostępowego, oraz od częstotliwości f sygnału radiowego,
- usytuowania obu anten łącza radiowego, tzn. od wysokości h_g zawieszenia anteny stacji głównej oraz od wysokości h_a położenia anteny stacji abonenckiej,
- względnego zawieszenia tych anten, odniesionego do średniej wysokości h_{sr} zabudowy terenu, tzn. odpowiednio $(h_g - h_{sr})$ oraz $(h_a - h_{sr})$,
- różnicy $(h_g - h_a)$ wysokości zawieszenia obu anten, oraz
- stopnia s przysłonięcia pierwszej strefy Fresnela, określonego zależnością:

$$s = \frac{4\left(\frac{h_g+h_a}{2} - h_{sr}\right)^2}{\lambda}. \tag{5}$$

Ponadto, dla potrzeb analizowania i projektowania radiowych systemów dostępowych w warunkach miejskich przyjęto wyróżniać cztery następujące charakterystyczne sytuacje propagacyjne, określone skrótowymi nazwami:

- LOS1, gdy zachodzi bezpośrednia widoczność pomiędzy antenami łącza, przy czym antena stacji abonenckiej jest usytuowana powyżej średniej wysokości zabudowy terenu,
- LOS2, gdy także spełniona jest bezpośrednia widoczność pomiędzy tymi antenami, jednakże antena stacji abonenckiej znajduje się poniżej średniej wysokości zabudowy terenu,
- NLOS1, gdy nie zachodzi bezpośrednia widoczność pomiędzy antenami łącza, przy czym antena stacji abonenckiej jest usytuowana powyżej średniej wysokości zabudowy terenu, oraz
- NLOS2, gdy także nie zachodzi bezpośrednia widoczność pomiędzy antenami łącza, jednakże antena stacji abonenckiej znajduje się poniżej średniej wysokości zabudowy terenu.

Opisany powyżej podział sytuacji propagacyjnych został zweryfikowany statystycznie na poziomie ufności wynoszącym 95%, co potwierdziło zasadność przyjęcia tego podziału.

Uzupełniając powyższą charakterystykę omawianego środowiska propagacyjnego, z punktu widzenia wpływu gęstości zabudowy infrastruktury miejskiej, przyjęto rozróżniać dwie następujące kategorie tego środowiska, tzn.:

- kategorię I oznaczającą śródmieścia miast, tj. obszary o zwartej i gęstej zabudowie w postaci przylegających do siebie względnie wysokich budynków, oraz
- kategorię II oznaczającą tereny podmiejskie o niezbyt gęstej zabudowie o charakterze osiedlowym, składającej się z bloków mieszkalnych oraz domów jednorodzinnych.

4. NOWY WIELOWARIANTOWY MODEL EMPIRYCZNY

4.1. GENEZA I SPOSÓB TWORZENIA MODELU

Nowe analityczne ujęcie tłumienia propagacyjnego w radiowym łączu dostępowym zostało opracowane przy użyciu wielowymiarowej analizy regresji z wieloma zmiennymi niezależnymi, na podstawie przeszło 18 tysięcy pomiarów wykonanych w rzeczywistych warunkach eksploatacyjnych.

Dla potrzeb tej analizy zdefiniowano wielowariantową funkcję błędu, w której jako zmienne niezależne przyjęto opisane czynniki propagacyjne. Czynniki te wpływają na mechanizm propagacji fal radiowych w łączach dostępowych i tym samym determinują tłumienie propagacyjne sygnału radiowego w takich łączach.

Wielowariantowość powyższej funkcji błędu wynika z przyjętej klasyfikacji sytuacji propagacyjnych, charakterystycznych dla radiowych systemów dostępowych w mieście.

W rezultacie poszczególne warianty analitycznego zapisu funkcji błędu ΔL przedstawiają się w następujący sposób [4]:

a) dla LOS1 ($h_a \geq h_{sr}$)

$$\begin{aligned} &\Delta L\left(l, h_g, h_a, h_g - h_{sr}, s\right) = L_{pom} - 32,4 - 20\lg f = \\ &= a\lg l + g\lg h_g + k\lg h_a + r_{gs}\lg\left(h_g - h_{sr}\right) + p\lg s + C, \end{aligned} \tag{6}$$

b) dla NLOS1 ($h_a \geq h_{sr}$)

$$\begin{aligned} &\Delta L\left(l, h_g, h_a, h_g - h_{sr}, \frac{h_g - h_a}{2}\right) = L_{pom} - 32,4 - 20\lg f = \\ &= a\lg l + g\lg h_g + k\lg h_a + r_{gs}\lg\left(h_g - h_{sr}\right) + r_{ga}\lg\frac{h_g - h_a}{2} + C, \end{aligned} \tag{7}$$

c) dla LOS2 oraz NLOS2 ($h_a \geq h_{sr}$)

$$\begin{aligned} &\Delta L\left(l, h_g, h_a, h_g - h_{sr}, h_{sr} - h_a, \frac{h_g - h_a}{2}\right) = L_{pom} - 32,4 - 20\lg f = \\ &= a\lg l + g\lg h_g + k\lg h_a + r_{gs}\lg\left(h_g - h_{sr}\right) + r_{sa}\lg(h_{sr} - h_a) + r_{ga}\lg\frac{h_g - h_a}{2} + C, \end{aligned} \tag{8}$$

przy czym L_{pom} wyraża pomierzone wartości tłumienia propagacyjnego, natomiast $a, g, k, p, r_{gs}, r_{ga}, r_{sa}$ są współczynnikami regresji – których wartości wyznaczono, zaś C oznacza wartość stałą.

4.2. ZAPIS ANALITYCZNY TŁUMIENIA PROPAGACYJNEGO

W wyniku przeprowadzonych badań opisanych wariantów funkcji błędu, wyznaczono wartości występujących w nich współczynników regresji dla poszczególnych sytuacji propagacyjnych. Otrzymano w ten sposób zapisy analityczne tłumienia propagacyjnego, zachodzącego w łączach dostępowych w określonych warunkach propagacyjnych w mieście.

4.2.1. W WARUNKACH LOS

Tłumienie propagacyjne sygnału w radiowych łączach dostępowych w mieście, pracujących w warunkach bezpośredniej widoczności anten, należy wyznaczać w następujący sposób:

a) w przypadkach, gdy wysokość h_a anteny stacji abonenckiej jest wyższa od średniej wysokości h_{sr} zabudowy terenu, na podstawie zapisu funkcji błędu (6), przy użyciu poniższej zależności:

$$L_{LOS1} = 23 + 20\lg f + 16,5\lg l + (22,1\lg h_g - 10,3\lg h_a) + 8,45\lg\left(h_g - h_{sr}\right) - 5,3\lg s, \tag{9}$$

przy czym występująca w wyrażeniu (6) stała $C = -9,4$, co po uwzględnieniu wartości 32,4 daje wynik końcowy: 32,4 - 9,4 = 23,

b) natomiast w przypadkach, gdy wysokość h_a anteny stacji abonenckiej jest niższa od średniej wysokości h_{sr} zabudowy terenu, na podstawie zapisu funkcji błędu (8), przy użyciu poniższej zależności:

$$\begin{aligned} L_{LOS2} &= 16,3 + 20\lg f + 18,1\lg l + (19,1\lg h_g - 6,7\lg h_a) + \\ &+ 12\lg\left(h_g - h_{sr}\right) + 0,6\lg(h_{sr} - h_a) - 16,2\lg\frac{h_g - h_a}{2}, \end{aligned} \tag{10}$$

przy czym występująca w wyrażeniu (8) stała $C = -16,1$, co po uwzględnieniu wartości 32,4 daje wynik końcowy: 32,4 - 16,1 = 16,3.

Występujące w powyższych wzorach parametry techniczne łącza radiowego wyrażone są w następujących jednostkach: $f[MHz]$ oraz $l[km]$, natomiast wszystkie wysokości $h_{()}$[m].

4.2.2. W WARUNKACH NLOS

Z kolei tłumienie propagacyjne sygnału w radiowych łączach dostępowych w mieście, pracujących w warunkach przysłonięcia linii bezpośredniej widoczności anten, należy wyznaczać w następujący sposób:

c) w przypadkach, gdy wysokość h_a anteny stacji abonenckiej jest wyższa od średniej wysokości h_{sr} zabudowy terenu, na podstawie zapisu funkcji błędu (7), przy użyciu poniższej zależności:

$$L_{NLOS1} = 108,6 + 20\lg f + 21,8\lg l + (-35\lg h_g + 16,6\lg h_a) + \\ - 26,3\lg\left(h_g - h_{sr}\right) + 23,9\lg\frac{h_g - h_a}{2}, \tag{11}$$

przy czym występująca w wyrażeniu (7) stała $C = 76,2$, co po uwzględnieniu wartości 32,4 daje wynik końcowy: 32,4 + 76,2 = 108,6,

d) zaś w przypadkach, gdy wysokość h_a anteny stacji abonenckiej jest niższa od średniej wysokości h_{sr} zabudowy terenu, na podstawie zapisu funkcji błędu (8), przy użyciu poniższej zależności:

$$L_{NLOS2} = 83,1 + 20\lg f + 15,8\lg l + (19,1\lg h_g - 20\lg h_a) + \\ - 47,2\lg\left(h_g - h_{sr}\right) + 0,3\lg(h_{sr} - h_a) + 34,4\lg\frac{h_g - h_a}{2}, \tag{12}$$

przy czym występująca w wyrażeniu (8) stała $C = 50,7$, co po uwzględnieniu wartości 32,4 daje wynik końcowy: 32,4 + 50,7 = 83,1.

Ponadto, podobnie jak w p. 4.2.1 parametry techniczne łącza radiowego wyrażone są w tych jednostkach.

4.3. WERYFIKACJA EKSPERYMENTALNA – OCENA PRZYDATNOŚCI

Ocenę przydatności opracowanego wielowariantowego modelu wyznaczania tłumienia propagacyjnego wykonano na podstawie zebranych obszernych danych pomiarowych. W ocenie tej zastosowano te same kryteria weryfikacji eksperymentalnej jak to miało miejsce w przy określaniu przydatności wybranych znanych modeli propagacyjnych. Tzn. także w omawianej weryfikacji miarą przydatności badanego modelu były błędy: średni (*ME*) i średni kwadratowy (*MSE*), pomiędzy wartościami tłumienia propagacyjnego pomierzonymi i obliczonymi przy użyciu opracowanego modelu. Zestawienie tak obliczonych błędów dla wszystkich możliwych charakterystycznych sytuacji propagacyjnych, uwzględniając przy tym także kategorie zabudowy środowiska propagacyjnego, oraz dla wszystkich przypadków pomiarowych łącznie, zestawiono w tabeli 2. Ponadto dla potrzeb porównania, na rys. 8 pokazano rozrzuty wartości tłumienia propagacyjnego pomierzone i obliczone przy użyciu nowego modelu – także dla wszystkich przypadków pomiarowych.

Jak widać, wartości tłumienia obliczone przy użyciu wielowariantowego modelu empirycznego dobrze przystają do danych pomiarowych i to niezależnie od typu sytuacji propagacyjnej oraz kategorii zabudowy środowiska propagacyjnego – świadczą o tym nie tylko znikome wartości obliczonych błędów średnich, lecz także standardowe błędy estymacji. Jest oczywiste, że rozrzut wartości pomierzonych wokół wartości teoretycznych, określony wartością standardowego błędu estymacji, jest relatywnie tym

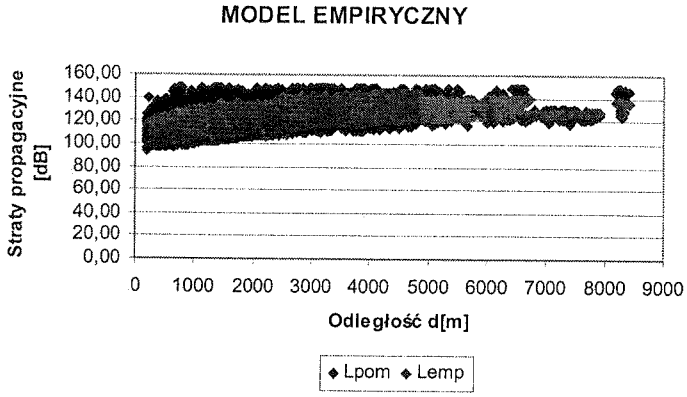
większy im rzeczywiste warunki propagacyjne bardziej odbiegają od idealnych warunków w wolnej przestrzeni. W omawianych warunkach propagacyjnych najlepiej pod tym względem jest w sytuacjach typu LOS – gdzie błąd ten jest rzędu 3 dB, zaś relatywnie najgorzej w sytuacjach typu NLOS – gdzie błąd ten jest dwukrotnie większy.

Tabela 2

Zestawienie obliczonych błędów – nowy model propagacyjny, ME – błąd średni, SSE – błąd średni kwadratowy (standardowy błąd estymacji)

The calculated errors for the new propagation model, ME – mean error, SSE – mean square error (standard error of estimation)

	LOS				NLOS				ŁĄCZNIE	
	LOS1		LOS2		NLOS1		NLOS2			
	ME	SEE	ME	SEE	ME	SEE	ME	SEE	ME	SEE
Kat. I	0,175	3,408	0,091	2,999	0,080	6,453	-0,200	5,335	0,019	4,829
Kat. II	-0,138	2,689	-0,167	3,077	-0,129	6,127	0,670	7,057	-0,032	5,039
ŁĄCZNIE	0,000	3,026	0,000	3,027	0,000	6,333	0,000	5,776	0,000	4,961
	ME=0,000 SEE = 3,026				ME=0,000 SEE=6,110					

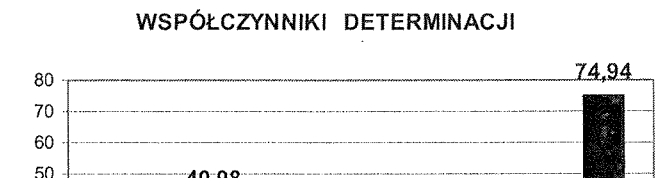


Rys. 8. Rozrzut wartości tłumienia propagacyjnego pomierzonych i obliczonych przy użyciu nowego modelu empirycznego, wg zależności (9–12) – dla wszystkich danych pomiarowych, Lpom – wartości pomierzone, Lemp – wartości obliczone

Fig. 8. The scatter of the propagation loss values measured and calculated using the new empirical model in accordance with the equations (9–12) – for all measurement data, Lpom – measured values, Lemp – calculated values

Na zakończenie, w celu porównania nowego modelu z dotychczas stosowanymi, na rys. 9 zebrano wartości współczynników determinacji dla poszczególnych modeli. Obrazuje to stopień ich dopasowania do rzeczywistych warunków propagacyjnych występujących w dostępowych łączach radiowych. Jak widać, najlepszy pod tym względem jest zaproponowany model wielowariantowy, który wyjaśnia wszystkie zbadane sytuacje propagacyjne w 75%, przy czym dla sytuacji typu LOS1 wartość tego współczynnika wyraźnie przekroczyła 80%. Na tym tle widać niedopasowanie dotychczas

stosowanych modeli propagacyjnych, czego przyczyna tkwi w genezie ich powstania, o czym jest mowa na wstępie tego artykułu.



Rys. 9. Współczynniki determinacji dla zbadanych modeli propagacyjnych – dla wszystkich danych pomiarowych, 1 – model wolnej przestrzeni, 2 – model COST 231 Walfisha-Ikegamiego, 3 – model Okumury-Haty, 4 – model COST 231 Haty, 5 – model Egli, 6 – nowy wielowariantowy model empiryczny

Fig. 9. The determination coefficients for the tested models – for all measurement data, 1 – free space model, 2 – COST 231 Walfish-Ikegami model, 3 – Okumura-Hata model, 4 – COST 231 Hata model, 5 – Egli model, 6 – new multi-variant empirical model

5. PODSUMOWANIE

Na podstawie analizy przeprowadzonych obszernych badań pomiarowych tłumienia propagacyjnego w radiowych systemach dostępowych w mieście, można jednoznacznie stwierdzić, że na wartość tego tłumienia obok tak oczywistych czynników jak częstotliwość i długość trasy propagacji, istotny wpływ mają także następujące okoliczności:

- spełnienie bądź też niespełnienie warunku bezpośredniej widoczności obu anten radiowego łącza dostępowego (LOS lub NLOS),
- usytuowanie anteny stacji abonenckiej względem zabudowy środowiska propagacyjnego, wyrażone poprzez wysokość zawieszenia tej anteny względem średniego poziomu zabudowy na obszarze objętym zasięgiem działania stacji bazowej obsługującej rozpatrywaną stację abonencką, oraz
- stopień przysłonięcia pierwszej strefy Fresnela, określony zależnością (5).

Mając to na uwadze opracowano nowe wielowariantowe ujęcie analityczne omawianego sposobu modelowania tłumienia propagacyjnego w radiowych systemach dostępowych w mieście. Model ten został zweryfikowany eksperymentalnie, co potwierdziło jego przydatność do analiz propagacyjnych niezbędnych w procedurze projektowania tych systemów.

6. BIBLIOGRAFIA

1. R.J. Katulski, A. Kiedrowski: *Wyznaczanie tłumienia propagacyjnego w systemie dostępowym w warunkach miejskich*, Przegląd Telekomunikacyjny i Wiadomości Telekomunikacyjne, rocznik LXXVI, nr 11, 2003, ss. 527-531.
2. R.J. Katulski, A. Kiedrowski: *Analiza projektowa radiowej stacji dostępowej w terenie zabudowanym*, Zeszyty Naukowe Wydziału ETI Politechniki Gdańskiej, seria: Technologie Informacyjne, nr 1, 2003, ss. 281-287.
3. R.J. Katulski, W. Pawłowski, A. Kiedrowski: *Testing of selected propagation models for radio access systems EMC analysis*, Proc. of the XIVth Int. Wroclaw Symp. on Electromagnetic Compatibility, Wrocław, 2002, pp. 43-46.
4. A. Kiedrowski: *Specyfika propagacji fal radiowych w systemie dostępowym w warunkach miejskich*, Praca doktorska, Instytut Łączności w Warszawie, 2004.
5. J. Walfish, H.L. Bertoni: *A theoretical model of UHF propagation in urban environments*, IEEE Trans. on Antennas and Propagation, vol. AP-36, no. 12, 1988, pp. 1788-1796.
6. F. Ikegami, S. Yoshida: *Analysis of multi-path propagation structure in urban mobile radio environments*, IEEE Trans. on Antennas and Propagation, vol. AP-28, no. 4, 1980, pp. 531-537.
7. F. Ikegami, S. Yoshida, T. Takeuchi, M. Umehira: *Propagation factors controlling mean field strength on urban streets*, IEEE Trans. on Antennas and Propagation, vol. AP-32, no. 8, 1984, pp. 822-829.
8. D.J. Cichon, T. Kurner, W. Wiesbeck: *Modellierung der wellenausbrietung in urbanen gelande*, Frequenz, vol. 47, no. 1-2, 1993, pp. 2-10.
9. COST 231 TD (90) 119 Rev. 1: *Urban transmission loss models for mobile radio in the 900- and 1800- MHz bands*, Hague – Netherlands, September 1991.
10. ITU-R, Doc. 3K/15-E: *Propagation data and prediction methods for the planning of short-range outdoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz*, 1999.
11. Y. Okumura, E. Ohmori, T. Kawano, K. Fukuda: *Field strength and its variability in VHF and UHF land-mobile service*, Review of the Electrical Communication Laboratory, vol. 16, no. 9-10, 1968, pp. 825-873.
12. M. Hata: *Empirical formula for propagation loss in land mobile radio services*, IEEE Trans. on Vehicular Technology, vol. VT-29, no. 3, 1980, pp. 317-325.
13. S.R. Saunders, F.R. Bonar: *Explicit multiple-building diffraction attenuation function for mobile radio wave propagation*, Electronic Letters, vol. 27, no. 14, 1991, pp. 1267-1277.
14. S.R. Saunders, F.R. Bonar: *Prediction of mobile radio wave propagation over buildings of irregular heights and spacings*, IEEE Trans. on Antennas and Propagation, vol. AP-42, no. 2, 1994, pp. 137-144.
15. F. Ikegami, T. Takeuchi, S. Yoshida: *Theoretical prediction of mean field strenght for urban mobile radio*, IEEE Trans. on Antennas and Propagation, vol. AP-39, no. 3, 1991, pp. 299-302.
16. N. Cardona, P. Moller, F. Alonso: *Applicability of Walfish-type urban propagation models*, Electronics Letters, vol. 31, no. 23, 1995, pp. 1971-1973.

R.J. KATULSKI, A. KIEDROWSKI

MODELLING OF THE PROPAGATION LOSS IN URBAN RADIO ACCESS SYSTEMS

Summary

A problem of the propagation loss calculations in radio access systems working in urban areas has been presented. There are some well known propagation models describing the loss in urban environment, but these models are accurate for mobile radio links, not for fixed access links.

First, these models are verified by measurements way in urban radio access systems. The presented results of the verification show than these models are not suitable for propagation analysis of the access systems.

Next, the main propagation mechanism elements existed in the radio access systems have been characterized. The four typical types of the urban propagation environment, called LOS1, LOS2, NLOS1 and NLOS2, have been classified. Taking into account the classification the new empirical propagation model is proposed. The model has form of four mathematical expressions for calculation of the propagation loss in radio access systems, due to the types of the propagation environment. Based on propagation measurement investigations made in different urban conditions, the proposed expressions by way multidimensional regression analysis have been obtained.

These measurements cases were divided in two groups due to categories of the urban areas, i.e.:

- category I of the urban areas, with compact and high buildings in cities centers, and
- category II of the areas, with low buildings in residential parts of the cities.

The proposed new model of the propagation loss was empirically tested for its usefulness for design procedure of the radio access systems in urban areas. The obtained results presented in figs. 8 and 9, and in tab. II, show good usefulness of the model for propagation analysis of the access systems.

Keywords: propagation of the radio waves, propagation loss modeling, radio access systems

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2006, 52, z. 2, ss. 211–228

Forward link power allocation in multiservice DS-CDMA wireless systems

RAFAŁ KRENZ

Poznań University of Technology
Piotrowo 3A, 60-965 Poznań, Poland
rkrenz@et.put.poznan.pl

Otrzymano 2005.04.11
Autoryzowano 2005.07.25

A reasonable forward link power allocation in multiservice DS-CDMA wireless systems is an important issue due to wide range of bandwidth required from different users. This paper considers several power allocation algorithms, which take into account various Quality of Service indicators e.g. the maximum probability of blocking and the maximum waiting time. Algorithms for both circuit-switched and packet data transmission mixed with voice traffic are considered. The SIR requirements and base station power constraints are analysed and the results of computer simulation of the proposed algorithms are presented.

Keywords: CDMA, forward link, resource allocation, circuit-switched, packet-switched

1. INTRODUCTION

Code Division Multiple Access (CDMA) has been recognized as a very attractive technique for wireless communications for some time now. Its advantages over other multiple access schemes include higher spectral reuse efficiency, greater immunity to multipath fading, gradual overload capability (soft capacity limit), simple exploitation of sectorisation and voice activity and more robust hand-off procedures [1]. First commercial wireless communication system based on CDMA was developed by Qualcomm and accepted as a North American second generation cellular system standard (cdmaOne or IS-95) [2]. Wideband CDMA has been selected as a radio interface for third generation systems both in North America (Cdma2000) and Europe (UMTS), as well as Japan [3], [4]. The introduction of these systems have been delayed a little due to the economic reasons, however, now they are fully operational.

Third generation wireless communication systems enable not only voice communication but also data transfer and multimedia applications. Second generation wireless communication systems, such as cdmaOne or GSM, allow data communication, but its bit rate is quite limited. To enable efficient wireless Internet access, video transfer, etc., much higher rates are necessary. A wide range of bit rates (up to 2 Mbit/s) can be readily implemented in a CDMA system by changing processing gain or assigning multiple codes to a given user [5], [6].

Many papers have been devoted to the reverse link (mobile to base) of the CDMA wireless systems (e.g. [7], [8], [9]). In second and third generation CDMA systems it is usually the forward link (from base to mobile), which is limiting the capacity of the system, even when throughput requirements on both links are symmetric (as in the case of voice service). That situation becomes even worse in multiservice CDMA systems, where the links are often asymmetric, with much higher throughput demands on forward links, which may often be required to carry large files or data streams at high transmission rates. The required transmitted power is in general proportional to the requested rate and may be very high for users located near the cell boundary.

This paper considers the forward link (base to mobile) of a multiservice CDMA wireless communication system only. It addresses different aspects of the base station power allocation and call management. The paper is organized as follows. In Section II the system structure and propagation models used throughout the paper are introduced. Analysis of the signal-to-interference-ratio (SIR) requirements and base station power constraints are given in Section III. Section IV and V presents the proposed power allocation algorithms for circuit-switched and packet data transmission, respectively. Results based on computer simulations are discussed in Section VI. Finally, Section VII summarizes the results of the work and discusses possible directions of future research.

2. PROPAGATION AND SYSTEM MODELS

In a large area environment (i.e. macrocell environment) there are several sources of variations of the received power. In this work the single slope path loss model [10] is used to describe the mean path loss and the large-scale fading (shadowing caused by obstacles in the propagation path between the mobile and the base station) is modeled as log-normal with spatial correlation of the shadows [12]. The third source of received power variations is multipath fading (small scale fading). It is not explicitly accounted for in this work.

A uniform, hexagonal cell layout with a base station at the center of every cell is considered throughout this work. The analysis assumes one ring of surrounding cells (interfering base stations). Mobile users are uniformly distributed in the home cell and they are moving at a certain speed v. The power transmitted from base stations surrounding the home cell is assumed to be equal (equally loaded cells). However,

these cells are not necessarily fully loaded. Finally, it is assumed that the system is interference limited and that background noise is negligible.

Both voice and data call (packet) arrivals are modeled as independent Poisson processes with arrival rates λ_v calls/s and λ_d packets/s, respectively. The service time for voice calls is exponentially distributed with mean $1/\mu_v$ s/call. For circuit-switched data transmission the service time is exponentially distributed with mean $1/\mu_d$ s/call, while for packet data transmission the mean packet size of δ_d kB/packet is specified. Each admitted user exhibits an ON/OFF activity pattern with activity factors β_v and β_d, respectively (with $\beta_d = 100\%$ for packet data transmission). Active voice users transmit at the basic bit rate, R_o, while data traffic users require transmission rates $R_i = r_i R_0$ with a probability p_i, where $i = 1, ..., M$ and M is the number of data rate classes defined in the system and r_i is associated with i-th data rate class.

3. BASE STATION POWER ALLOCATION IN MULTISERVICE CDMA

In principle, if different data users in a given cell request different data rates, the allocated power should be proportional to the requested rate. Additionally, the average power allocated to more distant users needs to be higher, because of the more severe inter-cell interference. Since a base station's total power available to all mobiles active in its cell is limited, the necessary power that needs to be allocated to a user may not be available at the time of an access request. Such event is more likely for a more distant user requesting a high data rate. Therefore, it is crucial to develop base station power allocation algorithms that ensure required quality of service (QoS).

The received average signal-to-interference-ratio in a multicell CDMA system with bandwidth W and transmission rate R can be expressed as:

$$SIR = \frac{W}{R} \cdot \frac{P_{req}}{(I_{\text{int}})_{in} + (I_{\text{int}})_{out}} \tag{1}$$

where P_{req} is the required average power of the received signal, $(I_{int})_{in}$ is the interference from within the home cell and $(I_{int})_{out}$ is the interference from the surrounding cells.

In order to obtain a given quality of transmission (e.g. $FER < 1\%$) a certain minimum value of $SIR \triangleq \varepsilon_{req}$ is required at the mobile station. Assuming that i-th base station radiates total power $P_{Ti} \triangleq k_i P_{\max}$, (1) may be rewritten for the j-th mobile user as:

$$\varepsilon_{req(j)} \leqslant \frac{W}{R_o} \cdot \frac{\alpha_{0j}^2 P_j}{\alpha_{0j}^2 \left(P_{max} - P_j\right) + \sum_{i=1}^{K_c - 1} \alpha_{ij}^2 P_{Ti}} \tag{2}$$

where P_j denotes the power which needs to be transmitted by the base station to the j-th mobile station, when data is transmitted at the basic rate R_o, α_{ij} is the overall path

gain from the i-th base station (with 0 denoting home base station) to the j-th mobile station and K_c is the number of interfering base stations.

The number of users K_u connected to every base station is limited by its maximum radiated power and must satisfy the following inequality:

$$\sum_{j=1}^{K_u} \beta_j P_j r_j \leqslant P_{max} \tag{3}$$

with β_j and r_j denoting activity factor and rate factor for the j-th user, respectively. If (3) is not satisfied the incoming calls are blocked due to insufficient power at the base station.

Solving (2) for P_j gives:

$$P_j = \frac{\frac{R}{W}\varepsilon_{req(j)} P_{max}}{1 + \frac{R}{W}\varepsilon_{req(j)}} \left[1 + \alpha_{0j}^{-2} \sum_{i=1}^{K_c - 1} k_i \alpha_{ij}^2 \right] \tag{4}$$

Finally, substitution of (4) into (3) yields the call blocking condition:

$$\sum_{j=1}^{K_u} \beta_j r_j \left(1 + \alpha_{0j}^{-2} \sum_{i=1}^{K_c - 1} k_i \alpha_{ij}^2 \right) > \frac{W}{R} \frac{1}{\varepsilon_{req}} + 1 \tag{5}$$

where equal ε_{req} has been assumed for all mobile stations. Equation (5) may be readily evaluated by computer simulation of respective random variables (r_j, α_{0j}, α_{ij}) to obtain probability of blocking and system capacity for a wide range of propagation models and other system parameters.

4. POWER ALLOCATION ALGORITMS FOR MIXED VOICE/CIRCUIT-SWITCHED DATA TRAFFIC

As mentioned earlier a base station transmitted power allocation algorithm must be implemented in the system to ensure required QoS. It is usually defined by an acceptable probability of blocking, which may be set to different values for voice and data calls (e.g. 2% and 5%). However, due to higher power requirements for data users it may be necessary to allow such users to wait for access to the system until the necessary power is available. If the incoming calls are queued, the other QoS indicators may be the average and maximum waiting times, specified separately for each class for certain probability of blocking.

For streamed data services it may be possible to reduce the required transmission rate if the user accepts a lower quality of service (e.g. lower image resolution or lower sound quality). This of course increases cell capacity and creates another QoS indicator, the probability of call degradation. The aforementioned QoS indicators have

been used in the proposed power allocation algorithms, which are described in the following paragraphs.

The first and simplest algorithm, denoted as CSD1, is shown in Fig. 1. The incoming voice and data calls are accepted as long as the necessary power is available at the home base station. If the incoming call requires the amount of power, which is not available at the time of the request, it is discarded and a blocking event is declared. Due to movement of mobile users a call dropping may occur if the power balance becomes insufficient. The QoS indicator is the probability of blocking, which is specified separately for voice and data calls.

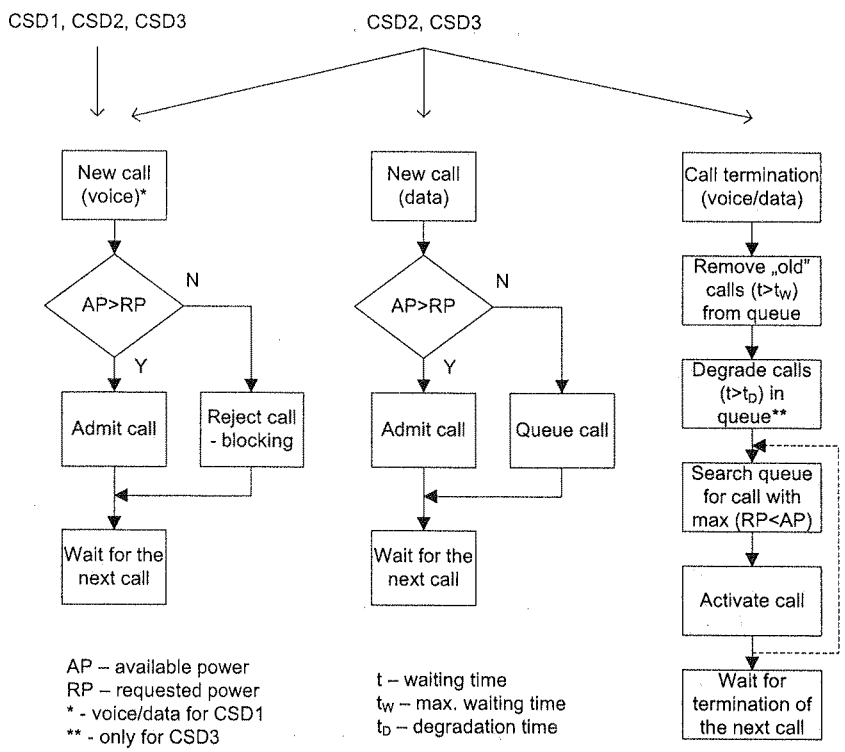


Fig. 1. Flowchart for algorithms CSD1-CSD3

Rys. 1. Algorytmy CSD1-CSD3

In algorithm CSD2, the incoming voice and data calls are processed separately (Fig. 1). If the new voice call requires the amount of power, which is not available at the time of the request, it is discarded and the blocking event is declared. However, the strategy for data calls is different: if the incoming data call can not be accepted immediately, it is put in a queue. For each class of data calls a maximum waiting time $t_W^{(class)}$ is specified. After termination of any active call the queue is searched and the call requiring the most (but available) power is activated. The search is repeated until

power balance available at the base station is not sufficient to handle any more queued calls. During the search the waiting time of each call is examined. If the waiting time exceeds the limit, $t_W^{(class)}$, for a given class, the call is removed from the queue and the blocking event is declared. If the available power becomes insufficient due to the active user movement, calls may be dropped. The QoS indicators are: probability of blocking, maximum and average waiting times for each class.

Algorithm CSD3 is an extended version of CSD2. It adds the feature of call degradation (Fig. 1). For each class of data calls a degradation time $t_D^{(class)}$ is specified. After removing "old" calls the queue is searched again and the waiting times are examined. If the waiting time exceeds $t_D^{(class)}$, the call is degraded, i.e. moved to a lower class (with lower transmission rate). Similarly, if the power balance becomes insufficient due to movement of the active users, calls may be degraded. In both cases only one degradation to the next lower class is allowed per call request. The specific QoS indicator for this algorithm is the probability of call degradation for each class.

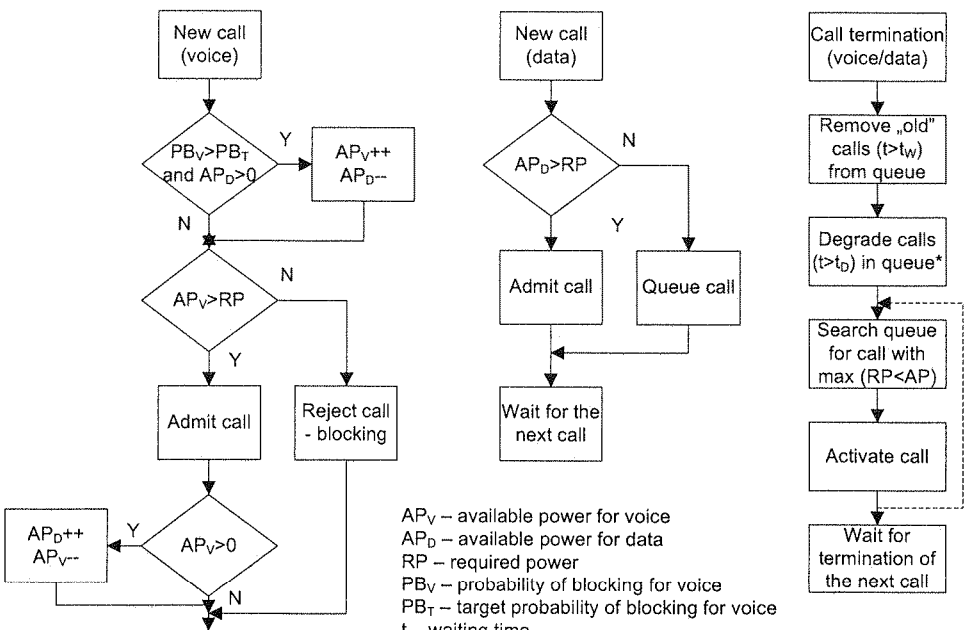


Fig. 2. Flowchart for algorithms CSD4-CSD5

Rys. 2. Algorytmy CSD4-CSD5

In the latter two algorithms the probability of blocking of data calls can be controlled by setting the maximum waiting time for each class. However, this is not possible

for voice calls, which can not be queued and should have priority over data calls. To ensure the required QoS for voice calls independently of the data traffic intensity, algoritm CSD4 has been proposed, in which the base station power resources are divided into two pools (Fig. 2). One of the pools is assigned to voice calls, the other to data calls. The probability of blocking of voice calls is constantly monitored and depending on it the voice call pool can shrink or grow by increasing or reducing the power available for data calls. The strategy for data calls remains unchanged.

Algorithm CSD5 (Fig. 2) is a modification of CSD4. It includes the call degradation procedure described above (see algorithm CSD3).

5. POWER ALLOCATION ALGORITMS FOR MIXED VOICE/PACKET DATA TRAFFIC

The algorithms for mixed voice/packet data traffic are based on these presented above, however, due to the bursty nature of data transmission in packet (connection-less) mode some modifications have to be introduced.

In algorithm BD1, presented in Fig. 3, incoming voice calls and data packets are accepted as long as the necessary power is available at the home base station. If the incoming call/packet requires the amount of power, which is not available at the time of the request, it is discarded and a blocking event is declared. Due to movement of mobile users a call/packet dropping may occur if the available power becomes insufficient. The QoS indicator is the probability of blocking, which is specified separately for voice calls and data packet transfer.

Algorithm BD2 applies a more sophisticated strategy (Fig. 3). If the incoming call/packet can not be accepted immediately, it is put in a queue. For each class of calls/packet transfers a maximum waiting time $t_W^{(class)}$ is specified. After termination of any active call/packet transfer the queue is searched and the voice call requiring the most (but available) power is activated. The search is repeated until power balance available at the base station is not sufficient to handle any more queued calls. During the search the waiting time of each call/packet is examined. If the waiting time exceeds the limit, $t_W^{(class)}$, for a given class, the call/packet is removed from the queue and the blocking event is declared. Next the search is repeated for data packet transfers. If the available power becomes insufficient due to the active user movement, calls/packet may be dropped. The QoS indicators are: probability of blocking, maximum and average waiting times for each class.

Algorithm BD3 adds the feature of packet transfer degradation (Fig. 3). For each class of data packet transfers a degradation time $t_D^{(class)}$ is specified. After removing "old" calls/packets the queue is searched again and the waiting times are examined. If the waiting time exceeds $t_D^{(class)}$, the packet transfer is degraded, i.e. moved to a lower class (with lower transmission rate). Similarly, if the power balance becomes insufficient due to movement of the active users, packet transfers may be degraded. In both cases only one degradation to the next lower class is allowed per packet request.

The specific QoS indicator for this algorithm is the probability of packet degradation for each class.

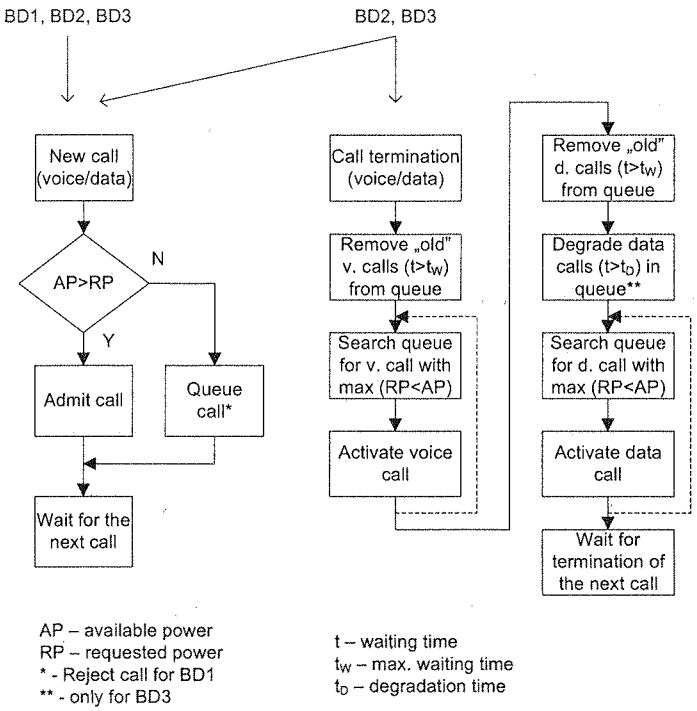


Fig. 3. Flowchart for algorithms BD1-BD3

Rys. 3. Algorytmy BD1-BD3

In algorithm BD4 the base station power resources are divided into two pools in order to reduce the influence of data traffic intensity on the probability of blocking and waiting times of voice calls and vice versa (Fig. 4). One of the pools is assigned to voice calls, the other to data packet transfers. The probability of blocking of voice calls is constantly monitored and depending on it the voice call pool can shrink or grow by increasing or reducing the power available for data packet transfers. In each group (voice/data) the strategy for calls/packets in a queue is identical to the one described for BD2.

Algorithm BD5 (Fig. 4) includes the packet transfer degradation procedure described above for data packet transfers (algorithm BD3).

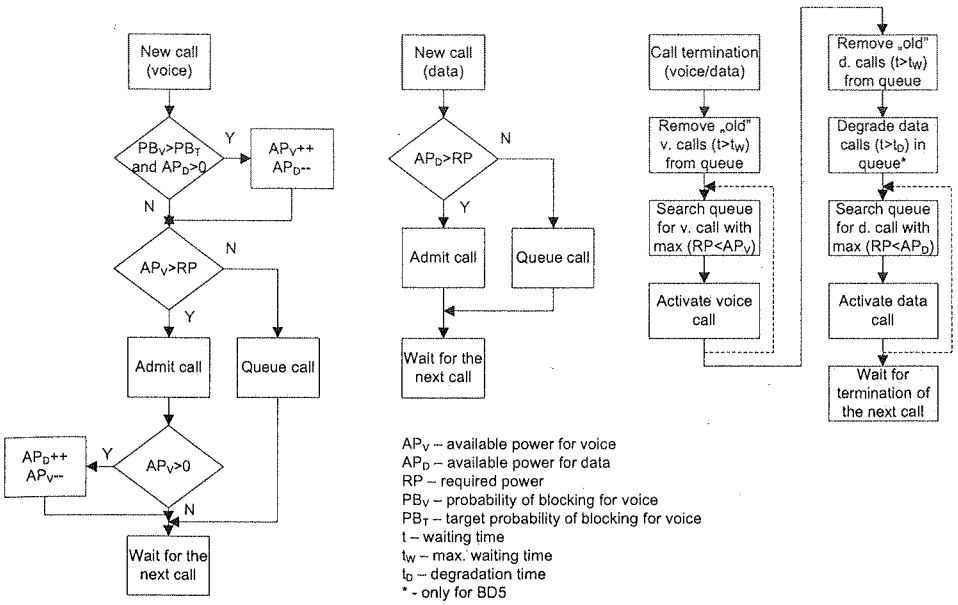


Fig. 4. Flowchart for algorithms BD4-BD5

Rys. 4. Algorytmy BD4-BD5

6. SIMULATION RESULTS

6.1. CASE I – CIRCUIT-SWITCHED DATA TRANSMISSION

To evaluate the performance of the proposed power allocation algorithms a series of computer simulations was run. A wideband CDMA system with a chip rate of 3.6864 Mchip/s and basic bit rate 9.6 kb/s was assumed. Four classes of data calls with rates 28.8 kb/s, 57.6 kb/s, 115.2 kb/s and 230.4 kb/s were specified. The required signal-to-noise-ratio was assumed to be 5.0 dB and the activity factors for voice and data calls were set to 0.4 and 0.6, respectively. The mean service time for voice calls $1/\mu_v = 200s$ and for data calls $1/\mu_d = 100s$. The path loss exponent equal to 3 was chosen [9] and the standard deviation of the shadowing process equal to 8.0 dB was assumed. The velocity of mobile stations was 10 m/s and the correlation of the shadowing process was equal to 0.6 over 100m distance. Data calls were split among the four classes: class 1 – 50%, class 2 – 30%, class 3 – 15%, class 4 – 5% of the overall data traffic. The target probabilities of blocking for voice and data calls were set to 2% and 5%, respectively. To obtain reliable results at least 10^6 calls were simulated for each algorithm and parameter set. The results for voice traffic of 40 Erl are discussed in the following.

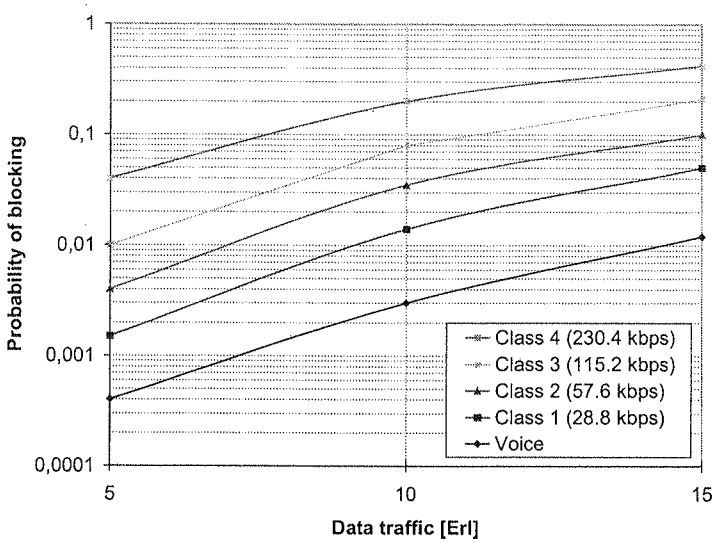


Fig. 5. Probability of blocking vs. offered data traffic for algorithm CSD1

Rys. 5. Prawdopodobieństwo blokady w zależności od natężenia ruchu dla algorytmu CSD1

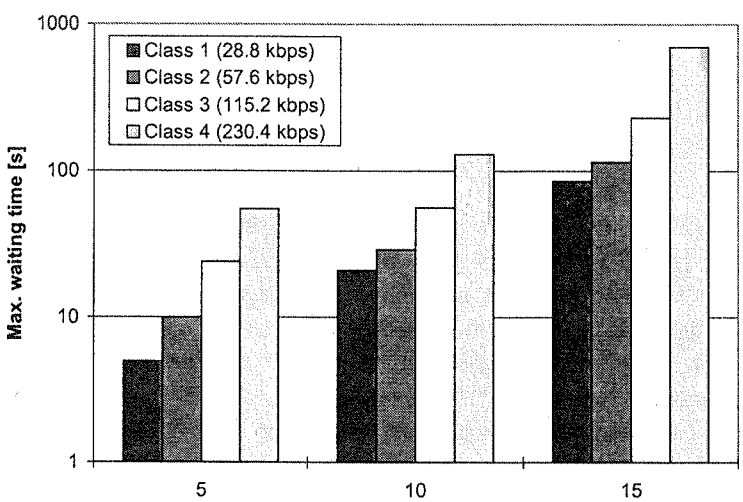


Fig. 6. Distribution of the maximum waiting times vs. offered data traffic for algorithm CSD5

Rys. 6. Rozkład maksymalnego czasu oczekiwania w zależności od natężenia ruchu dla algorytmu CSD5

In Fig. 5 the probability of blocking versus offered data traffic for algorithm CSD1 is shown. As expected, the probability of blocking is the lowest for voice calls (requiring lowest transmission rate, and consequently lowest power) and increases with the required transmission rate. Even for a moderate data traffic of 10 Erlangs the probabilities of blocking for Class 3 and Class 4 calls are unacceptably high (8% and 20%, respectively). When the more sophisticated algorithms are applied, the assumed 5% blocking for data calls can be easily maintained. Fig. 6 shows the distribution of the maximum waiting times $t_W^{(class)}$ versus offered data traffic for algorithm CSD5.

Distribution of the maximum waiting times versus loading of surrounding cells for algorithm CSD5 and data traffic equal to 15 Erl is presented in Fig. 7. It can be noted, that the maximum waiting times for each class are proportional to the loading of surrounding cells (which, of course, is strictly related to the level of intra-cell interference).

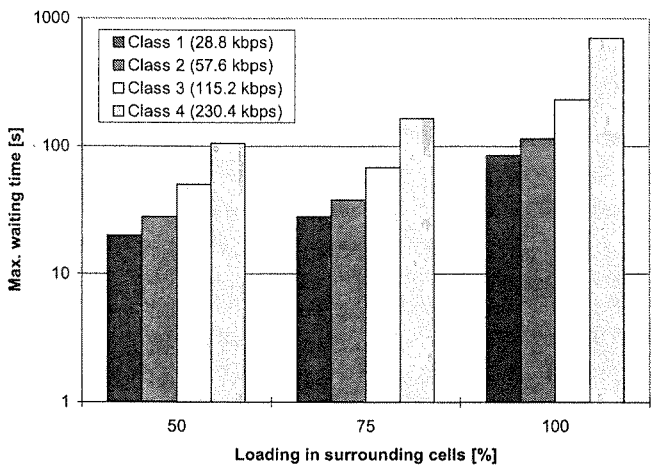


Fig. 7. Distribution of the maximum waiting times vs. loading of surrounding cells for algorithm CSD5, data traffic = 15 Erl

Rys. 7. Rozkład maksymalnego czasu oczekiwania w zależności od obciążenia sąsiednich komórek dla algorytmu CSD5 przy natężeniu ruchu 15 Erl

All proposed algorithms are compared in Fig. 8. It shows the distribution of the maximum waiting times for data traffic equal to 15 Erl and fully loaded surrounding cells. It can be seen that the maximum waiting times for Algorithms CSD2 and CSD3 are shorter than for algorithms CSD4 and CSD5. However, only the latter guarantee the maximum allowed probability of blocking of voice calls for the data traffic intensities of interest. The corresponding probabilities of degradation and dropping for data calls are presented in Table 1 with the assumption $t_D^{(class)} = t_W^{(class)}/2$. Since algorithm CSD4 is characterised by high probability of call dropping (there is no call degradation) it

seems that algorithm CSD5 is best suited for possible implementation. The distribution of the actual waiting times for this algorithm is given in Fig. 9.

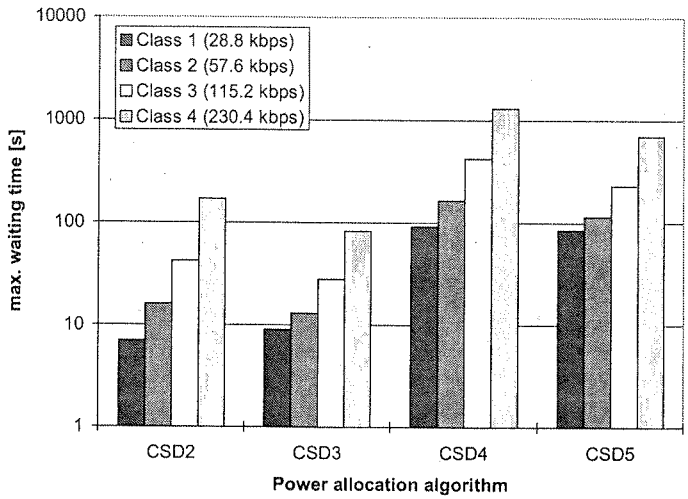


Fig. 8. Comparison of distribution of the maximum waiting times for Algorithms CSD2-CSD5, data traffic = 15 Erl

Rys. 8. Porównanie rozkładów maksymalnego czasu oczekiwania dla algorytmów CSD2-CSD5 przy natężeniu ruchu 15 Erl

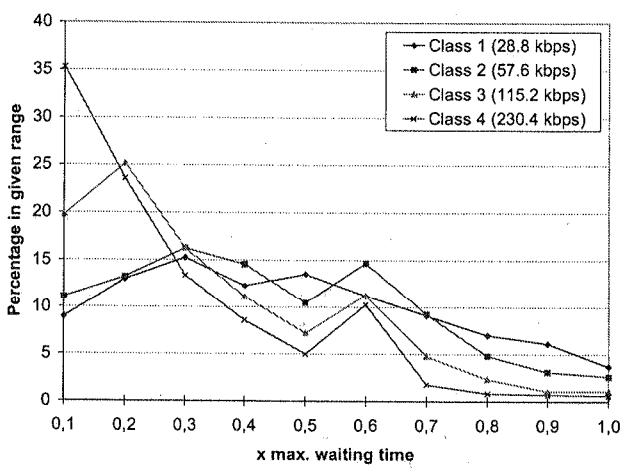


Fig. 9. Distribution of actual data call waiting times for algorithm CSD5 and data traffic = 10 Erl (for corresponding values of maximum waiting times see Fig. 6)

Rys. 9. Rozkład rzeczywistych czasów oczekiwania dla algorytmu CSD5 przy natężeniu ruchu 10 Erl

Tabela 1

Probabilities of call dropping/call degradation (CSD2-CSD5)

Prawdopodobieństwa zerwania oraz degradacji połączenia (CSD2-CSD5)

	Call dropping [%]				Call degradation [%]			
	class 1	class 2	class 3	class 4	class 1	class 2	class 3	class4
CSD2	0,1	2,7	17	49	0	0	0	0
CSD3	0	0	0,2	2,9	0	19	39	61
CSD4	0,1	2,6	15	37	0	0	0	0
CSD5	0	0	0,3	2,1	0	24	41	61

6.2. CASE II – PACKET DATA TRANSMISSION

In this case a wideband CDMA system with a chip rate of 3.6864 Mchip/s and basic bit rate 8 kb/s was assumed. Four classes of data calls with rates 8 kb/s, 32 kb/s, 64 kb/s and 144 kb/s were specified. The mean service time for voice calls $1/\mu_v = 100s$ and the mean packet size $\delta_d = 12kB$ were assumed. The other parameters remained unchanged. Only the results for voice traffic of 40 Erl are discussed in the following.

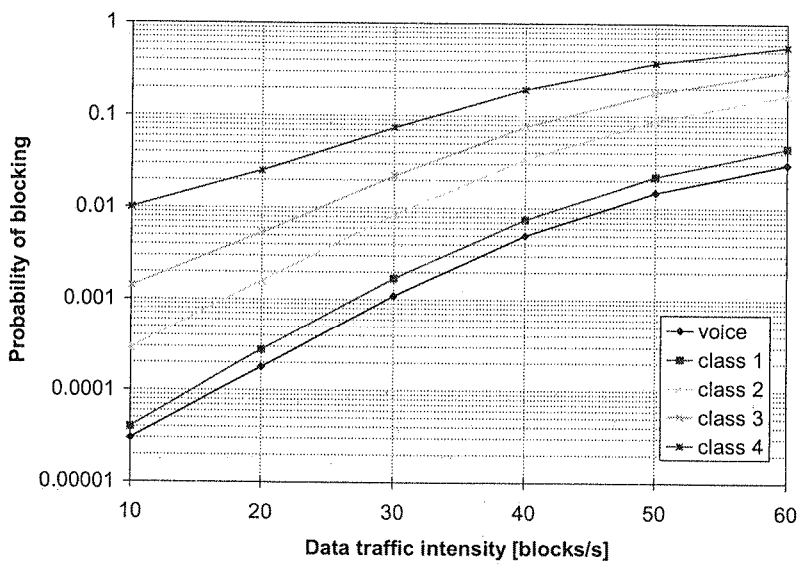


Fig. 10. Probability of blocking vs. offered data traffic for algorithm BD1

Rys. 10. Prawdopodobieństwo blokady w zależności od natężenia ruchu dla algorytmu BD1

In Fig. 10 the probability of blocking versus offered data traffic for algorithm BD1 is shown. As expected, the probability of blocking is the lowest for voice calls (requiring lowest transmission rate, and consequently lowest power) and increases with the required transmission rate. Even for a moderate data traffic of 40 packets/s the probabilities of blocking for Class 3 and Class 4 data transfers are unacceptably high (8% and 20%, respectively).

When the more sophisticated algorithms are applied, the assumed 2% blocking for voice calls and 5% blocking for data transfers can be easily maintained. Fig. 11 shows the distribution of the maximum waiting times $t_W^{(class)}$ versus offered data traffic for Algorithm BD3.

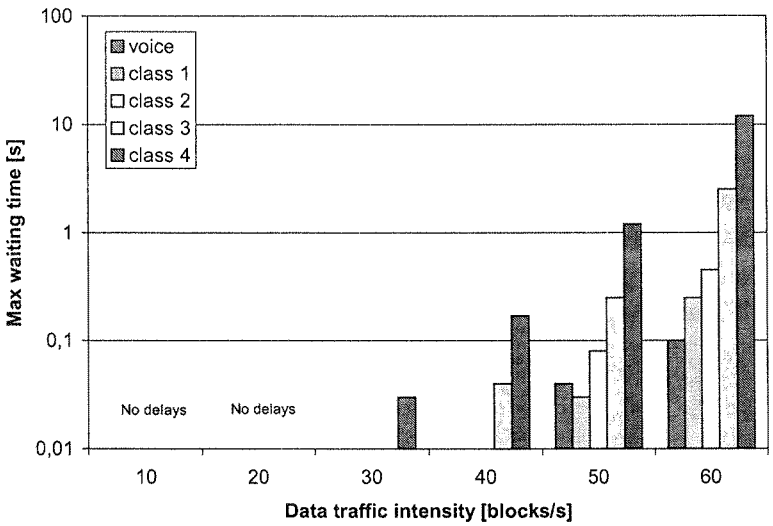


Fig. 11. Distribution of the maximum waiting times vs. offered data traffic for algorithm BD3

Rys. 11. Rozkład maksymalnego czasu oczekiwania w zależności od natężenia ruchu dla algorytmu BD3

All proposed algorithms are compared in Fig. 12. It shows the distribution of the maximum waiting times for data traffic equal to 50 packets/s and fully loaded surrounding cells. It can be seen that the maximum waiting times for Algorithms BD2 and BD3 are shorter than for algorithms BD4 and BD5. This results from less efficient power allocation from the two distinct pools.

The corresponding probabilities of degradation and dropping for data transfers are presented in Table 2 with the assumption $t_D^{(class)} = t_W^{(class)}/2$. Since the probability of call dropping is higher for Algorithm BD2 then for algorithm BD3 it seems that the last one is best suited for possible implementation in the case under consideration.

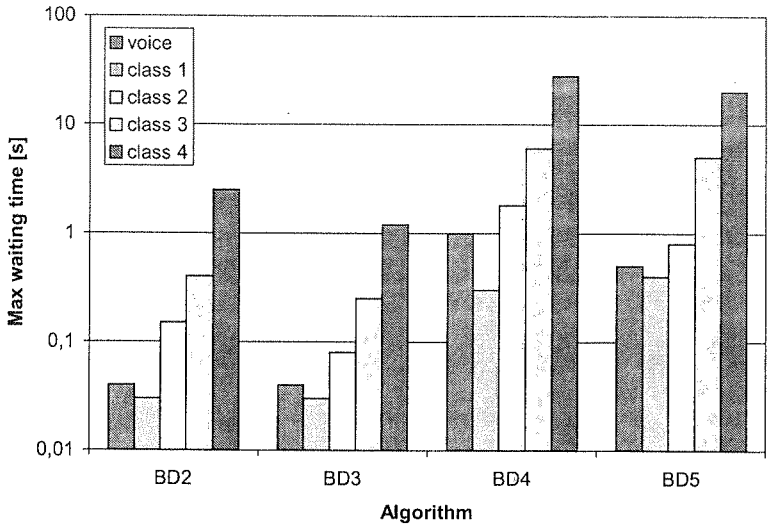


Fig. 12. Comparison of distribution of the maximum waiting times for algorithms BD2-BD5, data traffic = 50 packets/s

Rys. 12. Porównanie rozkładów maksymalnego czasu oczekiwania dla algorytmów BD2-BD5 przy natężeniu ruchu 50 pakietów/s

Tabela 2

Probabilities of call dropping/call degradation (BD2-BD5)

Prawdopodobieństwa zerwania oraz degradacji połączenia (BD2-BD5)

	Call dropping [%]				Call degradation [%]			
	class 1	class 2	class 3	class 4	class 1	class 2	class 3	class4
BD2	0	0	0,002	0,009	0	0	0	0
BD3	0	0	0	0	0	11,9	12	13,8
BD4	0	0	0,001	0,003	0	0	0	0
BD5	0	0	0	0	0	26,3	45,3	23

Finally, Fig. 13 shows the distribution of the mean waiting times for data traffic equal to 50 packets/s and fully loaded surrounding cells. Again, algorithms BD2 and BD3 show their advantages over algorithms BD4 and BD5 with the mean waiting times lower at least 10 times.

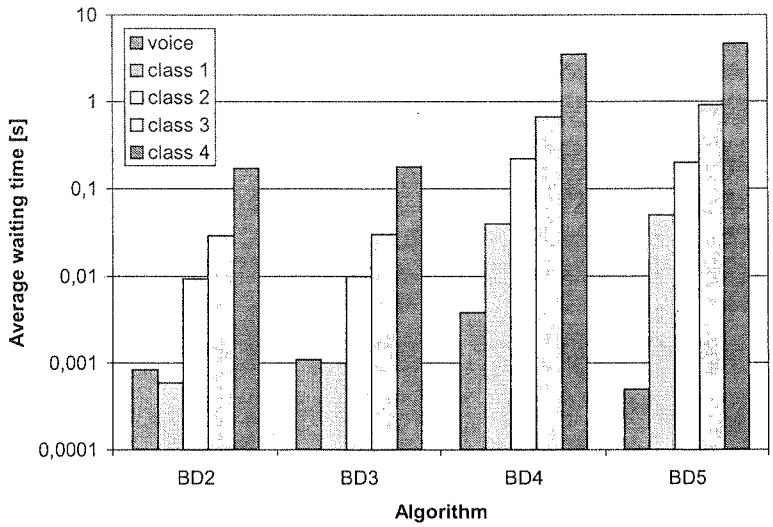


Fig. 13. Comparison of distribution of the mean waiting times for algorithms BD2-BD5, data traffic = 50 packets/s

Rys. 13. Porównanie rozkładów średniego czasu oczekiwania dla algorytmów BD2-BD5 przy natężeniu ruchu 50 pakietów/s

7. CONCLUSIONS

Several power allocation algorithms for a multiservice CDMA wireless system have been proposed and analyzed. Their application in an integrated voice/data system can ensure desired quality of service. It has been shown that simple algorithms (like algorithms CSD1 and BD1) can not guarantee the required probability of blocking, especially for high rate data calls, and more complex schemes must be applied. Assuming that users can wait for access to the system and the required data rates are flexible over a certain range, it has been shown that algorithm CSD5 is best suited in switched-circuit data mode, while for system exploiting packet data transmission algorithm BD3 offers the best performance. Future work should investigate performance of similar algorithms for other types of mixed traffic.

8. REFERENCES

1. W. Lee: *"Overview of cellular CDMA"*, IEEE Trans. Veh. Technol., vol. 40, pp. 291-302, May 1991.
2. TIA/EIA/IS-95A, *"Mobile Station – Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System"*, Telecommunications Industry Association, Arlington, VA, May 1995.
3. C. Smith, D. Collins: *"3G Wireless Networks"*, McGraw-Hill, 2002.

4. B. Walke, P. Seidenberg, M.P. Althoff: *"UMTS – The Fundamentals"*, Wiley, 2003.
5. K.B. Letaief, J. C-I Chuang, R.D. Murch: *"Multicode high-speed transmission for wireless personal communications"*, in Proc. IEEE Globecom '95, pp. 1835-1839, Nov. 1995, Singapore.
6. H. Zhang, D. Rutkowski: *"Orthogonal sequence division modulation – a novel method for future broadband radio services"*, in Proc. IEEE VTC'95, pp. 810-814, July 1995, Chicago.
7. A. Sampath, P.S. Kumar, J.M. Holtzman: *"Power control and resource management for a multimedia CDMA wireless system"*, in Proc. PIMRC '95, pp. 21-25, Sept. 1995, Toronto.
8. N.B. Mandayam, J. Holtzman, S. Barberis: *"Erlang capacity for an integrated voice/data DS-CDMA wireless system with variable bit rate sources"*, ibid., pp. 1078-1082.
9. A. Sampath, N.B. Mandayam, J.M. Holtzman: *"Analysis of an access control mechanism for data traffic in an integrated voice/data wireless CDMA system"*, in Proc. VTC' 96, pp. 1448-1452, May 1996, Atlanta.
10. W.C.Y. Lee: *Mobile Communications Engineering*, 2nd ed. McGraw-Hill, New York, 1998.
11. T.S. Rappaport, L.B. Milstein: *"Effects of radio propagation path loss on DS-CDMA cellular frequency reuse efficiency for the reverse channel"*, IEEE Trans. on Veh. Technol., vol. 41, pp. 231-241, Aug. 1992.
12. M. Gudmundson: *"Correlation model for shadow fading in mobile radio systems"*, Electronics Letters, vol. 27, pp. 2145-2146, Nov. 1991.

R. KRENZ

PRZYDZIAŁ MOCY NA ŁĄCZU „W DÓŁ" W SYSTEMACH KOMÓRKOWYCH ZE ZWIELOKROTNIENIEM KODOWYM DS-CDMA

Streszczenie

Systemy komórkowe trzeciej generacji (3G), takie jak UMTS czy Cdma2000, wykorzystują metodę wielodostępu ze zwielokrotnieniem kodowy (CDMA). Uzyskuje się dzięki temu większą niż w przypadku zwielokrotnienia TDMA/FDMA pojemność systemu, a także poprzez stosowaniu transmisji szerokopasmowej z bezpośrednim rozpraszaniem widma (DS) lepszą odporność na zaniki selektywne i zakłócenia wąskopasmowe.

W przypadku łącza „w dół" o pojemności decyduje w znacznym stopniu optymalny przydział mocy stacji bazowej do poszczególnych kanałów, którymi realizowana jest transmisja do stacji ruchomych. Znaczenie tego zagadnienia jest często niedoceniane, dlatego w niniejszym artykule podjęto próbę opracowania i weryfikacji algorytmów przydziału mocy w stacji bazowej dla transmisji głosu oraz danych, zarówno z komutacją łączy, jak i z komutacją pakietów. Po krótkim wprowadzeniu w Rozdziale 2 przedstawiono model analizowanego systemu komórkowego. Rozdział 3 zawiera matematyczna analizę zagadnienia przydziału mocy w stacji bazowej. W rozdziałach 4 i 5 przedstawiono natomiast opracowane przez autora algorytmy przydziału mocy, odpowiednio dla transmisji danych z komutacją łączy (CSDx) oraz z komutacją pakietów (BDx). Wyniki badań symulacyjnych, które wykorzystano do weryfikacji zaproponowanych algorytmów, zostały przedstawione w rozdziale 6.

Pierwszą grupę (CSD1/BD1) stanowią najprostsze algorytmy, w których w przypadku braku zasobów nadchodzące połączenie głosowe lub danych jest odrzucane (blokada). Wskaźnikiem oceny jakości (QoS) jest tu prawdopodobieństwo wystąpienia blokady.

W drugiej grupie algorytmów dla usługi transmisji danych wprowadzono możliwość oczekiwania (kolejkowanie) na zasoby stacji bazowej w przypadku ich braku w momencie pojawienia się żądania połączenia. Jeżeli zasoby nie zostaną przydzielone połączeniu przez określony czas, zostaje ono odrzucone (blokada). Oprócz prawdopodobieństwa wystąpienia blokady jakość usług jest w tym przypadku oceniana również na podstawie średniego oraz maksymalnego czasu oczekiwania.

W trzeciej grupie uzupełniono powyższe algorytmy o możliwość zmniejszenia szybkości transmisji danych (w stosunku do żądanej) po przekroczeniu określonego czasu oczekiwania (CSD3/BD3). Dodatkowym wskaźnikiem jakości jest więc prawdopodobieństwo degradacji połączenia.

Aby zmniejszyć wpływ połączeń danych na połączenia głosowe w kolejnej grupie algorytmów (CSD4/BD4) wydzielono część zasobów wyłącznie na potrzeby połączeń głosowych. Ich wielkość jest na bieżąco regulowana (kosztem zasobów dla połączeń danych), tak, aby nie przekroczyć określonego prawdopodobieństwa blokady dla połączeń głosowych.

W ostatniej grupie algorytmów (CSD5/BD5) do poprzedniego rozwiązania dodano możliwość degradacji (obniżenia szybkości transmisji) połączeń danych, podobnie, jak w grupie 3.

Zaproponowane algorytmy zostały zweryfikowane z wykorzystaniem symulacji komputerowej. Konfigurację oraz parametry modelowanego systemu przedstawiono odpowiednio w rozdziałach 2 i 6. Badania te wykazały, że zastosowanie prostych algorytmów (np. CSD1/BD1) prowadzi do przekroczenia założonego prawdopodobieństwa blokady już przy niewielkich natężeniach ruchu oferownego, w szczególności dla transmisji danych o większych szybkościach (Rys. 5/Rys. 11). Szczegółowa analiza uzyskanych wyników pozwala na stwierdzenie, że w przypadku transmisji danych z komutacją łączy najlepszym (z punktu widzenia weryfikowanych wskaźników jakości) spośród zaproponowanych jest algorytm CSD5 (Rys. 6–Rys. 9, Tabela 1), natomiast dla pakietowej transmisji danych algorytm BD3 (Rys. 11–Rys. 13, Tabela 2).

Słowa kluczowe: CDMA, łącze „w dół", przydział zasobów, komutacja łączy, komutacja pakietów

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2006, 52, z. 2, ss. 229–243

Estymacja położenia przy użyciu bezśladowego filtru Kalmana

STANISŁAW KONATOWSKI

Wojskowa Akademia Techniczna, Instytut Radioelektroniki Wydziału Elektroniki
ul. Kaliskiego 2, 00-908 Warszawa
skonatowski@wat.edu.pl

Otrzymano 2004.09.07
Autoryzowano 2005.07.11

Proces estymacji położenia w zintegrowanych systemach nawigacyjnych jest często realizowany na nieliniowych modelach systemów. Nieliniowość dynamiki obiektu, którego pozycję należy estymować, wymaga stosowania odpowiednich filtrów. Powszechnie przyjętym rozwiązaniem jest rozszerzony filtr Kalmana oparty na linearyzacji funkcji nieliniowych. W artykule przedstawiono ideę bezśladowego filtru Kalmana wykorzystującego przekształcenie bezśladowe. Zaprezentowano wyniki badań symulacyjnych, które wykazują lepszą dokładność estymacji położenia przy użyciu bezśladowego niż rozszerzonego filtru Kalmana.

Słowa kluczowe: bezśladowy filtr Kalmana, przekształcenie bezśladowe, model nieliniowy, zintegrowany system nawigacyjny

1. WPROWADZENIE

W zintegrowanych systemach nawigacyjnych wykorzystuje się dane pochodzące z różnych urządzeń nawigacyjnych. Najbardziej rozpowszechnioną metodą integracji danych jest łączenie urządzeń autonomicznych z radiotechnicznymi. Czujniki autonomiczne charakteryzują się prostą konstrukcją i większą niezawodnością niż urządzenia radiotechniczne. Są one bardziej odporne na zakłócenia czynne i bierne dzięki wykorzystaniu pierwotnych zjawisk fizycznych. Podstawową wadą systemów autonomicznych jest ich niska długoterminowa dokładność – błędy narastają w czasie w wyniku operacji całkowania. Problem pogarszającej się z czasem dokładności nie dotyczy urządzeń radiotechnicznych – w szczególności systemów satelitarnych. Ich dokładność jest lepsza niż systemów autonomicznych, lecz są one bardziej zawodne i nieodporne na zakłócenia

czynne i bierne. Poważnym mankamentem systemów radiotechnicznych jest możliwość utraty ciągłości informacji w wyniku działania zakłóceń lub zaniku sygnałów [6].

Podstawowymi metodami wykorzystywanymi w integracji danych nawigacyjnych są kompensacja i filtracja – metody te przenikają się na wzajem i uzupełniają. W metodzie kompensacji wyróżniony jest jeden przyrząd nawigacyjny, którego dane wyjściowe są kompensowane danymi z innych czujników. W metodzie filtracji dane pochodzące z różnych czujników poddawane są filtracji w celu pozbycia się zakłóceń, a następnie służą do wyznaczania pozycji lub innego elementu nawigacyjnego za pomocą odpowiedniego algorytmu [1, 3]. W praktyce, stosowane filtry Kalmana działają w czasie dyskretnym i pełnią rolę estymatorów błędów lub elementów nawigacyjnych. Za pomocą filtru Kalmana estymuje się stan obiektu (lub błędu) w chwili k-tej na podstawie pomiarów z chwili k-1. Filtry Kalmana wykorzystują informację o dynamice obiektu (systemu). Znajomość dynamiki obiektu, a także jej odpowiednie zamodelowanie jest kluczowym problemem w implementacji tych filtrów. W zależności od dynamiki systemu stosować można różne rodzaje filtrów. Dla systemów z dynamiką liniową zasadnym jest użycie zwykłego filtru Kalmana. W systemach z dynamiką nieliniową konieczne jest przeprowadzenie linearyzacji modelu i w takim przypadku powszechnie stosowany jest rozszerzony filtr Kalmana EKF (*Extended Kalman Filter*) [1–3]. Linearyzacja odbywa się przy użyciu pochodnych cząstkowych nieliniowej funkcji stanu lub przez jej rozwinięcie w szereg Taylora.

Alternatywą dla rozszerzonego filtru Kalmana jest bezśladowy filtr Kalmana UKF (*Unscented Kalman Filter*) [2, 7–11]. Jest to rekursywny filtr estymujący, którego własności dobrze spełniają wymagania modeli silnie nieliniowych. W odróżnieniu do EKF, filtr bezśladowy nie linearyzuje modelu, ale operuje na parametrach statystycznych poddanych nieliniowym przekształceniom wektorów stanu i pomiarowego. Podstawą działania UKF jest *przekształcenie bezśladowe* UT (*Unscented Transform*) [9–11].

2. PRZEKSZTAŁCENIE BEZŚLADOWE

Przekształcenie bezśladowe jest metodą obliczania statystyk zmiennej losowej poddanej nieliniowemu przekształceniu przy założeniu, że wygodniej jest estymować rozkład prawdopodobieństwa, niż dowolną funkcję nieliniową [1–2, 4–5, 8–10]. Aby obliczyć wartość średnią i wariancję n-wymiarowej zmiennej losowej, powstałej w wyniku jej nieliniowego przekształcenia, wyznacza się zbiór $2n + 1$ ważonych punktów sigma $S_i = \{W_i, \chi_i\}$. Punkty te są tak dobierane, aby dokładnie odzwierciedlały wartość średnią i wariancję zmiennej losowej $\mathbf{x}$. Zakłada się, że zmienna $\mathbf{x} \in \Re^n$ ma rozkład Gaussowski $N(\bar{\mathbf{x}}, \mathbf{P}_x)$, a jej składowe poddane są przekształceniu przez funkcję nieliniową $\mathbf{y} = f(\mathbf{x})$. Dąży się do określenia możliwie najdokładniejszej aproksymacji rozkładu prawdopodobieństwa zmiennej $\mathbf{y}$.

W przekształceniu bezśladowym obliczanie zbioru punktów sigma $\{\chi_i\}$ i wag W_i odbywa się zgodnie z następującymi zależnościami [6, 8–11]:

$$\chi_0 = \bar{\mathbf{x}}, \qquad W_0 = \kappa/(n+\kappa) \qquad \text{dla} \qquad i = 0, \tag{1}$$

$$\chi_i = \bar{\mathbf{x}} + \left(\sqrt{(n+\kappa)\,\mathbf{P}_{\mathbf{x}}}\right)_i, \qquad W_i = 1/[2(n+\kappa)], \quad \text{dla} \qquad i = 1, \ldots, n, \tag{2}$$

$$\chi_i = \bar{\mathbf{x}} - \left(\sqrt{(n+\kappa)\,\mathbf{P}_{\mathbf{x}}}\right)_{i-n}, \qquad W_i = 1/[2(n+\kappa)] \quad \text{dla} \qquad i = n+1, \ldots, 2n, \tag{3}$$

gdzie: κ – parametr skalujący,
$\left(\sqrt{(n+\kappa)\,\mathbf{P}_{\mathbf{x}}}\right)_i$ – i-ty wiersz bądź kolumna pierwiastka kwadratowego macierzy $\mathbf{P}_{\mathbf{x}}$,
W_i – wagi związane z i-tym punktem w taki sposób, że $\sum_{i=0}^{2n} W_i = 1$.
Każdy punkt sigma jest następnie propagowany przez funkcję nieliniową

$$\mathbf{Y}_i = f(\chi_i) \quad \text{dla} \quad i = 0, \ldots, 2n. \tag{4}$$

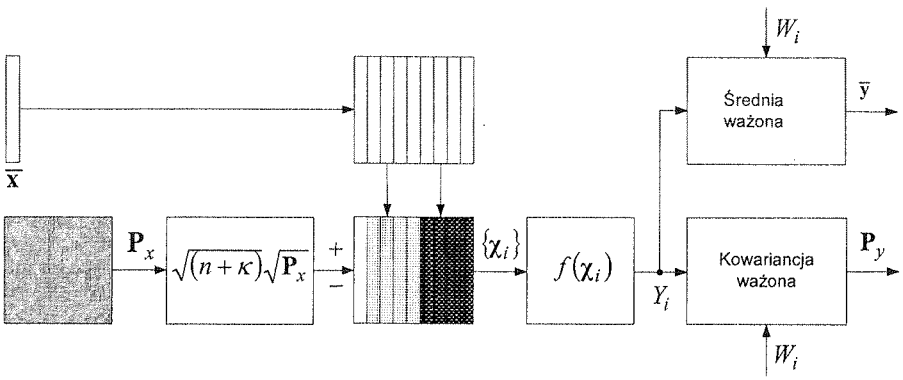
Poszukiwana wartość średnia $\bar{\mathbf{y}}$ i kowariancja $\mathbf{P}_y$ zmiennej losowej $\mathbf{y}$ wyznaczane są z wyrażeń:

$$\bar{\mathbf{y}} = \sum_{i=0}^{2n} W_i \mathbf{Y}_i, \tag{5}$$

$$\mathbf{P}_y = \sum_{i=0}^{2n} W_i (\mathbf{Y}_i - \bar{\mathbf{y}})\ (\mathbf{Y}_i - \bar{\mathbf{y}})^{\mathrm{T}}. \tag{6}$$

Otrzymuje się w ten sposób parametry rozkładu wyjściowej zmiennej losowej $\mathbf{x}$. Dokładność wyznaczonej wartości średniej i kowariancji jest na poziomie rozwinięcia funkcji nieliniowej w szereg Taylora do wyrazów rzędu drugiego włącznie. Błędy wprowadzają wyrazy wyższe rozwinięcia, lecz są one ważone poprzez dobór parametru κ [5]. Rysunek 1 przedstawia ideę przekształcenia bezśladowego.

W przypadku bardzo silnych nieliniowości punkty sigma mogą być przeskalowane od lub do wartości średniej wcześniejszego rozkładu (*a priori*) poprzez odpowiedni dobór parametru κ. Odległość i-tego punktu sigma od wartości średniej $\bar{\mathbf{x}}$ jest proporcjonalna do $\sqrt{n+\kappa}$. Gdy $\kappa = 0$ odległość ta jest proporcjonalna do $\sqrt{n}$. W przypadku gdy $\kappa > 0$, punkty są przeskalowane od $\bar{\mathbf{x}}$, a gdy $\kappa < 0$ punkty są przeskalowane do $\bar{\mathbf{x}}$. Dla szczególnego przypadku $\kappa = 3 - n$, pożądana niezmienniczość skalowania wymiarowego jest uzyskiwana przez eliminację wpływu n. Jednak dla wyrażenia $\kappa = 3 - n$ mniejszego od zera i wagi $W_0 < 0$ obliczona kowariancja może być niedodatnio półokreślona [5, 8, 10].



Rys. 1. Idea przekształcenia bezśladowego [10]

Fig. 1. Idea of the Unscented Transform [10]

Aby wyeliminować ten problem wprowadza się ważone przekształcenie bezśladowe SUT (*scaled unscented transformation*) [8], które zamienia oryginalny zbiór punktów sigma $S = \{W,\chi\}$ na zbiór przekształcony $S' = \{W',\chi'\}$ według zależności:

$$W_i' = \begin{cases} W_i^{(m)} = (W_0 - 1)/\alpha^2 + 1 & \text{dla } i = 0 \\ W_i^{(c)} = W_i^{(m)} + 1 - \alpha^2 + \beta^2 & \text{dla } i = 0 \\ W_i^{(m)} = W_i^{(c)} = W_i/\alpha^2 & \text{dla } i \neq 0 \end{cases}, \tag{7}$$

$$\chi_i' = \mathbf{x}_0 + \alpha\,(\mathbf{x}_i - \mathbf{x}_0) \quad \text{dla} \quad i = 0,\ ...,\ 2n,$$

gdzie α jest dodatnim parametrem ważącym nowych punktów sigma, który może być wybrany dowolnie małym, by minimalizować efekty wyższych rzędów rozwinięcia szeregu Taylora. Indeks górny (m) występuje przy wagach wykorzystywanych w procesie obliczeń wartości średnich, zaś indeks górny (c) przy wagach stosowanych do obliczeń macierzy kowariancji zmiennych losowych.

Dobór punktów sigma i ich ważenie może być realizowane w jednym kroku, aby zmniejszyć liczbę obliczeń, poprzez parametr skalujący λ

$$\lambda = \alpha^2\,(n + \kappa) - n. \tag{8}$$

Wtedy doboru punktów sigma dokonuje się zgodnie z zależnościami:

$$\chi_0 = \bar{\mathbf{x}}, \tag{9}$$

$$\chi_i = \bar{\mathbf{x}} + \left(\sqrt{(n+\lambda)\,\mathbf{P}_\mathbf{x}}\right)_i \quad \text{dla } i = 1,\ ...,\ n, \tag{10}$$

$$\chi_i = \bar{\mathbf{x}} - \left(\sqrt{(n+\lambda)\mathbf{P}_\mathbf{x}}\right)_{i-n} \quad \text{dla} \quad i = n+1, \ ..., \ 2n. \tag{11}$$

Związane z tymi punktami wagi obliczane są następująco [8]:

$$W_0^{(m)} = \lambda/(n+\lambda), \tag{12}$$

$$W_0^{(c)} = \lambda/(n+\lambda) + 1 - \alpha^2 + \beta, \tag{13}$$

$$W_i^{(m)} = W_i^{(c)} = 1/[2(n+\lambda)] \quad \text{dla } i = 1, \ ..., \ 2n. \tag{14}$$

Stałe α, β, κ są parametrami przekształcenia spełniającymi następujące założenia:

$$0 < \alpha \leqslant 1, \quad \kappa \geqslant 0, \quad \beta \geqslant 0. \tag{15}$$

Parametr κ decyduje o odległości punktów sigma do średniej $\bar{\mathbf{x}}$, a jego optymalną wartością w większości zastosowań jest zero. Parametr dodatni α minimalizuje efekty wyższych rzędów rozwinięcia szeregu Taylora. Stała β steruje wagą zerowego punktu sigma i często przyjmuje się ją równą 2 [5, 10].
Punkty sigma poddawane są nieliniowemu przekształceniu przez funkcję f

$$\mathbf{Y}_i = f(\chi_i) \quad \text{dla} \quad i = 0, \ ..., \ 2n. \tag{16}$$

Wtedy poszukiwana wartość średnia $\bar{\mathbf{y}}$ i kowariancja $\mathbf{P}_y$ zmiennej losowej $\mathbf{y}$ wyznaczane są z zależności:

$$\bar{\mathbf{y}} = \sum_{i=0}^{2n} W_i^{(m)} \mathbf{Y}_i, \tag{17}$$

$$\mathbf{P}_y = \sum_{i=0}^{2n} W_i^{(c)} (\mathbf{Y}_i - \bar{\mathbf{y}}) \ (\mathbf{Y}_i - \bar{\mathbf{y}})^{\mathrm{T}}. \tag{18}$$

3. ALGORYTM FILTRU UKF

Przedstawione powyżej przekształcenie bezśladowe jest podstawą działania bezśladowego filtru Kalmana, który operuje na następujących modelach stanu i pomiarowym:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \ \mathbf{u}_{k-1}, \ \mathbf{w}_{k-1}) \quad \text{dla} \quad \mathbf{w}_{k-1} \sim N(\mathbf{0}, \ \mathbf{Q}_{k-1}), \tag{19}$$

$$\mathbf{y}_k = h(\mathbf{x}_k, \ \mathbf{v}_k) \quad \text{dla} \quad \mathbf{v}_k \sim N(\mathbf{0}, \ \mathbf{R}_k). \tag{20}$$

W algorytmie bezśladowego filtru Kalmana używane są następujące oznaczenia:
- $\mathbf{x}_k = \mathbf{x}(t_k)$ jest n-wymiarowym wektorem stanu w chwili t_k,
- $\mathbf{y}_k$ jest p-wymiarowym wektorem pomiarowym w chwili t_k,
- $f(\mathbf{x}, \mathbf{u}, \mathbf{w})$ oznacza nieliniową funkcję stanu, opisującą zachowanie dynamiczne systemu. Charakteryzuje zmiany stanu systemu pomiędzy chwilami t_{k-1} i t_k,
- $\mathbf{u}$ jest wektorem sterowania (wejściowym systemu),
- $\mathbf{w}$ jest wektorem zakłóceń stanu,
- $\mathbf{Q}$ jest macierzą kowariancji zakłóceń stanu. Określa stopień niepewności w modelu dynamicznym podczas przejścia od chwili t_{k-1} do t_k,
- $h(\mathbf{x}, \mathbf{v})$ oznacza nieliniową funkcję pomiarową,
- $\mathbf{v}$ jest p-wymiarowym wektorem zakłóceń pomiarowych,
- $\mathbf{R}$ jest macierzą kowariancji błędów pomiarowych o wymiarach $p \times p$,
- $\mathbf{P}$ jest macierzą kowariancji wektora stanu o wymiarach $n \times n$.

W filtrze UKF wektor stanu jest zdefiniowany jako złożenie stanu oryginalnego i zmiennych szumu. Oznacza to, że jest on powiększony o wektory zakłóceń $\mathbf{w}$ i $\mathbf{v}$ do wymiaru n_a o kowariancjach odpowiednio $\mathbf{Q}$ i $\mathbf{R}$

$$\mathbf{x}_k^a = \begin{bmatrix} \mathbf{x}_k^{\mathrm{T}} & \mathbf{w}_k^{\mathrm{T}} & \mathbf{v}_k^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \tag{21}$$

a wektor punktów sigma ma postać:

$$\mathcal{X}^a = \begin{bmatrix} (\mathcal{X}^x)^{\mathrm{T}} & (\mathcal{X}^w)^{\mathrm{T}} & (\mathcal{X}^v)^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}. \tag{22}$$

Algorytm działania UKF, podobnie jak filtru Kalmana, przedstawia się następująco:

1. inicjalizacja – obliczenie wartości średniej i kowariancji wektora stanu w chwili początkowej:

$$\hat{\mathbf{x}}_0 = E[\mathbf{x}_0], \quad \mathbf{P}_0 = \left[\left(\mathbf{x}_0 - \hat{\mathbf{x}}_0\right) \left(\mathbf{x}_0 - \hat{\mathbf{x}}_0\right)^{\mathrm{T}} \right], \tag{23}$$

$$\hat{\mathbf{x}}_0^a = E[\mathbf{x}^a] = \begin{bmatrix} \hat{\mathbf{x}}_0^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \tag{24}$$

$$\mathbf{P}_0^a = E\left[\left(\mathbf{x}_0^a - \hat{\mathbf{x}}_0^a\right) \left(\mathbf{x}_0^a - \hat{\mathbf{x}}_0^a\right)^{\mathrm{T}} \right] = \begin{bmatrix} \mathbf{P}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix}, \tag{25}$$

2. aktualizacja dla $k \in \{1, 2, \ldots, \infty\}$:
   - obliczanie punktów sigma zgodnie z równaniami (9–11) przekształcenia UT:

$$\mathcal{X}_{k-1}^a = \left[\hat{\mathbf{x}}_{k-1}^a, \hat{\mathbf{x}}_{k-1}^a + \sqrt{(n_a + \lambda)\mathbf{P}_{k-1}^a}, \hat{\mathbf{x}}_{k-1}^a - \sqrt{(n_a + \lambda)\mathbf{P}_{k-1}^a} \right], \tag{26}$$

   - aktualizacja punktów sigma wektora stanu (21) na podstawie punktów: $\mathcal{X}_{k-1}^x$ dla wektora stanu i $\mathcal{X}_{k-1}^w$ dla wektora zakłóceń stanu:

$$\mathcal{X}^x_{k|k-1} = f\left(\mathcal{X}^x_{k-1},\ \mathbf{u}_{k-1},\ \mathcal{X}^w_{k-1}\right), \tag{27}$$

- przewidywanie wartości średniej wektora stanu (dla wag wyliczonych według zależności (12–14)):

$$\bar{\hat{\mathbf{x}}}_k = \sum_{i=0}^{2n_a} W_i^{(m)} \mathcal{X}^x_{i,k|k-1}, \tag{28}$$

- przewidywanie kowariancji wektora stanu:

$$\mathbf{P}_k^- = \sum_{i=0}^{2n_a} W_i^{(c)} \left(\mathcal{X}^x_{i,k|k-1} - \bar{\hat{\mathbf{x}}}_k\right) \left(\mathcal{X}^x_{i,k|k-1} - \bar{\hat{\mathbf{x}}}_k\right)^{\mathrm{T}}. \tag{29}$$

3. aktualizacja pomiarów (dla zaktualizowanych punktów sigma $\mathcal{X}^x_{k|k-1}$ wektora stanu i punktów sigma $\mathcal{X}^v_{k-1}$ wektora zakłóceń procesu pomiarowego):

$$\mathbf{Y}_{k|k-1} = h\left(\mathcal{X}^x_{k|k-1},\ \mathcal{X}^v_{k-1}\right), \tag{30}$$

$$\bar{\hat{\mathbf{y}}}_k = \sum_{i=0}^{2n_a} W_i^{(m)} Y_{i,k|k-1}, \tag{31}$$

$$\mathbf{P}_{y_k y_k} = \sum_{i=0}^{2n} W_i^{(c)} \left(Y_{i,k|k-1} - \bar{\hat{\mathbf{y}}}_k\right) \left(Y_{i,k|k-1} - \bar{\hat{\mathbf{y}}}_k\right)^{\mathrm{T}}, \tag{32}$$

$$\mathbf{P}_{x_k y_k} = \sum_{i=0}^{2n_a} W_i^{(c)} \left(\mathcal{X}_{i,k|k-1} - \bar{\hat{\mathbf{x}}}_k\right) \left(Y_{i,k|k-1} - \bar{\hat{\mathbf{y}}}_k\right)^{\mathrm{T}}, \tag{33}$$

- wyliczanie macierzy wzmocnień Kalmana:

$$\mathbf{K}_k = \mathbf{P}_{x_k y_k} \mathbf{P}_{y_k y_k}^{-1}, \tag{34}$$

- obliczanie skorygowanej wartości średniej wektora stanu:

$$\bar{\hat{\mathbf{x}}}_k = \bar{\hat{\mathbf{x}}}_k + \mathbf{K}_k \left(\mathbf{y}_k - \bar{\hat{\mathbf{y}}}_k\right), \tag{35}$$

- obliczanie skorygowanej macierzy kowariancji wektora stanu:

$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k \mathbf{P}_{y_k y_k} \mathbf{K}_k^{\mathrm{T}}. \tag{36}$$

Algorytm UKF wymaga obliczenia pierwiastka kwadratowego macierzy. Jeśli operacja ta jest wykonana przy użyciu rozkładu na czynniki Cholesky'ego to rozwiązanie uzyskuje się kosztem $1/6\ n^3$ działań (n jest wymiarem wektora stanu). Jeśli macierze kowariancji są wyliczane rekursywnie to pierwiastek kwadratowy może być obliczony kosztem n^2 działań [7, 8, 10]. Algorytm filtru UKF (19–36) jest ogólną formą bezśladowego filtru Kalmana.

4. ALGORYTM FILTRU UKF Z ADDYTYWNYM SZUMEM O ZEROWEJ WARTOŚCI ŚREDNIEJ

Dla szczególnego (lecz często spotykanego) przypadku, w którym szumy procesu i pomiaru są czysto addytywne, złożoność obliczeniowa UKF może zostać znacznie zredukowana. W takim przypadku wymiar wektora stanu nie musi być zwiększany o wymiary wektorów szumów systemu. Zmniejsza to wymiar punktów sigma, jak również ich całkowitą liczbę. Realizuje się to przy użyciu prostej procedury addytywnej [9–10]. Operacje, które realizuje algorytm UKF z szumami addytywnymi są następujące:

- inicjalizacja zgodnie z zależnością (23),
- wyznaczenie punktów sigma dla $k \in \{1, 2, ..., \infty\}$:

$$\mathcal{X}_{k-1} = \left[\hat{\mathbf{x}}_{k-1},\ \hat{\mathbf{x}}_{k-1} + \left(\sqrt{(n+\lambda)\,\mathbf{P}_{k-1}}\right),\ \hat{\mathbf{x}}_{k-1} - \left(\sqrt{(n+\lambda)\,\mathbf{P}_{k-1}}\right)\right], \tag{37}$$

- aktualizacja:

$$\mathcal{X}^*_{k|k-1} = f\left(\mathcal{X}_{k-1},\ \mathbf{u}_{k-1}\right), \tag{38}$$

$$\hat{\mathbf{x}}_k^- = \sum_{i=0}^{2n} W_i^{(m)} \mathcal{X}^*_{i,k|k-1}, \tag{39}$$

$$\mathbf{P}_k^- = \sum_{i=0}^{2n} W_i^{(c)} \left(\mathcal{X}^*_{i,k|k-1} - \hat{\mathbf{x}}_k^-\right)\left(\mathcal{X}^*_{i,k|k-1} - \hat{\mathbf{x}}_k^-\right)^{\mathrm{T}} + \mathbf{Q}_{k-1}, \tag{40}$$

- ponowne przeliczenie punktów sigma:

$$\mathcal{X}_{k|k-1} = \left[\mathcal{X}^*_{k|k-1},\ \mathcal{X}^*_{0,k|k-1} + \left(\sqrt{(n+\lambda)\,\mathbf{Q}_{k-1}}\right),\ \mathcal{X}^*_{0,k|k-1} - \left(\sqrt{(n+\lambda)\,\mathbf{Q}_{k-1}}\right)\right], \tag{41}$$

- korekcja – wyznaczanie estymatorów wartości średniej i kowariancji wektora stanu:

$$\mathbf{Y}_{k|k-1} = h\left(\mathcal{X}_{k|k-1}\right), \tag{42}$$

$$\hat{\mathbf{y}}_k^- = \sum_{i=0}^{2n} W_i^{(m)} Y_{i,k|k-1}, \tag{43}$$

$$\mathbf{P}_{y_k y_k} = \sum_{i=0}^{2n} W_i^{(c)} \left(Y_{i,k|k-1} - \hat{\mathbf{y}}_k^-\right)\left(Y_{i,k|k-1} - \hat{\mathbf{y}}_k^-\right)^{\mathrm{T}} + \mathbf{R}_{k-1}, \tag{44}$$

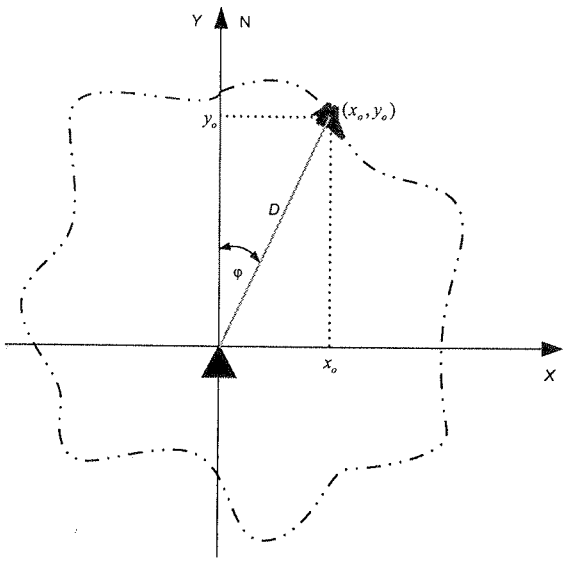
$$\mathbf{P}_{x_k y_k} = \sum_{i=0}^{2n} W_i^{(c)} \left(\mathcal{X}_{i,k|k-1} - \hat{\mathbf{x}}_k^-\right)\left(Y_{i,k|k-1} - \hat{\mathbf{y}}_k^-\right)^{\mathrm{T}}, \tag{45}$$

- pozostałe wyrażenia na macierz wzmocnień Kalmana oraz skorygowane: średnią wartość wektora stanu i kowariancję wektora stanu wylicza się z zależności (34–36).

5. BADANIA FILTRÓW

Celem badań jest porównanie jakości estymacji rozszerzonego filtru Kalmana EKF i bezśladowego filtru Kalmana UKF. Porównanie jakości estymacji położenia jest przeprowadzone dla obiektu, który porusza się wokół radiolatarni systemu bliskiej nawigacji SBN. Radiolatarnia SBN mierzy odległość do obiektu i jego azymut. Pomiary te podawane są na wejścia filtrów, których zadaniem jest estymowanie pozycji obiektu w prostokątnym układzie współrzędnych. Rysunek 2 przedstawia rozpatrywany model w formie graficznej.

Nieliniowość systemu wynika z konieczności przeliczenia współrzędnych biegunowych (odległość D i azymut φ) na współrzędne prostokątne (x, y).



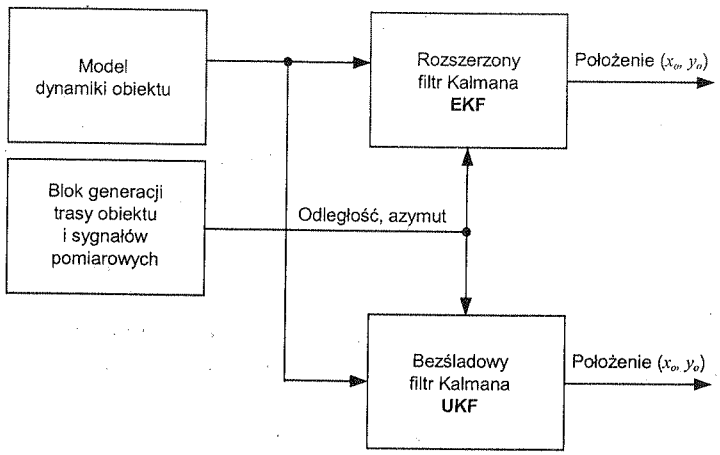
Rys. 2. Pomiar odległości i azymutu w systemie bliskiej nawigacji SBN

Fig. 2. Measurement of distance and azimuth in SBN

Procesy pomiaru odległości i azymutu obiektu przez system bliskiej nawigacji SBN oraz estymacji położenia obiektu przez dwa dyskretne filtry Kalmana (rozszerzony i bezśladowy) zostały zaprojektowane w środowisku MATLAB. Schemat blokowy układu do symulacji i badań przedstawiony jest na rysunku 3.

Na wejścia filtrów Kalmana podawane są sygnały azymutu i odległości z addytywnymi szumami pomiarowymi oraz informacja o modelu dynamiki obiektu. Dostarczane

są także pomiary bez szumów w celu uzyskania rzeczywistych pozycji (trasy) obiektu poruszającego się ze stałą prędkością. Sygnały pomiarowe są podstawą do estymowania pozycji obiektu i wykreślania jego trasy.



Rys. 3. Schemat układu do symulacyjnego badania filtrów

Fig. 3. Block diagram of the system for simulation

W badaniach przyjęto, że wektor stanu ma postać:

$$\mathbf{x} = [\ x_o \quad y_o \quad v_x \quad v_y\ ]^{\mathrm{T}}, \tag{46}$$

gdzie: x_o, y_o – współrzędne prostokątne obiektu,
v_x, v_y – składowe prędkości obiektu.
Macierz stanu ma postać:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \mathrm{T} & 0 \\ 0 & 1 & 0 & \mathrm{T} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{47}$$

gdzie T jest okresem między chwilami t_{k-1} i t_k.

W analizowanym systemie SBN występują zależności nieliniowe między współrzędnymi układu biegunowego – w którym dokonywane są pomiary, a współrzędnymi układu prostokątnego – w którym wyznaczana jest pozycja obiektu:

$$D = \sqrt{x_o^2 + y_o^2}, \quad \varphi = \arctan(y_o/x_o). \tag{48}$$

Zależności (48) posłużyły do skonstruowania nieliniowej funkcji pomiarowej $h(\mathbf{x})$

$$h(\mathbf{x}) = \begin{bmatrix} \sqrt{x_o^2 + y_o^2} \\ \arctan(y_o/x_o) \end{bmatrix}. \tag{49}$$

Wiadomo [1, 4], że w algorytmie rozszerzonego filtru Kalmana EKF, aby wyznaczyć macierz pomiarową $\mathbf{H}_{EKF}$ należy wyznaczyć pochodne cząstkowe nieliniowej funkcji pomiarowej $h(\mathbf{x})$ (49) względem wszystkich elementów wektora stanu (46):

$$\mathbf{H}_{EKF} = \frac{\partial h(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{x_o}{\sqrt{x_o^2 + y_o^2}} & \dfrac{y_o}{\sqrt{x_o^2 + y_o^2}} & 0 & 0 \\ \dfrac{-y_o}{x_o^2 + y_o^2} & \dfrac{x_o}{x_o^2 + y_o^2} & 0 & 0 \end{bmatrix}. \tag{50}$$

Zakłócenia pomiarowe odległości D i azymutu φ charakteryzowane są przez macierz kowariancji błędów pomiarowych $\mathbf{R}$:

$$\mathbf{R} = \begin{bmatrix} R_D & 0 \\ 0 & R_\varphi \end{bmatrix}, \tag{51}$$

gdzie: R_D – wariancja błędów pomiarowych odległości,
R_φ – wariancja błędów pomiarowych azymutu.

Algorytmy filtrów EKF (zależności 46–51) i UKF (zależności 23, 34–45) zostały zrealizowane w środowisku MATLAB. W filtrze UKF przyjęto następujące wartości parametrów [8, 10] przekształcenia bezśladowego: $\alpha = 0{,}5$, $\beta = 2$, $\kappa = 0$. Wartości początkowe wektora stanu zawierają elementy o wartościach zerowych

$$\mathbf{x}_0 = [\,0 \quad 0 \quad 0 \quad 0\,]^{\mathrm{T}}. \tag{52}$$

Początkowa macierz kowariancji wektora stanu ma postać:

$$\mathbf{P}_0 = \begin{bmatrix} 1000\ \mathrm{m}^2 & 0 & 0 & 0 \\ 0 & 1000\ \mathrm{m}^2 & 0 & 0 \\ 0 & 0 & 100\ (\mathrm{m/s})^2 & 0 \\ 0 & 0 & 0 & 100\ (\mathrm{m/s})^2 \end{bmatrix}. \tag{53}$$

Dla zapewnienia identycznych warunków podczas badania filtrów macierze kowariancji błędów stanu $\mathbf{Q}$ i błędów pomiarowych $\mathbf{R}$ dla rozpatrywanego systemu w obydwu filtrach mają jednakowe postacie.

W procesie symulacji ruchu obiektu wokół radiolatarni zostały wygenerowane sygnały pomiarowe przy użyciu następujących zależności:

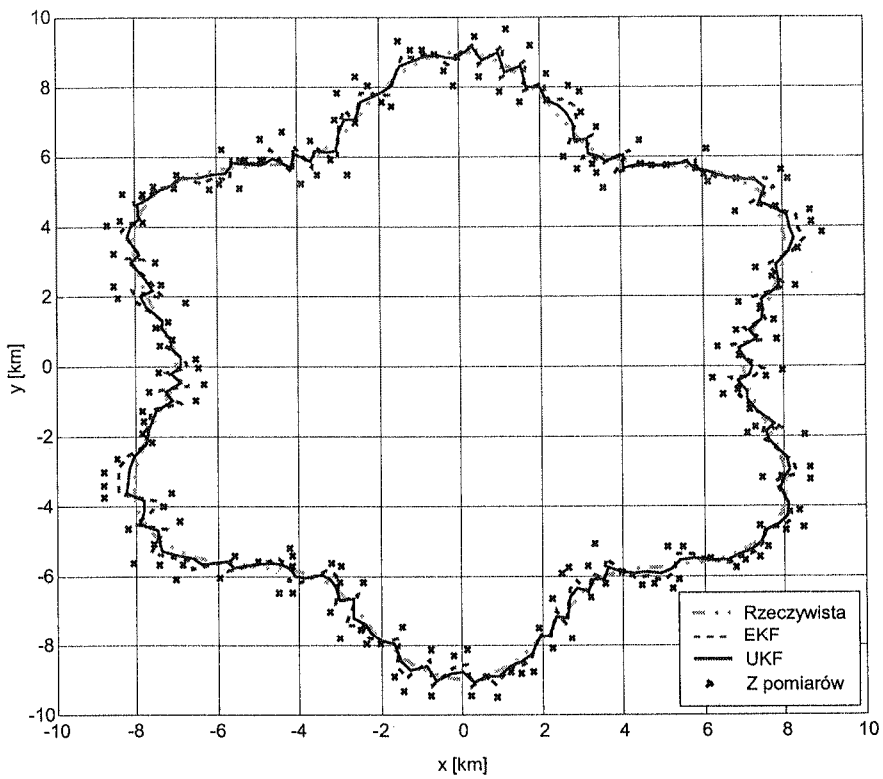
$$D = R + C\cos(Z\varphi) + v_D, \tag{54}$$

$$\varphi = \alpha_T + v_\varphi, \tag{55}$$

gdzie: R – stała odległość bazowa,
C, Z – współczynniki kształtu trasy,
α_T – kąt zmieniający się od 0° do 360° ze stałym krokiem $T = 2°$,
$v_{D,\varphi}$ – szumy odległości i azymutu uwzględniane podczas generacji trasy obiektu.

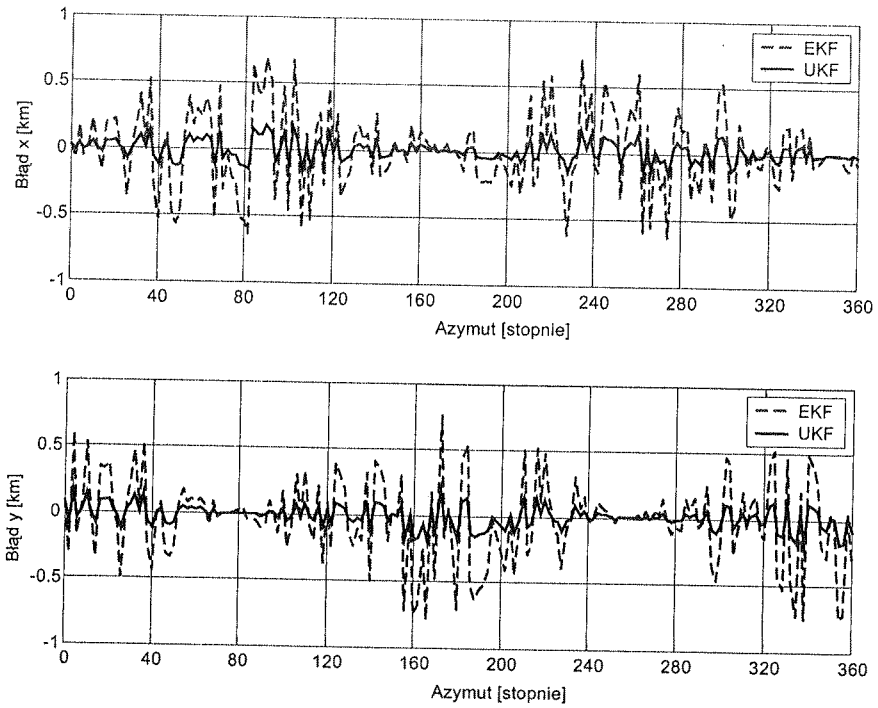
Przyjęto następujące wartości współczynników kształtu trasy obiektu: $C = 1$, $Z = 6$, odległość bazowa $R = 8$ km, skok azymutu $T = 2°$.

Rysunek 4 przedstawia trasy obiektu powstałe na podstawie estymowanych pozycji przez obydwa filtry, rzeczywistą i wynikającą bezpośrednio z pomiarów. Natomiast wartości błędów współrzędnych prostokątnych prezentuje rysunek 5. Błędy te są różnicą między pozycją rzeczywistą a estymowaną przez analizowane filtry.



Rys. 4. Trasa obiektu

Fig. 4. The route of the object



Rys. 5. Błędy współrzędnych biegunowych

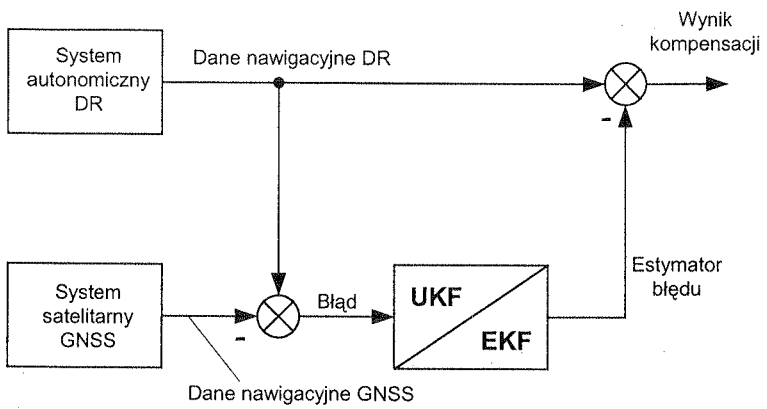
Fig. 5. Position errors (x and y coordinate) in rectangular coordinates

6. PODSUMOWANIE

Celem artykułu było przedstawienie istoty bezśladowego filtru Kalmana UKF, a także porównanie jego działania na tle rozszerzonego filtru Kalmana EKF. Został scharakteryzowany podstawowy jego algorytm oraz algorytm dla addytywnego szumu o zerowej wartości średniej. Stosowanie filtrów UKF jest niezbędne w poprawnym procesie przetwarzania danych nawigacyjnych, szczególnie w przypadku nieliniowych funkcji stanu i/lub pomiarowej systemu – występujących w zagadnieniach nawigacyjnych.

Przykładowe wyniki symulacyjnych badań porównawczych jakości estymacji położenia z wykorzystaniem rozszerzonego i bezśladowego filtru Kalmana pokazują, że zastosowanie bezśladowego filtru Kalmana UKF, jako algorytmu przetwarzania danych w systemie z nieliniową funkcją pomiarową, poprawia proces estymacji. Błędy położenia estymowane przez UKF we współrzędnych prostokątnych są mniejsze niż dla filtru EKF. Lepsza jakość estymacji położenia z wykorzystaniem filtru UKF pociąga za sobą konieczność wykonywania większej liczby obliczeń (szczególnie obliczania pierwiastków kwadratowych macierzy), co czyni go bardziej restrykcyjnym dla jednostek obliczeniowych w zintegrowanych systemach nawigacyjnych.

W artykule porównano wyniki pracy filtrów, jako algorytmów przetwarzania danych nawigacyjnych określających położenie obiektu. Kontynuacją badań bezśladowego filtru Kalmana może być jego zastosowanie w estymacji błędu zintegrowanego systemu nawigacyjnego DR/GNSS (rys. 6), wykorzystującego ideę kompensacji.



Rys. 6. Zintegrowany system nawigacyjny z kompensacją błędów

Fig. 6. Integrated navigation system based on the compensation mode

7. BIBLIOGRAFIA

1. R.G. Brown, P.Y.C. Hwang: *Introduction to Random Signals and Applied Kalman Filtering with MATLAB Exercises and Solutions*. John Wiley & Sons, Canada, 1997.
2. N.J. Gordon, B. Ristic, S. Arulampalam: *Beyond the Kalman Filter – Particle Filters for Tracking Applications*. Artech House, London, 2004.
3. M.S. Grewal, A.P Andrews: *Kalman filtering Theory and Practice Using MATLAB*. John Wiley & Sons, Canada, 2001.
4. S.J. Julier, J.K. Uhlmann, H.F. Durrant-Whyte: *A New Method for the Nonlinear Transformation of Means and Covariances in Flters and Estimators*. IEEE Transactions on Automatic Control, vol. 45, no. 3, March 2000, pp. 477-482.
5. S.J. Julier, J.K. Uhlmann: *A New Extension of the Kalman Filter to Nonlinear Systems*. Proceedings of AeroSense: The 11-th International Symposium on Aerospace/Defense Sensing, Simulations and Controls, vol. II Multi Sensor Fusion, Tracking and Resource Management, Orlando, Florida, 1997, pp. 182-193.
6. L.A. Klein: *Sensor Technologies and Data Requirements for ITS*. Artech House, London, 2001.
7. R. van der Merwe, A. Doucet, N. de Freitas, E.A. Wan: *The Unscented Particle Filter*. Technical Report CUED/F-INFENG/TR 380, Cambridge University Engineering Department, Cambridge, 2000.
8. R. van der Merwe, E.A. Wan: *The Square-root Unscented Kalman Filter for State and Parameter-estimation*. Proceedings of International Conference on Acoustics, Speech, and Signal Processing, vol. 6, Salt Lake City, May 2001, pp. 3461-3464.
9. R. van der Merwe, E.A. Wan: *The Unscented Kalman Filter*. Department of Electrical and Computer Engineering, Oregon Graduate Institute of Science and Technology, Beaverton, Oregon, 2001.

10. E.A. Wan, R. van der Merwe: *The Unscented Kalman Filter to appear in Kalman Filtering and Neural Networks.* Chapter 7. Edited by Simon Haykin, John Wiley & Sons, USA, 2001.
11. E.A. Wan, R. van der Merwe: *The Unscented Kalman Filter for Nonlinear Estimation.* Proc. IEEE Symp. Adaptive Systems for Signal Proc., Communication and Control (AS-SPCC), Lake Louise, Alberta, Canada, October 2000, pp. 153-158.

S. KONATOWSKI

POSITION ESTIMATION USING UNSCENTED KALMAN FILTER

Summary

In integrated navigation systems different kinds of Kalman Filter working as error estimators or navigation algorithms are widely used. These filters work in time-discrete mode. Kalman Filters utilize information about dynamics of the object (system). Knowledge about dynamics and its correct modelling is the main issue in implementation of the Kalman Filters. In systems with linear dynamics, it is adequately to use basic Kalman Filter. Systems with nonlinear dynamics require linearization of the system model and in such case Extended Kalman Filter (EKF) is generally accepted.

Unscented Kalman Filter (UKF) is an alternative for EKF. UKF is a recursive-estimating filter, which properties meet well requirements of strongly nonlinear systems. UKF does not linearize the model but manipulate on statistical parameters of nonlinear transformed state and measurement vector. UKF bases on Unscented Transform (UT). UT converts the state vector into a set of weighted Sigma Points. These points are than used in algorithms of UKF. The UKF algorithm is a set of equations, which are necessary to do prediction, innovation and correction steps.

Simulation results of position estimation using EKF and UKF show that UKF used as data processing algorithm gives better accuracy of estimation in system with nonlinear dynamics than EKF. Nonlinearity in system used in simulation causes by transformation of co-ordination systems. Such situation takes place very often in navigation. This shows that UKF is more suitable to systems with strong nonlinearities than EKF. Better accuracy of position estimation using UKF calls for large number of computations (especially evaluation of matrix square root), what makes it more demanding for computation units of integrated navigation systems. UKF may also be used to estimate errors in integrated navigation system based on the compensation mode.

Keywords: Unscented Kalman filter, unscented transform, nonlinear model, integrated navigation system

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2006, 52, z. 2, ss. 245–266

Wykorzystanie systemów pomiarowych z transmisją danych przez sieci telefonii komórkowej GSM oraz Internet

MICHAŁ MAĆKOWSKI

Instytut Elektroniki i Telekomunikacji, Politechnika Poznańska
ul. Piotrowo 3A, 60-965 Poznań
mmackow@et.put.poznan.pl

Otrzymano 2004.09.16
Autoryzowano 2005.07.29

W artykule przedstawiono rozproszone systemy pomiarowe wykorzystujące do transmisji danych sieci komórkowe GSM. Omówiono możliwości takich systemów. Opisano usługi transmisji danych dostępne w sieciach GSM, które można zastosować w systemach pomiarowych. Pokazano przykładowe rozwiązania systemów pomiarowych zbudowanych i sprawdzonych przez autora w Politechnice Poznańskiej. Zaprezentowano wyniki pomiarów efektywnej szybkości transmisji danych.

Słowa kluczowe: rozproszone systemy pomiarowe, transmisja danych, GSM, GPRS, EDGE, UMTS, Internet

1. WSTĘP

Zastosowanie do transmisji danych telefonii komórkowej GSM oraz Internetu stwarza nowe możliwości rozwiązań rozproszonych systemów pomiarowych, stanowiąc konkurencję dla innych systemów radiokomunikacyjnych stosowanych obecnie w telemetrii i telemechanice.

Bezprzewodowa transmisja danych pomiarowych jest alternatywą dla systemów przewodowych, szczególnie w przypadkach, gdy obiekt pomiaru znajduje się w znacznej odległości od centrali systemu pomiarowego, lub znajduje się w trudno dostępnym miejscu (np. tereny leśne, parki narodowe) oraz gdy obiekt pomiaru przemieszcza się po rozległym obszarze. Systemy pomiarowe z transmisją danych przez sieci GSM dobrze spełniają to zadanie. Ich podstawowymi zaletami są duży zasięg potencjalnego systemu (obszar działania sieci GSM), oraz brak wysokich nakładów inwestycyjnych przeznaczonych na budowę infrastruktury telekomunikacyjnej.

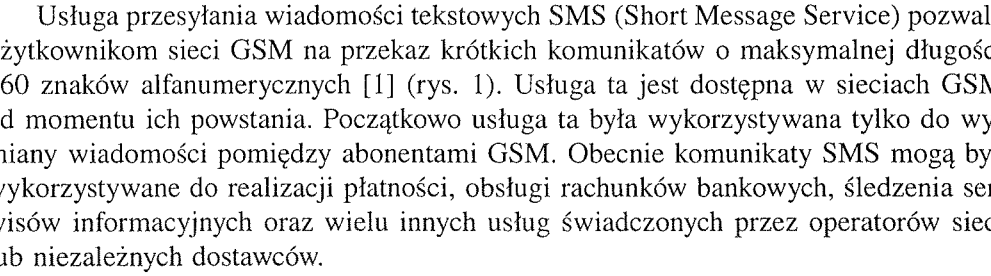
W systemach pomiarowych można wykorzystać wszystkie usługi transmisji danych dostępne w sieciach GSM. O wyborze konkretnego sposobu transmisji decyduje przeznaczenie systemu oraz jego wymagania określające ilość, częstość i szybkość transmitowanych danych w systemie.

Obecnie sieci GSM umożliwiają transmisję danych z wykorzystaniem różnych technologii: SMS, MMS, CSD (HSCSD), GPRS, EDGE oraz UMTS.

2. USŁUGI TRANSMISJI DANYCH W SIECIACH KOMÓRKOWYCH GSM

2.1. WIADOMOŚCI TEKSTOWE I MULTIMEDIALNE – SMS, MMS

Usługa przesyłania wiadomości tekstowych SMS (Short Message Service) pozwala użytkownikom sieci GSM na przekaz krótkich komunikatów o maksymalnej długości 160 znaków alfanumerycznych [1] (rys. 1). Usługa ta jest dostępna w sieciach GSM od momentu ich powstania. Początkowo usługa ta była wykorzystywana tylko do wymiany wiadomości pomiędzy abonentami GSM. Obecnie komunikaty SMS mogą być wykorzystywane do realizacji płatności, obsługi rachunków bankowych, śledzenia serwisów informacyjnych oraz wielu innych usług świadczonych przez operatorów sieci lub niezależnych dostawców.



Rys. 1. Ramka krótkiego komunikatu tekstowego SMS

Fig. 1. Short Massage Services (SMS) frame structure

Wymiana wiadomości SMS jest realizowana za pośrednictwem węzła sieci zwanego Centrum Obsługi Wiadomości SMS (SMSC, Short Message Service Center). Komunikaty SMS nie są przesyłane bezpośrednio pomiędzy abonentami, ale dwuetapowo przez centrum SMSC. Każdy komunikat jest wysyłany do SMSC, gdzie jest przechowywany, a następnie przesyłany dalej do odbiorcy. W SMSC komunikaty są przechowywane do czasu ich dostarczenia, co umożliwia późniejsze przekazanie SMS'a odbiorcy, który jest aktualnie niedostępny. Centrum SMSC zapewnia również potwierdzenie dostarczenia komunikatu SMS.

Komunikaty SMS są dostarczane z opóźnieniem, typowo jest to kilka, kilkanaście sekund, ale jest możliwa nawet zwłoka kilku godzinna lub kilku dniowa, a w skrajnym przypadku komunikat może nie zostać dostarczony. Wielkość opóźnienia zależy od dostępności odbiorcy, aktualnego obciążenia centrum SMSC oraz natężenia ruchu

w sieci GSM. SMS'y są transmitowane przez kanały sygnalizacyjne interfejsu radiowego, które są głównie wykorzystywane do procedur systemowych i nie oferują dużej szybkości transmisji. W przypadku dużego natężenia ruchu w komórce oraz dużej liczby przesyłanych komunikatów SMS może dochodzić do "zatykania się" kanałów, co powoduje powstawanie opóźnień w transmisji SMS'ów.

Przesyłanie danych pomiarowych za pomocą komunikatów SMS umożliwia budowę systemów pomiarowych, w których nie jest wymagana duża częstość transmitowania danych (kilka, kilkanaście pomiarów na dobę). Przykładem wykorzystania komunikatów SMS mogą być systemy meteorologiczne, do przekazywania raportów ze zdalnych stacji pomiarowych. SMS'y mogą być również wykorzystywane w systemach pomiarowych do transmisji awaryjnej, w przypadku, gdy transmisja przez podstawowy kanał transmisyjny (CSD lub GPRS) będzie niemożliwa.

Wiadomości multimedialne MMS (Multimedia Message Service) mogą zawierać: sformatowany tekst, zdjęcia, obrazy, animacje, wideo, dźwięk oraz prezentacje multimedialne. Podobnie jak komunikaty SMS, mogą być wymieniane pomiędzy abonentami sieci GSM lub mogą być wysyłane w postaci poczty elektronicznej. MMS może zawierać kilkaset kilobajtów danych, co pozwala na jednorazowe przesłanie dużej liczby danych pomiarowych w dowolnej postaci: np.: pliku tekstowego, arkusza kalkulacyjnego z zarejestrowanymi wynikami pomiarów, oscylogramu, wykresu charakterystyki itp. Jednostkowa opłata za przesłanie komunikatu MMS jest naliczana za każde rozpoczęte 100 kB (sieć Plus GSM). Usługę MMS można, w systemach pomiarowych, wykorzystać do przesyłania okresowych (dziennych, tygodniowych, itp.) raportów z serii pomiarowych lub do przesyłania informacji w postaci graficznej np.: oscylogramy, obrazy termograficzne z kamer IR.

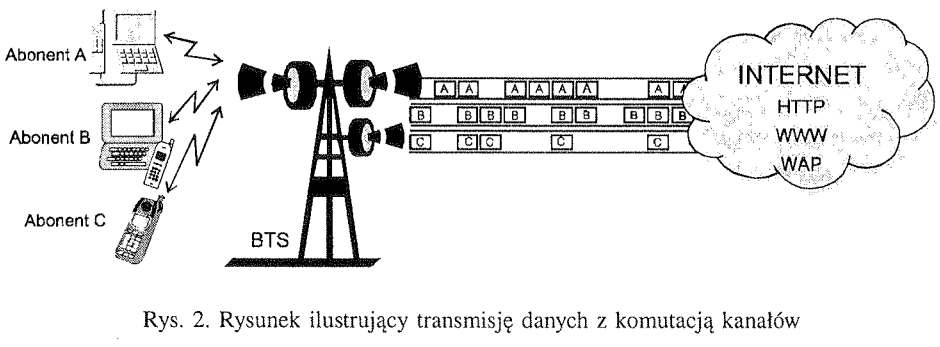
2.2. TRANSMISJA Z KOMUTACJA KANAŁÓW – CSD, HSCSD

Podstawowym sposobem transmisji danych, w czasie rzeczywistym, w sieciach komórkowych GSM jest transmisja danych z komutacją kanałów – CSD (Circuit Switched Data) [2]. Umożliwia ona transmisję z maksymalną przepływnością 9600 bit/s. Transmisja CSD umożliwia wymianę danych zarówno wewnątrz sieci GSM oraz między GSM a innymi sieciami telekomunikacyjnymi i informatycznymi.

W sieciach GSM drugiej generacji zwiększono maksymalną prędkość transmisji danych z komutacją kanałów z 9600 bits/s w trybie CSD do 57,6 kbits/s w trybie HSCSD (High Speed Circuit Switched Data).

Podstawowa transmisja danych w GSM (CSD) wykorzystuje po jednej szczelinie czasowej do nadawania i do odbioru, w pojedynczej ramce TDMA, co umożliwia transmisję z maksymalną prędkością 9600 bits/s. W trybie HSCSD, poprzez zastosowanie efektywniejszej metody kodowania, zwiększono przepływność transmisji danych w pojedynczej szczelinie czasowej do 14,4 kbits/s. Ponadto HSCSD umożliwia jednoczesne wykorzystanie kilku szczelin czasowych, zarówno dla danych nadawanych jak i odbieranych. Teoretyczna maksymalna przepływność wynosić 115,2 kbits/s,

przy wykorzystaniu 8 szczelin czasowych i kodowaniu 14,4 kbits/s. W praktyce można wykorzystać maksymalnie 4 szczeliny czasowe dla każdego kierunku, co daje 57,6 kbits/s. Tworzone połączenia mogą być niesymetryczne, gdy wykorzystujemy różną liczbę szczelin czasowych dla transmisji „w dół" i „w górę". Szybkości możliwe do uzyskania w HSCSD zależą przede wszystkim od operatora sieci GSM oraz od możliwości posiadanego terminalu GSM. Rozpoczęcie transmisji z komutacją kanałów wymaga nawiązania połączenia telefonicznego pomiędzy urządzeniami nadawczym i odbiorczym. Na cały czas trwania połączenia zostaje przydzielony osobny kanał transmisyjny, niezależnie od ilości i częstości transmitowanych danych (rys. 2).



Rys. 2. Rysunek ilustrujący transmisję danych z komutacją kanałów

Fig. 2. Switching channels data transmission

2.3. PAKIETOWA TRANSMISJA DANYCH – GPRS

Transmisja pakietowa GPRS (General Packet Radio Service) została wprowadzona w roku 2000, w systemie GSM drugiej generacji w tzw. fazie 2+. W trybie transmisji GPRS wszystkie transmitowane dane są dzielone na pakiety. Każdy pakiet danych zawiera adres przeznaczenia, co umożliwia transmisję pakietów nadawanych przez różnych użytkowników za pomocą współdzielonych kanałów transmisyjnych. Pakiety mogą być przenoszone niezależnie, różnymi trasami, zwykle z pewnym opóźnieniem. Pakietowa transmisja GPRS umożliwia zarówno transmisję punkt – punkt jak i transmisję rozsiewczą: punkt – wiele punktów.

Maksymalne prędkości transmisji w GPRS obecnie wynoszą: 53,6 kbit/s dla danych odbieranych przez stacje ruchomą i 26,8 kbit/s dla danych nadawanych. Są to maksymalne prędkości transmisji deklarowane przez operatorów sieci GSM oraz obsługiwane przez dostępne na rynku terminale. Teoretyczna maksymalna przepływność wg dokumentów ETSI wynosi 171,2 kbits/s. Prędkość transmisji danych w GPRS jest uzależniona od dwóch parametrów: od schematu kodowania kanałowego, określającego prędkość transmisji w pojedynczej szczelinie ramki TDMA oraz od liczby szczelin w ramce TDMA, które można jednocześnie wykorzystać [1]. W trybie GPRS definiuje się cztery schematy kodowania kanałowego o różnych przepływnościach (tabela 1). Większa przepływność jest wynikiem zastosowania mniejszej liczby bitów przezna-

czonych do korekcji błędów. Wybór sposobu kodowania przez stację bazową BTS zależy głównie od jakości interfejsu radiowego. Obecnie, w praktyce wykorzystuje się jedynie schematy CS-1 oraz CS-2.

Tabela 1

Schemat kodowania zdefiniowane w GPRS

The GPRS coding schemas

Schemat kodowania	CS-1	CS-2	CS-3	CS-4
maksymalna przepływność pojedynczego kanału [kbit/s]	**9,05**	**13,4**	**15,6**	**21,4**

Drugim parametrem transmisji GPRS odpowiedzialnym za przepływność jest klasa transmisji wielokanałowej (tzw. multislot class), którą może obsługiwać stacja ruchoma. Określa ona liczbę kanałów, które stacja może wykorzystywać do wysyłania i odbierania danych (tabela 2) w pojedynczej ramce TDMA.

Tabela 2

Przykładowe klasy transmisji wielokanałowej

The example of multislot class transmission

Klasa transmisji wielokanałowej	Maksymalna liczba kanałów (slotów)		
	w dół (Rx)	w górę (Tx)	suma
1	1	1	2
4	**3**	**1**	**4**
5	**2**	**2**	**4**
6	**3**	**2**	**4**
8	**4**	**1**	**5**
10	**4**	**2**	**5**
11	4	3	5
12	4	4	5
18	8	8	n.d.
19	6	2	n.d.
20	6	3	n.d.
29	8	8	n.d.

Aktualnie dostępne na rynku terminale GSM głównie należą do następujących klas transmisji wielokanałowej: 4, 5, 6, 8 i 10 oraz mogą, w większości, obsługiwać dowolny schemat kodowania.

Szybkość i jakość transmisji zależy również od jakości usługi QoS (ang. Quality of Service). W QoS zdefiniowano cztery klasy opóźnień (tabela 3) [3].

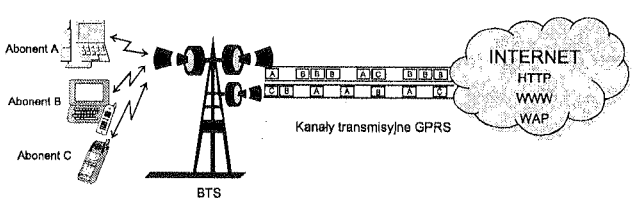
Tabela 3

Klasy opóźnień zdefiniowane dla GPRS

The delay classes defined for GPRS

Klasa	Pakiet 128 bitów		Pakiet 1024 bity	
	Średnie opóźnienie	95% opóźnienie	Średnie opóźnienie	95% opóźnienie
1	< 0,5 s	< 1,5 s	< 2 s	< 7 s
2	< 5 s	< 25 s	< 15 s	< 75 s
3	< 50 s	< 250 s	< 75 s	< 375 s
4	Najmniejsze możliwe	Najmniejsze możliwe	Najmniejsze możliwe	Najmniejsze możliwe

Najistotniejszą cechą różniącą transmisję z komutacją kanałów HSCSD, a transmisję z komutacją pakietów GPRS, pod względem możliwości ich zastosowania w systemach pomiarowych, jest sposób wykorzystania dostępnych zasobów sieciowych. W HSCSD kanał transmisyjny jest przydzielony abonentowi na wyłączność przez cały czas trwania połączenia, niezależnie od jego aktywności (ilości wymienianych danych) i w tym czasie nie jest dostępny dla innych użytkowników sieci. GPRS umożliwia współdzielenie tych samych kanałów transmisyjnych przez wielu abonentów. Dany użytkownik wykorzystuje zasoby sieci tylko na czas wymiany danych, w pozostałym czasie zasoby te są dostępne dla innych użytkowników sieci (rys. 3). Takie rozwiązanie pozwala zmienić sposób naliczania opłat za transmisję: w przypadku GPRS abonent płaci za ilość wysłanych i odebranych danych, z kolei korzystając z transmisji z komutacją kanałów (CSD, HSCSD) abonent płaci za całkowity czas połączenia. To pozwala użytkownikowi wykorzystującemu transmisję GPRS na stałe połączenie z siecią, adres IP przydzielony jego terminalowi jest stale widoczny w sieci.



Rys. 3. Ilustracja transmisji danych pakietowych GPRS

Fig. 3. Packet data transmission – GPRS (General Packet Radio Service)

Istotną zaletą technologii GPRS jest reakcja na przeciążenie sieci GSM, polegająca na automatycznym zmniejszeniu przepływności poszczególnych połączeń, a nie bloko-

waniem dostępu do sieci lub zerwaniem istniejącego połączenia – jak ma to miejsce w tradycyjnych sieciach GSM z komutacją kanałów (CSD, HSCSD).

2.4. PRZYSPIESZONA, RADIOWA TRANSMISJA DANYCH – EDGE

Technologia EDGE (ang. Enhanced Data Rates for Global Evolution) stanowi pośredni etap w ewolucji cyfrowych komunikacji bezprzewodowej z drugiej generacji (GSM/GPRS) do trzeciej (UMTS) [4]. W wyniku zmiany sposobu modulacji i kodowania kanałowego zwiększono maksymalną, teoretyczną prędkość transmisji danych do 473,6 kbits/s. Oprócz stosowanej w GSM modulacji dwuwartościowej GMSK (ang. Gaussian Minimum Shift Keying) w EDGE wykorzystuje się również ośmiowartościową modulację 8-PSK (ang. Phase Shift Keying). Modulacja 8-PSK jest bardziej efektywna i pozwala przesłać dane z maksymalną prędkością 59,2 kbits/s w pojedynczej szczelinie czasowej, a w przypadku wykorzystania modulacji GMSK wartość ta wynosi 17,2 kbits/s. Ze względu na zmienną jakość kanału radiowego wprowadzono 9 schematów kodowania i modulacji (tabela 4). Procedura sterowania jakością łącza radiowego LQS (ang. Link Quality Control) pozwala na adaptacyjną zmianę wykorzystywanego schematu kodowania i modulacji w zależności od jakości łącza radiowego, optymalizując przepływność w pojedynczej szczelinie czasowej.

Tabela 4

Modulacja i schematy kodowania w EDGE

Modulation and coding schemas in EDGE

Schemat	Typ modulacji	Przepływność szczeliny [kbits/s]	Maksymalna teoretyczna przepływność [kbits/s]	Sprawność kodowania
MCS-9	8PSK	59,2	473	1,0
MCS-8	8PSK	54,4	435	0,92
MCS-7	8PSK	44,8	358	0,76
MCS-6	8PSK	27,2	234	0,49
MCS-5	8PSK	22,4	179,2	0,37
MCS-4	GMSK	17,2	141	1,0
MCS-3	GMSK	13,6	119	0,80
MCS-2	GMSK	11,2	90	0,66
MCS-1	GMSK	8,8	70,4	0,53

EDGE wykorzystuje taki sam zakres częstotliwości oraz kanały radiowe oddalone o 200 kHz jak standard GSM. Takie rozwiązanie ułatwia współdziałanie tradycyjnych sieci GSM z sieciami GSM/EDGE oraz upraszcza proces implementacji technologii EDGE w już istniejących sieciach GSM. Wprowadzenie technologii EDGE w sieciach GSM wymaga wprowadzenia modernizacji stacji bazowych BTS oraz dokonania

istotnych zmian w oprogramowaniu zarządzającym kanałami radiowymi i usługami. Aby nowa technologia była dostępna dla abonentów muszą oni posiadać odpowiednie terminale mobilne obsługujące technologie EDGE.

Technologia transmisji danych EDGE jest obecnie oferowana przez wszystkich, trzech polskich operatorów sieci GSM na terenie większych miast Polski. Dostępne na rynku terminale umożliwiają transmisję danych w technologii EDGE z maksymalną przepływnością 247 kbits/s dla danych pobieranych (downlink) oraz 123 kbits/s dla danych wysyłanych (uplink). Terminale te oprócz technologii EDGE obsługują również technologię transmisji GPRS. Przełączanie pomiędzy technologiami transmisji GPRS i EDGE odbywa się automatycznie, bez ingerencji użytkownika, w zależności od ich aktualnej dostępności. Znajdując się na obszarze zasięgu EDGE terminal wykorzystuje tą technologię do transmisji, a opuszczając obszar zasięgu EDGE, terminal przechodzi w tryb transmisji GPRS.

Standard EDGE zaimplementowany w obecnych sieciach GSM pozwala operatorom sieci na dostarczenie usług typowych dla UMTS.

2.5. UNIWERSALNY SYSTEM TELEFONII MOBILNEJ – UMTS

System UMTS (ang. Universal Mobile Telecommunications System) stanowi sieć komórkową trzeciej generacji, pracującej w paśmie 2 MHz. Najważniejszą cechą UMTS jest znaczące zwiększenie prędkości transmisji danych do 2 Mbits/s. UMTS jest systemem komunikacji globalnej, łączącym naziemne systemy telefonii komórkowej z satelitarnymi systemami komunikacji mobilnej. Na obszarach zaludnionych są wykorzystywane systemy naziemne, a na obszarach niezamieszkałych i słabo zaludnionych (np. morza, oceany, pustynie, rozległe tereny górskie i leśne gdzie nie istnieje infrastruktura naziemna) będzie wykorzystywana łączność satelitarna. Globalny zasięg, zróżnicowane systemy komunikacji oraz duża niezawodność systemu UMTS wymuszają hierarchiczny podział terytorialny na strefy o określonej wielkości (tabela 5). Najmniejszą strefę stanowi pikokomórka – obszar o promieniu nieprzekraczającym 100 m, zwykle zawierający się wewnątrz pojedynczego budynku np. biurowce, hotele, dworce, lotniska, itp. oraz w miejscach o największym natężeniu ruchu telekomunikacyjnego. Odbiorcami są abonenci stacjonarni oraz piesi. Maksymalna prędkość transmisji w pikokomórce ma sięgać 2 Mbits/s. Większa strefa to mikrokomórka – obszar o promieniu do 1 km umieszczony na terenach miejskich np. małe osiedla mieszkaniowe, ulice miast, stadiony itp. Komunikacja w mikrokomórce jest możliwa z abonentami stacjonarnymi oraz poruszającymi się z prędkościami do 100 km/h. Wraz ze wzrostem odległości od stacji bazowej oraz zwiększanie prędkości przemieszczania się terminala maleje maksymalna prędkość transmisji. Kolejną strefę stanowi makrokomórka, jest to obszar o promieniu do 35 km. Makrokomórki mogą obejmować np. dzielnice miast lub całe miasta, główne trasy komunikacyjne (autostrady), umożliwiają komunikację z abonentami poruszającymi się z dużymi prędkościami (do 500 km/h). Największą strefą jest megakomórka (hiperkomórka). Obszar megakomórki ma promień od 100 do

5000 km i obejmuje obszary o umiarkowanym trafiku oraz obszary słabozaludnione. Na obszarze megakomórki komunikacja może być realizowana za pomocą łączności naziemnej oraz satelitarnej. Na obszarze megakomórki maksymalna przepływność nie powinna spaść poniżej wartości 144 kbits/s.

Tabela 5

Hierarchia komórek radiowych w UMTS

Hierarchy radio cells in UMTS

Typ komórki	Promień komórki [km]	Charakter komórki	Maksymalna szybkość przemieszczania się terminala [km/h]	Maksymalna przepływność transmisji
Pikokomórka	≤ 0,1	Obszary o dużym trafiku, wewnątrz budynków: biurowce, dworce, lotniska	≤ 10	2 Mb/s
Mikrokomórka	≤ 1	Otwarte tereny o dużym trafiku: osiedla, ulice miast, stadiony	≤ 100	384 kb/s ÷ 2 Mb/s
Makrokomórka	≤ 35	Dzielnice miast, miasta, główne arterie komunikacyjne (autostrady)	≤ 500	144 kb/s ÷ 384 kb/s
Megakomórka	100 ÷ 5000	Rozległe obszary o małym trafiku: regiony, kraje	Bez limitów	144 kb/s

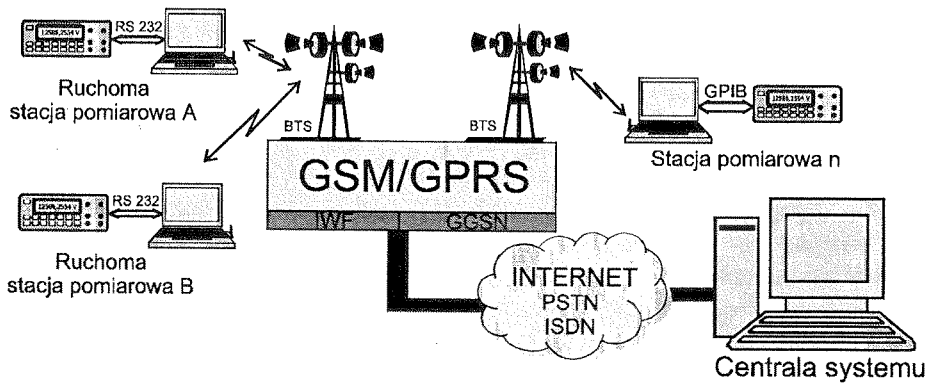
Wraz z rozwojem UMTS będą dostępne nowe usługi telekomunikacyjne oraz ulepszone te oferowane przez obecne systemy komunikacji mobilnej. Wśród wielu różnorodnych usług oferowanych przez systemy UMTS znajdziemy: telefonię, szybką transmisję danych, telekonferencje, wideorozmowy, wideokonferencje, mobilny Internet, dostęp do baz danych, gazetę elektroniczną pozycjonowanie i nawigację w terenie.

Technologia UMTS została, po raz pierwszy w Polsce, udostępniona komercyjnie przez sieć Plus GSM we wrześniu 2004 roku. Wśród pierwszych, dostępnych usług, charakterystycznych dla UMTS, znalazły się szybka transmisja danych oraz wideorozmowa. Początkowo UMTS był dostępny jedynie w centrum Warszawy.

3. SYSTEMY POMIAROWE Z TRANSMISJA DANYCH W SIECIACH GSM

Usługi transmisji danych dostępne w sieciach komórkowych GSM oraz UMTS można wykorzystać w rozproszonych systemach pomiarowych [5]. Na rysunku 4 pokazano schemat przykładowego systemu pomiarowego z transmisją danych w sieci GSM. System taki może składać się z kilku stacji pomiarowych, mobilnych lub stacjonar-

nych, wyposażonych w modemy GSM, które umożliwiają im nawiązywanie połączeń i wymianę informacji z centralą systemu. Transmisja danych pomiarowych w systemie może przebiegać za pomocą łączy telekomunikacyjnych (PSTN, ISDN), niepublicznych sieci komputerowych oraz Internetu. Wybór sposobu komunikacji w systemie zależy od usługi transmisji danych w GSM. Wykorzystując transmisję z komutacją kanałów (CSD, HSCSD), centrala systemu może łączyć się ze stacjami pomiarowymi za pomocą sieci informatycznej, Internetu lub przez nawiązanie połączenia telefonicznego z wybraną stacją przez dowolną sieć telefoniczną: komórkową, stacjonarną analogową lub cyfrową. System z pakietową transmisją danych (GPRS, EDGE, UMTS) wymaga połączenia centrali systemu przez prywatną sieć teleinformatyczną lub przez Internet. W przypadku transmisji przez Internet centrala systemu lub stacje pomiarowe powinny posiadać stały, publiczny adres IP, który umożliwi ich wzajemną lokalizację w sieci.



Rys. 4. Przykładowy system pomiarowy z transmisją danych w sieci GSM/GPRS

Fig. 4. Example of he measuring system with data transmission in GSM/GPRS cellular network

Wymiana danych pomiędzy centralą a stacjami za pomocą sieci komputerowych i Internetu wymaga stosowania odpowiednich, sieciowych protokołów transmisji do najpopularniejszych należą protokoły: TCP oraz UDP [6].

Protokół TCP (Transmission Control Protocol) jest protokołem transmisji gwarantowanej, zorientowanym połączeniowo. Umożliwia zestawienie połączenia, w którym dane są niezawodnie przesyłane. Niezawodność transmisji jest realizowana za pomocą metody nazwanej retransmisją z potwierdzeniem PAR (Positive Acknowledgment with Retransmission). Odbiorca, każdorazowo po otrzymaniu pakietu danych, wysyła nadawcy potwierdzenie poprawnie odczytanego pakietu. Nadawca, przez określony czas, oczekuje na potwierdzenia odbioru i po jego otrzymaniu wysyła kolejne pakiety danych. Jeżeli potwierdzenie nie nadejdzie, w ciągu określonego czasu, nadawca ponowne wysyła dane. Protokół TCP zapewnia całkowitą obsługę transmisji danych w warstwie transportu w modelu warstw TCP/IP. TCP posiada mechanizmy pozwalające dzielić większe porcje danych na segmenty przed ich wysłaniem oraz ponownie je składać po stronie odbiorczej, również wówczas, gdy poszczególne segmenty danych

docierają w niewłaściwej kolejności. Do popularnych aplikacji protokołu TCP należą: FTP, Telnet, serwery pocztowe SMTP, POP serwery WWW oraz HTTP.

Zastosowanie protokołu TCP w systemach pomiarowych pozwala na znaczne uproszczenie oprogramowania, ponieważ TCP całkowicie obsługuje transmisję, programista nie musi rozbudowywać oprogramowania o mechanizmy kontrolne. Stosowanie protokołu TCP wymaga wysłania - oprócz danych - także informacji sterujących, zapewniających kontrolę procesu komunikacji, co powoduje zmniejszenie efektywności transmisji i prędkość transmisji jednocześnie zwiększając obciążenia łącza transmisyjnego.

Protokół Datagramów Użytkownika UDP (User Datagram Protocol) realizuje wymianę danych bez gwarancji odbioru. UDP realizuje transmisję bez wcześniejszego nawiązywania połączenia między stacjami wymieniającymi dane. Protokół ten nie posiada mechanizmów potwierdzania dostarczanych danych, dlatego oprogramowanie systemów wykorzystujących UDP musi zapewnić: retransmisje zagubionych i uszkodzonych danych, fragmentację i ponowne składanie większych strumieni danych. Nagłówek UDP jest wyposażony w sumę kontrolną, która umożliwia, sprawdzenie czy odczytany datagram nie został uszkodzony podczas transportu. Zaletą UDP jest małe obciążenie transmitowanych danych informacjami kontrolnymi, co wraz z brakiem potwierdzenia pozwala na efektywniejsze, niż w przypadku TCP, wykorzystanie dostępnych zasobów sieciowych przy jednoczesnym mniejszym ich obciążeniu. Wśród popularnych aplikacji wykorzystujących UDP są: TFTP, protokoły zarządzania siecią: SNMP, DNS oraz aplikacje typu RealAudio (np. internetowe stacje radiowe i telewizyjne).

W systemach pomiarowych, gdzie dużą rolę odgrywa poprawność transmisji, lepszym wyborem jest zastosowanie protokołu TCP.

O wyborze sposobu transmisji z komutacją łączy (CSD, HSCSD) czy pakietów (GPRS, EDGE, UMTS) decyduje charakter systemu pomiarowego (szczególnie ilość i częstość transmitowanych danych) oraz dostępność usług transmisji na obszarze działania systemu.

Cechy pakietowej transmisji danych predestynują ją do zastosowania w systemach pomiarowych, w których wymagana jest stała, ciągła łączność stacji pomiarowych z centralą, a dane transmitowane są nieokresowo lub często w niewielkich pakietach. Przykłady zastosowań pakietowej transmisji danych GPRS/EDGE w systemach pomiarowych:

- W pomiarach meteorologicznych, automatyczna akwizycja wyników pomiarów temperatury, wilgotności, siły i kierunku wiatru, itp. z rozmieszczonych na dużym obszarze lub poruszających się stacji pomiarowych,
- Do monitorowania alarmowego, np. przekroczenia poziomów alarmowych zbiorników wodnych, wykrywanie pożarów na dużych obszarach leśnych,
- Do sygnalizacji antywłamaniowej, przez automatyczne informowanie o włamaniu odpowiednich służb publicznych,
- W pomiarach geodezyjnych, kartograficznych do bezpośredniej transmisji danych z odbiorników systemu nawigacji satelitarnej GPS,

- Do nadzoru floty pojazdów w przedsiębiorstwach kurierskich, transportowych, komunikacji publicznej,
- W energetyce do nadzoru stacji transformatorowych, rozdzielczych, itp.

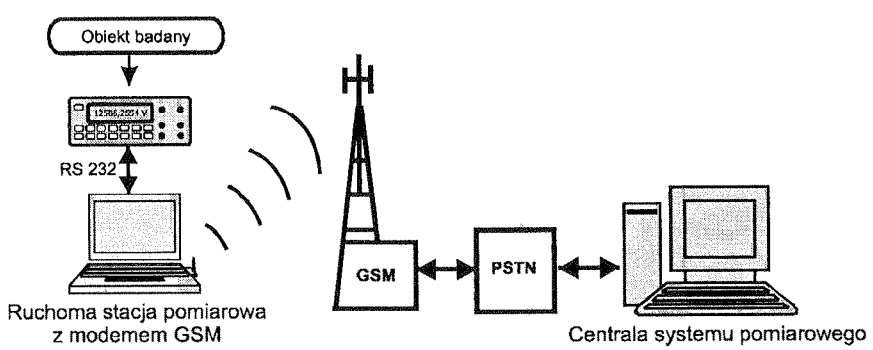
Szybka komutowana transmisja danych (HSCSD) może być wykorzystana w systemach niewymagających ciągłego połączenia pomiędzy centralą systemu a stacjami pomiarowymi, połączenie jest nawiązywane tylko na czas przesłania danych pomiarowych i instrukcji sterujących. HSCSD można stosować w systemach transmitujących dane strumieniowe, wymagające stałej, dużej prędkości transmisji w czasie rzeczywistym, np.:

- pomiary biomedyczne np. zdalna transmisja sygnału EKG,
- transmisja danych pomiarowych ze zdalnych mierników.

4. SYSTEMY POMIAROWE ZBUDOWANE I SPRAWDZONE W POLITECHNICE POZNAŃSKIEJ

4.1. SYSTEM Z KOMUTOWANĄ TRANSMISJĄ DANYCH: HSCSD – PSTN

Pierwszym systemem pomiarowym z transmisją danych przez sieć GSM, zbudowanym przez autora w Politechnice Poznańskiej był system z komutowaną transmisją w trybie HSCSD (rys. 5) [7, 8]. Opracowany system był zbudowany z centrali systemu oraz mobilnej stacji pomiarowej. Centralę systemu stanowił komputer PC dołączony do stacjonarnej sieci telefonicznej PSTN za pomocą analogowego modemu V.90. Stację pomiarową tworzył laptop wyposażony w modem HSCSD oraz przyrząd pomiarowy. Stosowano multimetry cyfrowe dołączane do laptopa łączem RS232 oraz kartę akwizycji danych: DAQCard-6024E firmy National Instruments. Oprogramowanie centrali i stacji napisano w graficznych środowiskach programistycznych (HP VEE 5.0 oraz LabView 6i), przeznaczonych do tworzenia aplikacji kontrolno-pomiarowych.



Rys. 5. Schemat blokowy systemu pomiarowego z transmisją przez GSM i PSTN

Fig. 5. Block scheme of the measuring system with HSCSD and PTSN transmission

Po nawiązaniu połączenia telefonicznego między komputerami (centralą systemu a stacją pomiarową) jest możliwa dwukierunkowa wymiana danych pomiarowych i instrukcji sterujących między centralą systemu a przyrządem pomiarowym.

Przedstawiony system pomiarowy badano w celu sprawdzenia jego dynamiki (czas nawiązania połączenia, maksymalna przepływność danych) oraz poprawności transmisji, czyli stabilności połączeń i błędów powstających w transmitowanych pakietach.

Wykorzystując multimetry cyfrowe do realizacji pomiarów, uzyskano maksymalną szybkość transmisji 20 pomiarów na sekundę. Stosunkowo niskie prędkości transmisji były spowodowane przez niską przepustowość łącza RS 232 (9600 bitów/s) łączącego multimetry z komputerem w stacji pomiarowej. Dodatkowo każdy transmitowany wynik pomiaru miał objętość kilkunastu bajtów.

Tych ograniczeń nie posiadał system, w którym do akwizycji pomiarów zastosowano kartę pomiarową, wykorzystującą 12 bitowy przetwornik A/C, o maksymalnej częstotliwości próbkowania 200 kHz. Zastosowanie karty pomiarowej w miejscu multimetrów pozwoliło na znaczne zwiększenie szybkości rejestrowanych i transmitowanych pomiarów, ponadto wynik pojedynczego pomiaru zawierał się w dwóch bajtach, co dodatkowo poprawiło efektywność transmisji. Przesyłając, w czasie rzeczywistym, pojedyncze wyniki pomiarów uzyskano maksymalną szybkością na poziomie 50 pomiarów na sekundę. Transmitując wyniki pomiarów w pakietach uzyskano znacznie większe prędkości próbkowania (tabela 6). W czasie rzeczywistym, bez opóźnień, transmitowano wyniki pomiarów próbkowanych z maksymalną częstotliwością 1639 Hz.

Tabela 6

Maksymalne częstotliwości próbkowania, którego wyniki transmitowano bez opóźnień

Maximum sampling rates whose results transmission without delays

Liczba pomiarów w pakiecie	1	10	100	1000
maksymalna częstotliwość próbkowania [S/s]	**49,8**	**490**	**1612**	**1639**

Badania systemu pomiarowego wykonywano przy różnym natężeniu ruchu w sieci GSM zarówno „w szczycie" jak i „po szczycie", w ciągu całego tygodnia (również w soboty i niedziele). W ten sposób sprawdzano wpływ natężenia ruchu w sieci GSM na zmiany prędkości transmisji. Zauważono, że ustawiana maksymalna, symetryczna prędkość połączenia 28,8 kbit/s była utrzymywana przez cały czas trwania połączenia. Spadki prędkości transmisji do 14,4 kbit/s występowały sporadycznie i niezależnie do pory nawiązywania połączenia. Czas nawiązywania połączenia wynosił od kilkunastu do 30 sekund. Połączenia między centralą systemu a stacją pomiarową trwały maksymalnie kilka minut. Wszystkie nawiązane połączenia były stabilne przez cały czas trwania i nie dochodziło do ich nieprzewidzianego zrywania, ponadto nie stwierdzono żadnych błędów w transmisji. Wszystkie wysłane przez stację pomiarową pakiety danych dotarły bezbłędnie do centrali. Jedynym mechanizmem zastosowanym do kon-

troli transmisji był bit kontroli parzystości w poszczególnych ramkach transmitowanych danych.

4.2. SYSTEMY POMIAROWE Z TRANSMISJĄ DANYCH PRZEZ SIEĆ GSM ORAZ INTERNET

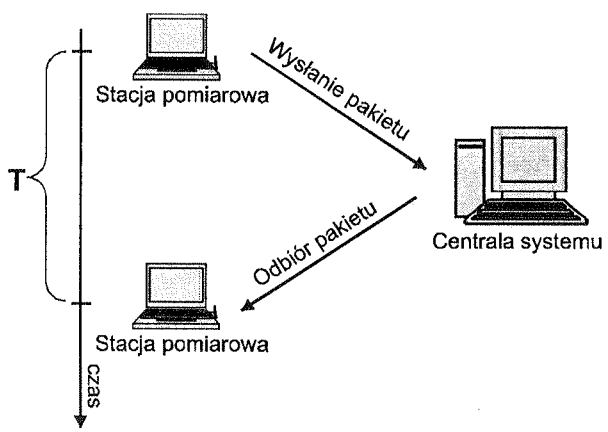
Przed budową systemów pomiarowych wykorzystujących transmisję przez sieci GSM i Internet zmierzono czas transmisji pakietów danych w systemie (tabela 7). Program uruchomiony w stacji pomiarowej cyklicznie wysyłał do centrali pakiety o zadanej długości, centrala natychmiast po odebraniu pakietu retransmitowała go powrotnie do stacji wysyłającej. Rejestrowano czas T (rys. 6) tej operacji. Przyjęto połowę czasu T jako czas transmisji pakietu w systemie. Na rysunku 7 przedstawiono wykres ilustrujący zarejestrowane czasy transmisji pakietów o różnych długościach. Na wykresie pokazano średni czas transmisji pakietu, o zadanej długości, obliczony z serii 100 pomiarów, oraz dodatkowo najkrótszy i najdłuższy czas transmisji w serii.

Tabela 7

Czasy transmisji pakietów w systemie

Times the transmission of the packets in the system

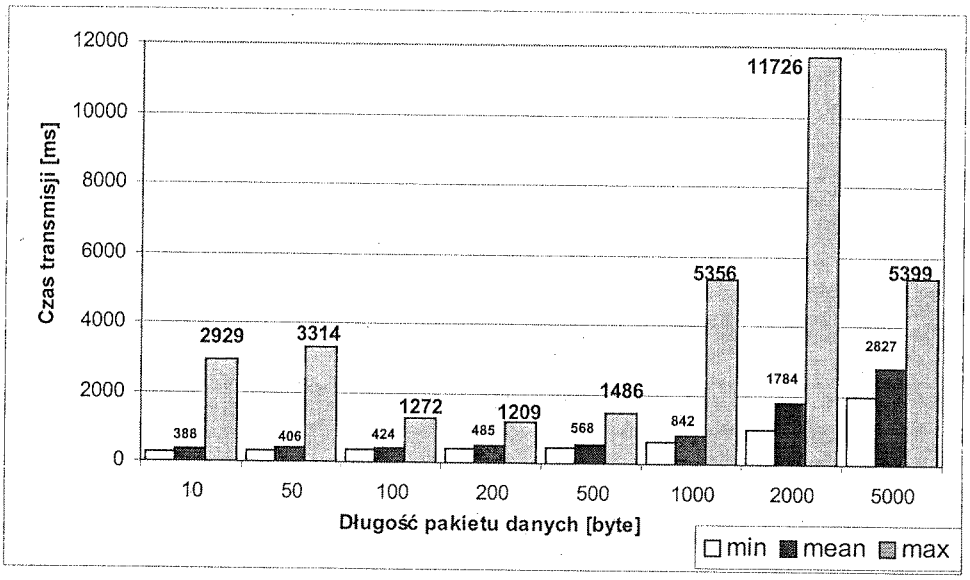
	Długość pakietu								
	10	**50**	**100**	**200**	**500**	**1000**	**2000**	**5000**	**[byte]**
najkrótszy czas. trans.	294	314	359	405	477	655	1028	1981	[ms]
średni czas trans.	**388,3**	**406**	**424**	**485**	**568**	**842,3**	**1784**	**2827**	**[ms]**
najdłuższy czas trans.	2929	3314	1272	1209	1486	5356	11726	5399	[ms]



Rys. 6. Pomiar czasu transmisji pakietów w systemie

Fig. 6. Measuring of the time transmission of packets in a system

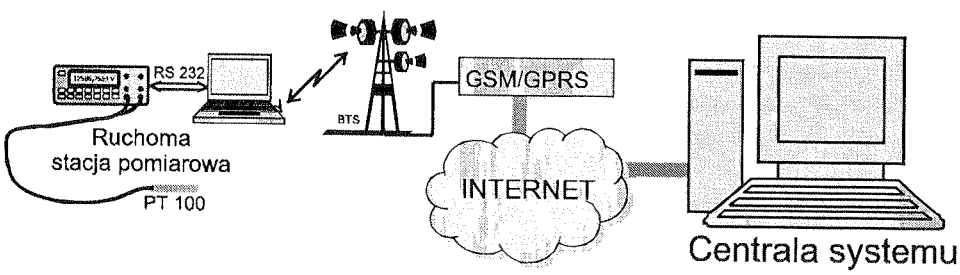
Analizując wyniki przedstawione w tabeli 7 oraz na wykresie (rys. 7) zauważono, że minimalny i średni czas transmisji pakietu wzrasta wraz z wielkością pakietu. Maksymalne czasy transmisji nie zależą od długości pakietu, ale od chwilowego natężenia trafiku w Internecie, który ma charakter losowy.



Rys. 7. Wykres ilustrujący czasy transmisji pakietów danych o zadanych długościach

Fig. 7. Timing of the data packet transmission with given lengths

Opracowano i sprawdzono dwa systemy pomiarowe z transmisją przez GSM i Internet [9]. Pierwszy z nich stanowił praktyczna realizację systemu przeznaczonego do zdalnych pomiarów temperatury (rys. 8). Drugi służył do określenia dynamicznych własności systemów pomiarowych wykorzystujących transmisją danych za pomocą sieci komórkowych oraz Internetu.



Rys. 8. Schemat systemu pomiarowego przeznaczonego do zdalnego pomiaru temperatury

Fig. 8. The scheme of measuring system intended to remote temperature measurement

System, przeznaczony do zdalnego pomiaru temperatury, składał się z pojedynczej, przenośnej stacji pomiarowej wyposażonej w modem GPRS oraz z centrali systemu połączonej, przez sieć LAN, z Internetem.

Stacja pomiarowa, wykorzystując transmisję GPRS, łączyła się z punktem dostępowym do Internetu, udostępnionym przez operatora sieci GSM, a następnie przez Internet z centralą systemu. Po nawiązaniu połączenia, wysyłała, w czasie rzeczywistym, wyniki pomiarów temperatury. Transmisja przebiegała według protokołu TCP/IP.

Mobilną stację pomiarową, złożono z notebooka wyposażonego w modem GSM/GPRS klasy 10, multimetru cyfrowego dołączonego przez interfejs RS 232C oraz czujnika temperatury. Po nawiązaniu połączenia z centralą, praca stacji pomiarowej polegała na wykonywaniu, w określonych chwilach, czasu pomiaru temperatury i wysyłaniu wyników pomiaru. Na rysunku 9 przedstawiono ramkę transmitowanych danych. Numer pomiaru dołączono do każdego pakietu danych wysyłanego do centrali, w celu sprawdzenia czy wszystkie wysłane pakiety danych dotarły do centrali. W celu późniejszej kontroli wszystkie wysyłane dane były zapisywane w pliku.

Nagłówek TCP	Numer pakietu	Separator: „;”	Wynik pomiaru	Znacznik końca ramki: „\r\n”

Rys. 9. Format ramki transmitującej wynik pomiaru temperatury

Fig. 9. Result of temperature measurement – frame structure

Centralę systemu stanowi, odpowiednio oprogramowany, komputer stacjonarny klasy PC połączony z siecią Internet za pomocą łącza stałego o stałym, publicznym adresie IP. Zadaniem centrali systemu, po zgłoszeniu się stacji pomiarowej, jest rejestracja i wizualizacja przychodzących wyników pomiarów.

Przedstawiony system pomiarowy umożliwiał transmisję z prędkością do 10 wyników pomiarów w jednostce czasu. Po przeprowadzeniu serii prób i porównaniu danych nadanych przez stację pomiarową z danymi zarejestrowanymi przez centralę systemu nie stwierdzono błędów w transmisji – wszystkie wysłane dane zostały bezbłędnie odebrane przez centralę systemu.

Kolejny system pomiarowy, zbudowany w celu określenia dynamiki systemów z transmisją danych przez Internet i sieć GSM powstał w wyniku modyfikacji przedstawionego wcześniej systemu, która głównie polegała na zamianie urządzenia pomiarowego oraz zmianie oprogramowania. Do rejestracji pomiarów wykorzystano kartę pomiarową w miejscu multimetru cyfrowego. Stacja pomiarowa rejestrowała i transmitowała wyniki pomiarów napięcia. System mógł transmitować dane wykorzystując transmisję pakietową GPRS oraz transmisję z komutacją kanałów HSCSD. Oprogramowanie stacji umożliwiało rejestrację wyników pomiarów, o zadanych parametrach (częstotliwość próbkowania, rozdzielczość przetwarzania, zakres mierzonego napięcia) oraz transmisję wyników pomiarów pojedynczo lub w pakietach. Program pracował w pętli, cyklicznie rejestrując i wysyłając dane. Rejestrowane próbki są lokowane

w szeregowym buforze, z którego są usuwane po wysłaniu. Wyniki pomiarów przed wysłaniem uzupełniano informacjami kontrolnymi tj. numerem pakietu oraz liczbą próbek w pakiecie (rys. 10). Po każdym powtórzeniu pętli obserwowano liczbę pozostałych w pamięci (niewysłanych) próbek. Systematyczne zwiększanie się tej wartości świadczy o tym, że stacja nie nadąża transmitować, na bieżąco, danych rejestrowanych przez kartę pomiarową.

Nagłówek TCP	Numer pakietu (2 bajty)	Długość pola przeznaczonego na próbki (2 bajty)	Próbki (0 – 65 536 bajty)

Rys. 10. Format pakietu z danymi pomiarowymi

Fig. 10. Frame of the measurement data

Po przesłaniu serii pakietów danych obliczano średnią prędkość transmisji danych w systemie. Średnią prędkość transmisji obliczano z zależności (1), jako stosunek liczby wszystkich wysłanych bajtów danych do całkowitego czasu transmisji serii pakietów.

$$\bar{v} = \frac{N}{T}\,[byte/s], \tag{1}$$

gdzie: $\bar{v}$ – średnia prędkość transmisji,
N – liczba wszystkich wysłanych bajtów danych,
T – całkowity czas transmisji.

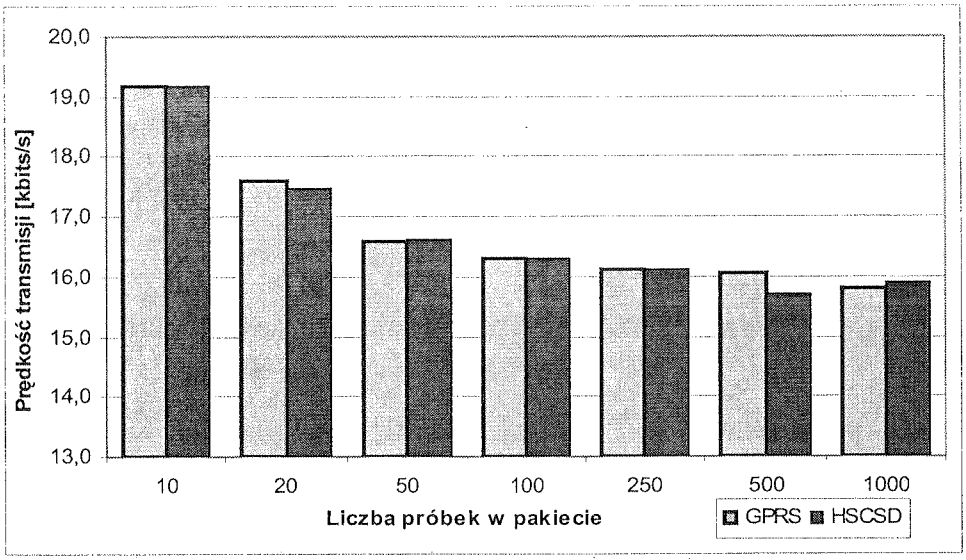
Otrzymane wyniki zaprezentowano w tabeli 8. Dane transmitowano w pakietach zawierających różną liczbę próbek. Wykorzystano technologię transmisji z komutacją kanałów HSCSD oraz transmisję pakietową GPRS. Badany sygnał próbkowano z częstotliwością 1 kHz. Pojedyncza próbka jest dwubajtowa, zatem przesłanie samego sygnału próbkowanego z częstotliwością 1 kHz wymaga pasma o przepływności 16 kHz, wyższe prędkości transmisji przedstawione w tabeli 8 wynikają z obecności w pakiecie dodatkowych czterech bajtów kontrolnych. Ilustracją tabeli 8 jest wykres przedstawiony na rysunku 11.

Tabela 8

Średnia prędkość transmisji danych (częstotliwość próbkowania wynosiła 1 kHz)

The average rate of data transmission in presented system (1 kHz frequency sampling)

	Liczba próbek pomiarowych w pakiecie						
	10	**20**	**50**	**100**	**250**	**500**	**1000**
GPRS [kbit/s]	19,19	17,60	16,58	16,32	16,13	16,06	15,79
HSCSD [kbit/s]	19,19	17,49	16,62	16,32	16,12	15,45	15,91



Rys. 11. Średnia prędkość transmisji danych w systemie w funkcji wielkości pakietu

Fig. 11. The average data rate in system vs. numbers of measurements in packet

Mierząc czas wykonania pętli, dokonywano pomiaru czasów wysyłania pojedynczych pakietów. Wyniki te przedstawiono w tabeli 9 i wykresach przedstawionych na rysunkach: 12 dla transmisji GPRS i 13 dla transmisji HSCSD). Wykresy przedstawiają wartości średnie i maksymalne.

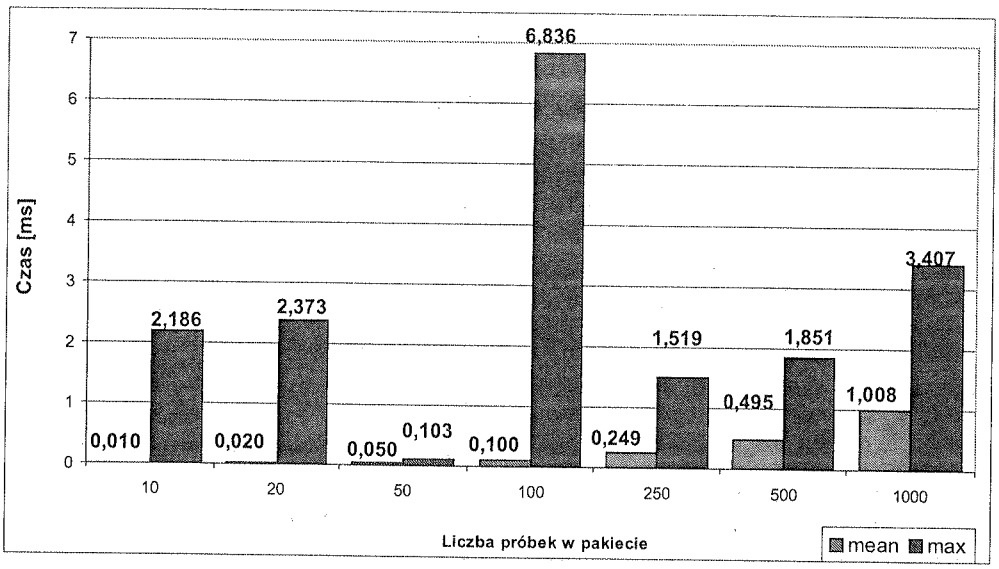
Tabela 9

Średnie i maksymalne wartości czasów transmisji pakietów w systemie

The average and maximum values of times of transmission packets in the system

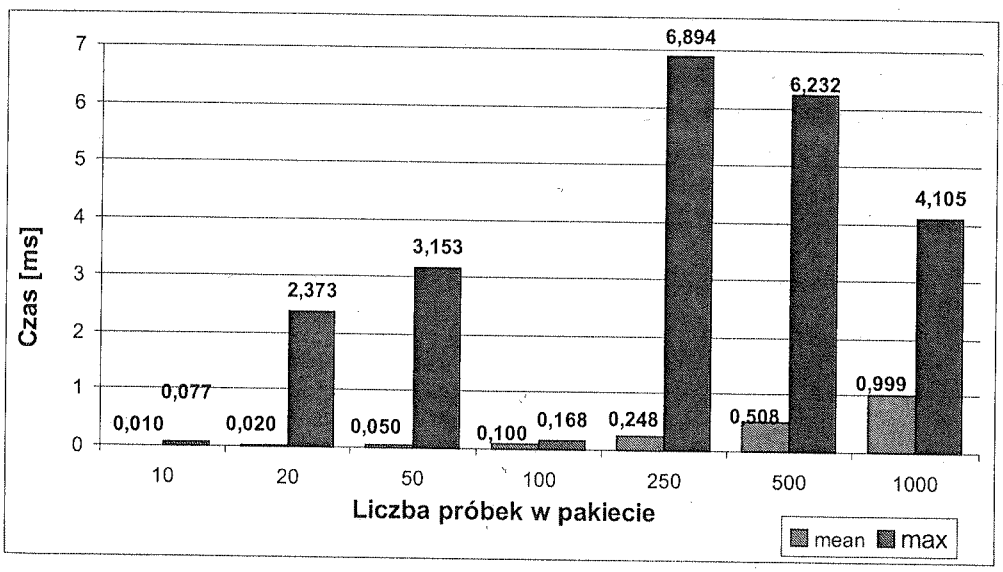
		Liczba próbek w pakiecie						
		10	**20**	**50**	**100**	**250**	**500**	**1000**
GPRS	mean [s]	0,010	0,020	0,050	0,100	0,249	0,495	1,008
	max [s]	2,186	2,373	0,103	6,836	1,519	1,851	3,407
HSCSD	mean [s]	0,010	0,020	0,050	0,100	0,248	0,508	0,999
	max [s]	0,077	2,373	3,153	0,168	6,894	6,232	4,105

Wartości średnie czasów wysyłania pakietów pokrywają się w obu sposobach transmisji (GPRS i HSCSD) z wartościami teoretycznymi, które możemy wyznaczyć dzieląc liczbę próbek w pakiecie przez częstotliwość próbkowania. Aby transmisja przebiegała na bieżąco czas wysyłania pakietów nie powinien być większy od tych wartości. Zwiększenie czasu transmisji pakietu powoduje powstawanie opóźnień w transmisji



Rys. 12. Czas wysyłania pojedynczego pakietu, transmisja GPRS

Fig. 12. Sending time the single packet in GPRS



Rys. 13. Czas wysyłania pojedynczego pakietu, transmisja HSCSD

Fig. 13. Sending time the single packet in HSCSD

i zwiększanie liczby nieodczytanych próbek w pamięci stacji. Ponieważ efektywna przepływność danych w Internecie ma charakter losowy w wyniku ciągłych zmian obciążenia sieci, niewielkie opóźnienia pojawiały się stosunkowo często. Po zakończeniu wysłania serii pakietów z danymi przez stację pomiarową, centrala systemu kończyła odbiór danych kilka, kilkanaście sekund później. Przedstawione na wykresie maksymalne, zarejestrowane czasy transmisji pojawiały się bardzo sporadycznie, niezależnie od zastosowanej technologii transmisji (GPRS, HSCSD), wielkości pakietu oraz pory dnia, w której realizowano pomiary.

W sposób analogiczny jak w systemie z transmisją danych pomiarowych przez łącze komutowane HSCSD i PSTN sprawdzono poprawność transmisji. Dane wysyłane przez stację rejestrowano, a następnie dokonano porównania z danymi zarejestrowanymi przez centralę systemu. W wyniku porównania nie stwierdzono błędów w transmisji i utraty danych.

Transmitowano sygnał próbkowany z maksymalną częstotliwością 1 kHz, przy wyższych częstotliwościach, w wyniku pojawiających się opóźnień w transmisji, system nie nadążał odczytywać i transmitować rejestrowanych próbek sygnału, co przy skończonej pojemności pamięci przeznaczonej na próbki powodowało utratę części danych.

5. PODSUMOWANIE

W artykule przedstawiono aktualnie dostępne usługi przeznaczone do transmisji danych oferowane przez sieci telefonii komórkowej GSM: przesyłanie wiadomości SMS, MMS oraz transmisję CSD, HSCSD, GPRS, EDGE oraz UMTS Usługi te znajdują szerokie zastosowanie w biznesie, usługach m-commerce, rozrywce, edukacji, transporcie oraz w przemyśle. Jedną z wielu możliwych aplikacji tych usług jest, przedstawione w artykule, wykorzystanie ich w systemach pomiarowych. Transmisja danych przez sieci GSM w rozproszonych systemach pomiarowych stanowi alternatywne rozwiązanie dla transmisji przewodowej oraz innych sposobów transmisji bezprzewodowej (transmisja radiomodemowa z licencjonowanymi łączami radiowymi, transmisja z łączem optycznym, laserowym), w sytuacjach, gdy budowa własnej infrastruktury teleinformacyjnej jest nieopłacalna lub niemożliwa.

Do niewątpliwych zalet systemów pomiarowych wykorzystujących transmisję w sieci GSM można zaliczyć: duży obszar działania (zasięg sieci GSM), mobilność, niskie koszty związane z wdrażaniem systemu pomiarowego oraz dużą elastyczność systemu umożliwiającą łatwą jego rekonfigurację.

Omówione systemy pomiarowe zostały opracowane, zbudowane i sprawdzone przez autora w Politechnice Poznańskiej. Systemy działały poprawnie zgodnie z założeniami.

Przedstawione wyniki pomiarów czasów transmisji danych w rzeczywistym systemie pozwalają na określenie opóźnień mogących wystąpić w transmisji wyników pomiarów w rozproszonych systemach pomiarowych, wykorzystujących sieci telefonii komórkowej GSM/GPRS oraz Internet do transmisji danych i instrukcji sterujących.

Dużą trudność stanowi oszacowanie teoretycznej wartości opóźnień transmisji, ponieważ ich wartość stale się zmienia w wyniku ciągłej zmiany trafiku. Doświadczalne wyznaczenie tych wartości pozwala określić ich wartości średnie i graniczne. Znajomość tych danych jest niezbędna dla projektantów systemów pomiarowych, których działanie odbywa się według określonych zależności czasowych.

Otrzymane wyniki charakteryzują się dużym rozrzutem wokół wartości średniej, dodatkowo zaobserwowano sporadyczne opóźnienia znacząco odbiegające od tej wartości (np. seria pomiarów z czwartku, transmisja pakietów o wielkości 1kB: przy wartości średniej wynoszącej poniżej 3 s został zarejestrowany pojedynczy pomiar opóźnienia o czasie ponad 68 s). Tak duże zmiany parametrów transmisji uniemożliwiają zastosowanie takiej transmisji danych w systemach pomiarowych czasu rzeczywistego.

Analizują zarejestrowane wyniki pomiarów czasów opóźnień w systemach z transmisją danych za pomocą Internetu, zauważono, że wyniki te są zbliżone zarówno dla transmisji komutowanej HSCSD oraz pakietowej GPRS. Jedyną istotną różnicą, jaką zauważono jest czas trwania połączenia między centralą systemu a stacją pomiarową. Wielokrotnie zestawiono połączenia z wykorzystaniem obu technologii transmisji. Czas trwania połączeń GPRS wynosił kilka godzin i nie zauważono przypadkowego zrywania połączenia, jak miało to miejsce przy transmisji HSCSD, wszystkie połączenia zrywały się po upływie kilkunastu minut, najdłuższe trwało 16 minut i 30 sekund. Autorowi nie udało się ustalić, co jest przyczyną zrywania połączeń. Uwzględniając to oraz koszty związane z transmisją danych, transmisja GPRS powinna znaleźć większe zastosowanie w praktycznych realizacjach systemów pomiarowych. Aktualnie wiele firm oferuje usługi telemetryczne oparte o ten sposób transmisji danych.

W swoich, przyszłych pracach autor planuje budowę i badania systemów pomiarowych wykorzystujących, obecnie wdrażane w Polsce, technologie EDGE oraz UMTS. Celem tych badań będzie między innymi oszacowanie rzeczywistych opóźnień w transmisji danych.

6. BIBLIOGRAFIA

1. A. Simon, M. Walczyk: *Sieci komórkowe GSM. Usługi i bezpieczeństwo*. Kraków, Xylab, 2002.
2. K. Wesołowski: *Systemy radiokomunikacji ruchomej*, wyd. 2 Warszawa, WKŁ, 1999.
3. K. Wesołowski: *Mobile Communication Systems*, West Sussex, Wiley, 2002.
4. A. Urbanek: *Vademecum Teleinformatyka II*. Warszawa, IDG Poland S.A., 2002.
5. W. Nawrocki: *Komputerowe systemy pomiarowe*. Warszawa, WKŁ, 2002.
6. B. Komar: *TCP/IP dla każdego*. Gliwice, Helion, 2002.
7. M. Maćkowski, W. Nawrocki: *Rozproszony system pomiarowy z transmisją danych w sieci GSM*. Elektronizacja, nr 3/2002, ss. 11-14.
8. M. Maćkowski: *Bezprzewodowa transmisja danych pomiarowych w sieci GSM*. Pomiary Automatyka Kontrola nr 7, 8/2002.
9. M. Maćkowski: *Zastosowanie pakietowej transmisji danych – GPRS w rozproszonych systemach pomiarowych*. Pomiary Automatyka Robotyka nr 7-8/2004.

M. MAĆKOWSKI

APPLICATION OF THE MEASURING SYSTEMS WITH DATA TRANSMISSION IN GSM CELLULAR NETWORK AND INTERNET

Summary

In this article distributed measuring systems using some GSM cellular network services were described. Such systems are alternative for wire systems, especially when measured object of measurement is located far from base station, also when object is located in difficult accessible terrain (large forest areas, jungles, national parks), also when object is moving over large area. Advantages such kind of systems are large range (area when GSM is available), mobility, low cost of introducing, flexibility and easily of reconfiguration. In measuring system can by use all available data transmission services known in GSM cellular networks. Selection one of them depend on final destinations and demands on data quantity, frequency and speed transmission in system. Currently GSM networks can to transmit data at maximum speed 26.4 kbit/s (GPRS) and 28.8 kbib/s (HSCSD) for data sending and adequately 53.6 kbit/s and 57.6 kbit/s for data receiving.

Possibilities and examples applications such kind of system were presented. Data transmission services, available in GSM network, were described. An application of such measuring system, designed and tested in Poznan University of Technology was shown. The results measurements of an effective data rate were presented.

Keywords: distributed measuring system, data transmission, GSM, GPRS, EDGE, UMTS, Internet

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 2006, 52, z. 2, ss. 267–285

Zastosowanie metody odpytywania wielocyklowego w cyfrowych interfejsach szeregowych opartych o architekturę master-slave

SŁAWOMIR ŻABA

Instytut Elektrotechniki i Elektroniki Przemysłowej, Politechnika Krakowska
ul. Warszawska 24, 33-155 Kraków
szaba@mail.pk.edu.pl

Otrzymano 2005.06.06
Autoryzowano 2005.09.09

We współczesnych urządzeniach elektronicznych wykorzystuje się połączenia sieciowe pomiędzy mikroprocesorami i innymi elementami układu. Jako architekturę protokołu sieciowego stosuje się najczęściej model master-slave. W artykule omówiono podstawowe zasady funkcjonowania trzech cyfrowych interfejsów szeregowych, pracujących w oparciu o ww. architekturę: I^2Cbus, SPI i 1-Wire. Następnie przedstawiono uogólniony model analizy czasowej takich systemów pod kątem spełnienia ograniczeń czasu rzeczywistego. Zaprezentowano także model odpytywania wielocyklowego jako alternatywę dla tradycyjnej metody odpytywania jednocyklowego stosowanej w magistralach o architekturze master-slave. Przedstawione zostały trzy algorytmy wytwarzające sekwencję odpytywania wielocyklowego: jednorodny oraz dwa priorytetowe – z użyciem metody Generalised Rate Monotonic Scheduling (GRMS) i Earliest Deadline First (EDF). Dokonano porównania tych trzech algorytmów, przedstawione wnioski zobrazowano przykładem.

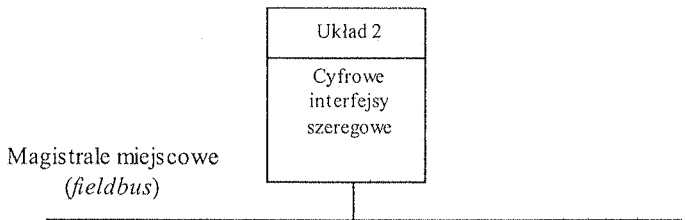
Słowa kluczowe: rozproszony system komputerowy, system czasu rzeczywistego, cyfrowe interfejsy szeregowe, odpytywanie wielocyklowe, metoda GRMS, EDF

1. WSTĘP

Współczesne koncepcje elektronicznych systemów sterowania polegają na rozpraszaniu sprzętu i inteligencji. Decentralizacja sprzętu na obszarze rzędu setek metrów czy kilometrów polega na zastosowaniu odpowiednich systemów transmisji danych bazujących na magistrali miejscowej (*fieldbus*) (rys.1). Magistrale miejscowe w odróżnieniu od zwykłych sieci komputerowych, spełniają szereg warunków systemów

czasu rzeczywistego. Do najpopularniejszych magistral przemysłowych możemy zaliczyć: PROFIBUS, CAN, InterBus, Modbus.

Postęp technologiczny sprawił, że rozproszona technika mikroprocesorowa znalazła się praktycznie wszędzie. Za pomocą sieciowych interfejsów mikrokontrolerów łączone są ze sobą komponenty znajdujące się w obrębie pojedynczego urządzenia elektronicznego (rys. 1).



Rys. 1. Rozproszone urządzenie elektroniczne czasu rzeczywistego

Fig. 1. Distributed real time electronic device

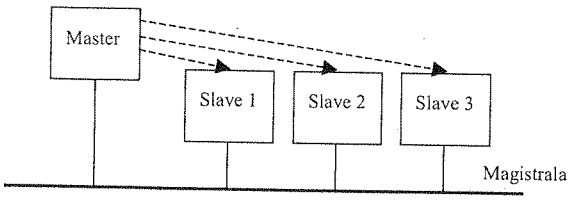
Praca w sieci umożliwia zastosowanie wielu tańszych i prostszych kontrolerów (niż jednego lub kilku bardziej kosztownych) oraz umieszczenie ich w miejscach najbardziej odpowiednich. Najchętniej wykorzystywane są interfejsy szeregowe, które pozwalają na stosowanie tanich dwóch, trzech kabli do transmisji danych. Spośród wielu cyfrowych interfejsów szeregowych, do najczęściej wykorzystywanych możemy zaliczyć: I^2CBus, SMBus, SPI, Microwire i 1-Wire. W dalszej części omówione zostaną interfejsy I^2C, SPI i 1-Wire, ponieważ SMBus jest praktycznie kompatybilne z I^2CBus (do częstotliwości 100 kHz), a Microwire jest kompatybilne z SPI.

W artykule nie są prezentowane szczegółowe (sprzętowe) mechanizmy funkcjonowania ww. interfejsów, ale skupiono się na protokołach wymiany danych w celu przedstawienia modelu analizy czasowej i modelu odpytywania wielocyklowego.

2. ARCHITEKTURA *MASTER-SLAVE*

Interfejsy szeregowe I^2Cbus, SPI i 1-Wire działają w oparciu o architekturę master-slave, stosując tzw. metodę odpytań (*polling*) [5]. Jest to bardzo popularna metoda, zarówno w systemach z magistralami miejscowymi, jak i cyfrowymi interfejsami szeregowymi, ze względu na determinizm i prostotę. W tej metodzie wydzielona jest

stacja master odpytująca stacje slave poprzez wysyłanie odpowiednich wiadomości, przekazując im w ten sposób zgodę na transmisję w sieci (rys. 2).



Rys. 2. Protokół odpytań

Fig. 2. Poling method

Największą zaletą tej metody jest łatwy sposób implementacji oraz fakt, że przy prawidłowym działaniu systemu nie ma możliwości wystąpienia kolizji wiadomości – wiadomości wysyłane są albo przez stację master, albo przez jedną stację slave, tą, która uzyskała zgodę na transmisję. Protokół ten jest idealny w przypadku centralnej akwizycji danych, dla których nie jest wymagana komunikacja typu peer-to-peer (każdy-z-każdym) i nie są stosowane globalne priorytety. Wadą tego rozwiązania jest to, że wystąpienie błędu w węźle master powoduje przerwanie komunikacji. Metoda odpytań zajmuje sporą część pasma transmisyjnego i wykazuje się słabą efektywnością. Niektóre warianty rozwiązań tego protokołu umożliwiają transmisję danych pomiędzy stacjami slave poprzez stację master. Niezawodność systemu zwiększa się poprzez użycie większej liczby węzłów master.

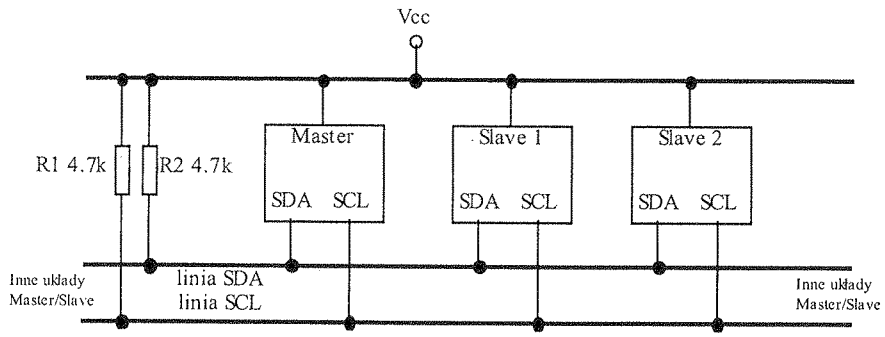
3. LOKALNE INTERFEJSY SZEREGOWE URZĄDZEŃ CYFROWYCH

3.1. INTERFEJS I^2CBUS

Interfejs Inter-Integrated Circuit Bus (I^2CBus) został opracowany w firmie Philips w celu synchronicznej komunikacji szeregowej pomiędzy urządzeniami, modułami w ramach urządzenia, jak również układami scalonymi na płytce drukowanej. Pozwala on na komunikację pomiędzy różnymi układami za pomocą dwuprzewodowej magistrali – linia SDA to linia danych, linia SCL to linia taktująca (rys. 3).
Układy dołączane do magistrali mogą pracować w dwóch trybach:

- ***Master*** – pełni funkcję nadrzędną poprzez inicjowanie transmisji i generowanie sygnału taktującego,
- ***Slave*** – wysyła lub odbiera dane po zaadresowaniu przez układ *master*.

Każde urządzenie magistrali ma swój unikatowy adres. Dane odbiera tylko ten układ, do którego są przeznaczone, a wysyła dane to urządzenie, którego adres został wysłany na magistralę przed przejściem układu *master* w tryb odbioru.



Rys. 3. Podstawowa struktura magistrali I²CBus

Fig. 3. I²CBus interface structure

Każda transmisja musi zaczynać się i kończyć charakterystyczną sekwencją stanów *Start* i *Stop* linii SCL i SDA. Po wysłaniu sygnału *Start* przez dany układ *Master*, a przed nadaniem sygnału *Stop*, żaden inny układ nadrzędny nie może przejąć kontroli nad magistralą. Transmisja danych odbywa się zawsze w formacie 8-bitowym z dodatkowym bitem potwierdzenia ACK (rys. 4).

Start	Adres slave	ACK	Dane (1 bajt)	ACK	Dane (1 bajt)	ACK	Stop	--------	Start

Rys. 4. Przykładowy format przesyłu danych I²CBus

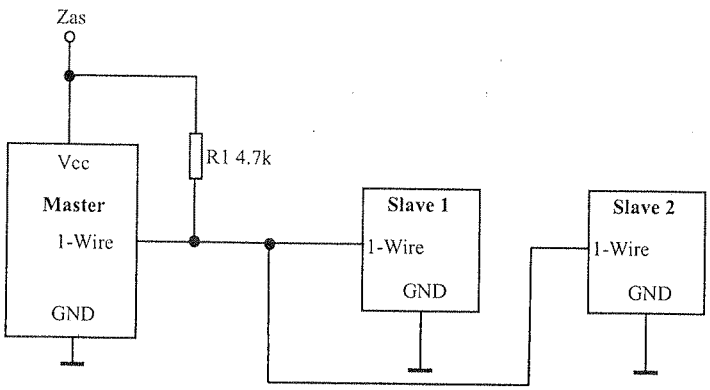
Fig. 4. I²CBus data structure

Bit potwierdzenia wysyłany jest przez układ, do którego przeznaczony był bajt danych.

3.2. INTERFEJS 1-WIRE

Interfejs 1-Wire opracowany przez firmę Dallas (obecnie Maxim) jest jednym z najnowszych cyfrowych interfejsów szeregowych. Jego specyficzną właściwością jest możliwość transmisji danych w dwóch kierunkach z wykorzystaniem jednego wyprowadzenia, którym prowadzone jest też zasilanie (oczywiście do prawidłowego funkcjonowania systemu potrzebna jest *masa*) – rys. 5.

Standard przewiduje pracę w systemie jednego układu *master* i dowolnej liczby układów *slave*. Układy *slave* ze względu na prostotę i brak dodatkowych sygnałów sterujących wyposażone są w interpretatory poleceń. Przesłanie określonej komendy (kodu) do układu slave powoduje wykonanie przez jego wewnętrzny automat sterujący sekwencji czynności odpowiadających poleceniu.



Rys. 5. Typowa struktura 1-Wire

Fig. 5. 1-Wire interface structure

Każda transakcja wymiany danych powinna obejmować trzy etapy:

- *Inicjalizacja.* Zerowanie układu *slave* oraz potwierdzenia przez *slave'a* aktywności w systemie.
- *Przesłanie rozkazu typu ROM.* Każdy układ 1-Wire ma niepowtarzalny, 8-bajtowy kod zapisany w wewnętrznej pamięci ROM. Nosi on nazwę ROM *Code* i może być utożsamiany z adresem układu. Komendy typu ROM umożliwiają zaadresowanie konkretnego układu, identyfikację układu lub pominięcie sprawdzania 64-bitowego kodu (używana jest przy pojedynczym układzie *slave* na magistrali).
- *Przesłanie komendy sterującej.* Zależy od typu układu i dostępnych funkcji układu (dane katalogowe).

Przykładowy format transmisji przedstawiono na rys. 6.

Inicjalizacja	Komenda ROM (1bajt)	*ROM Code* (8 bajtów)	Funkcja (1 bajt)	Dane (1 bajt)

Rys. 6. Przykładowy format transmisji 1-Wire

Fig. 6. 1-Wire data transmission

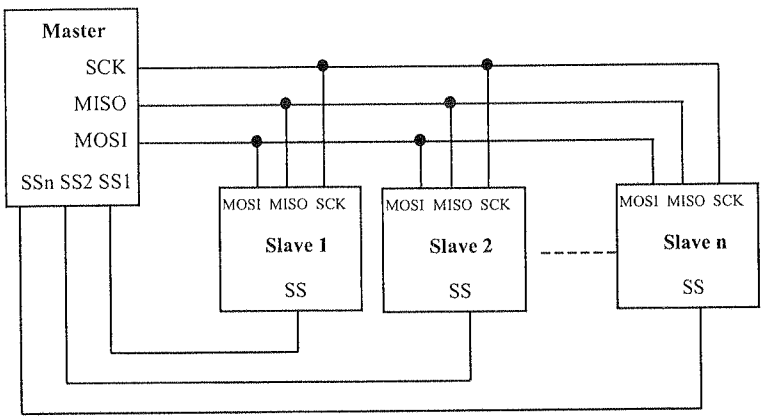
Założono, że komenda ROM to funkcja uaktywnienia układu, którego *ROM Code* jest dokładnie taki, jak sekwencja wysłana przez urządzenie Master. Natomiast funkcja sterująca to np. odczyt danych.

3.3. INTERFEJS SPI

Interfejs *Serial Peripherial Interface* (SPI) został opracowany przez firmę Motorola do synchronicznej, dwukierunkowej komunikacji pomiędzy układami scalonymi. Podobnie jak w innych interfejsach, zastosowano architekturę *master-slave*. Transmisja opiera się o czteroprzewodową linię z możliwością pełnego dupleksu:

- wybór układu slave (*Slave Select*) SS,
- linia sygnału zegarowego (*Serial Clock*) SCK,
- wyjście danych układu master (*Master Out-Slave In*) MOSI,
- wejście danych układu master (*Master In-Slave Out*) MISO.

Przykładową konfigurację magistrali SPI przedstawiono na rys. 7. Jeżeli układ master nie dysponuje wystarczającą liczbą wyjść SS, należy użyć dekodera.



Rys. 7. Struktura mgistrali SPI

Fig. 7. SPI interface structure

Ponieważ wybór układu *slave* następuje sprzętowo, to transmisja danych nie jest obarczona narzutem protokołu ani danymi nadmiarowymi – należy jednak pamiętać, że okupione to jest większą liczbą linii sygnałowych w stosunku do interfejsów I^2C i 1-Wire.

4. MODEL ANALIZY CZASOWEJ DLA SPEŁNIENIA OGRANICZEŃ CZASU RZECZYWISTEGO DLA CYFROWYCH INTERFEJSÓW SZEREGOWYCH

W dowolnym systemie sieciowym, każda wiadomość może należeć do jednej z trzech następujących grup:

- wiadomości okresowe (*periodic*) – dane do przesłania aktywowane są regularnie co odstęp czasu Δt – tzw. zmienne cykliczne,
- wiadomości nieokresowe (*aperiodic*) – dane aktywowane są nieregularnie, jednakże można określić minimalny czas pomiędzy kolejnymi aktywacjami,
- sporadyczne (*sporadic*) – dane aktywowane są nieoczekiwanie, bez określonej reguły.

W dalszej części rozważane będą wiadomości okresowe (najczęściej występujące w systemach sterowania). Dla wiadomości okresowej zdefiniowane zostaną następujące parametry:

- Czas przesłania c_i – czas trwania przesłania całej ramki danych (dane użytkowe plus dane nadmiarowe wprowadzane przez protokół sieciowy, np. bit potwierdzenia),
- Okres występowania t_i – przedział czasu, co który zostają aktywowane dane do wysłania,
- Ograniczenie czasowe (*deadline*) d_i – graniczny przedział czasu dla zrealizowania przesłania wiadomości; jeżeli dane do wysłania zostaną aktywowane w chwili czasu t_o, to ograniczenie czasowe wynosi $t_o + d_i$.

Wymaganie terminowego reagowania systemu (dotrzymania ograniczenia czasowego) jest jednym z ważniejszych wymagań stawianych systemom czasu rzeczywistego. Oznaczmy czas cyklu odpytania węzłów slave przez układ master jako docelowy czas odpytania TPT (*target polling time*). Aby w systemie dotrzymane zostały ograniczenia czasowe wartość TPT musi być dobrana w ten sposób, że:

$$TPT \leq d_{min} \tag{1}$$

gdzie: d_{min} jest najkrótszym ograniczeniem czasowym wiadomości w systemie.
Dla $d_i = t_i$ zachodzi $TPT \leq t_{min}$, gdzie t_{min} jest najkrótszym okresem występowania wiadomości w systemie.
Jeżeli warunek (1) nie będzie dotrzymany, to nie będą spełnione ograniczenia dla danych z okresem występowania t_{min}, a tym samym ograniczeniem d_{min}.

Oznaczamy przez z_M sumaryczny czas przesłania zapytań przez stację master. Czas z_M jest czasem potrzebnym do organizacji transmisji (np. dla magistrali I^2CBUS są to sygnały *Start*, *Stop* i adres układu *slave*). Pozostały czas, tzn. TPT- z_M może być efektywnie wykorzystany do transmisji danych przez stacje (węzły) slave. Czas, w którym węzeł i wyłącznie dysponuje magistralą oznaczony jest przez h_i i obliczany jest wg. wzoru:

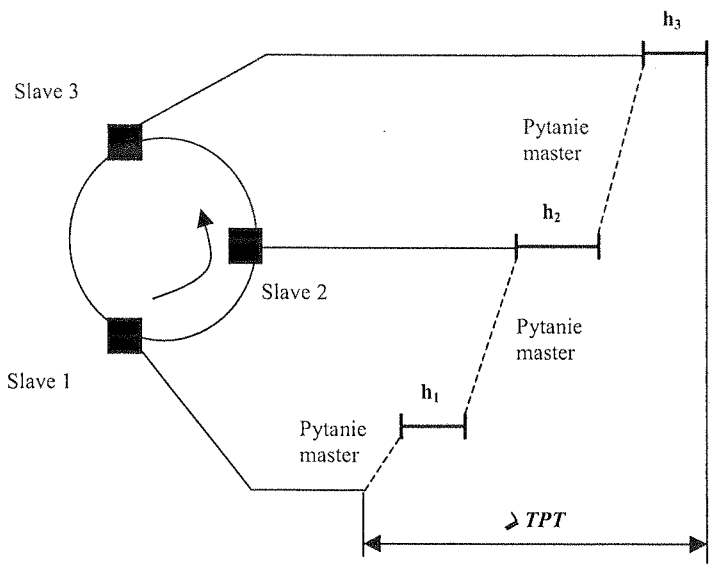
$$h_i = \frac{u_i}{u}(TPT - z_M) \tag{2}$$

gdzie: $u = u_1 + ... + u_n$, a u_i jest stopniem wykorzystania sieci przez stację i

$$u_i = \sum_{j \in P_j} \frac{c_j}{t_j} \tag{3}$$

gdzie: zbiór $P_i = \{P_{i1}, P_{i2}, \ldots, P_{im}\}$ oznacza pakiety danych (wiadomości) należące do węzła i,
Na rys. 8. przedstawiono omawianą sytuację.

Każdy węzeł *slave* może wysyłać dane przez czas nie większy od wartości h_i. W ten sposób nie zostanie przekroczona wartość TPT. Oczywiście powyższy postulat nie gwarantuje jeszcze, że w systemie zostaną dotrzymane ograniczenia czasowe, ale niespełnienie tego warunku powodowałoby, że w systemie na pewno nie byłyby dotrzymane ograniczenia dla wiadomości z ograniczeniem d_{min}, a tym samym cały system nie spełniał by ograniczeń RT (*real time*).



Rys. 8. Ilustracja docelowego czasu odpytań *TPT*

Fig. 8. Target polling time *TPT* ilustration

Aby w systemie były dotrzymane ograniczenia czasu rzeczywistego należy z reguły zastosować szeregowanie wiadomości w węzłach systemu sterującego. Znanych jest wiele metod, jak np.: GRMS, EDF, MLF, MUF [8, 9, 10]. Jednakże urządzenia slave dołączane do cyfrowych interfejsów szeregowych, to z reguły proste układy elektroniczne i nie jest możliwe zastosowanie szeregowania wiadomości – uruchomienie programu szeregującego wymaga przynajmniej obecności procesora z pamięcią.

Jednym ze sposobów polepszenia pracy systemu, bez konieczności stosowania metod szeregowania wiadomości w poszczególnych węzłach systemu rozproszonego jest wprowadzenie modelu odpytywania wielocyklowego.

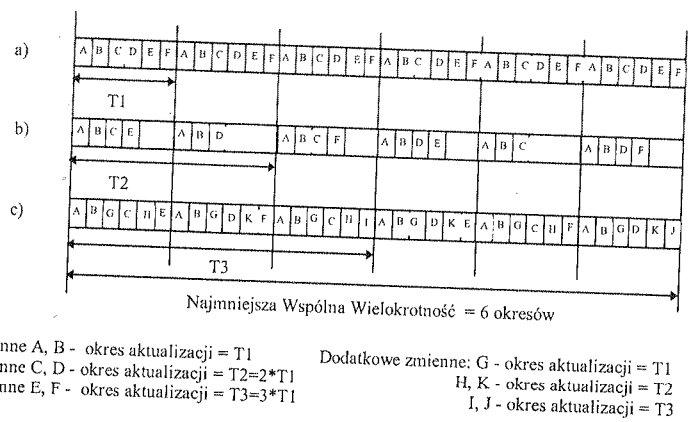
5. MODEL ODPYTYWANIA WIELOCYKLOWEGO

Jak wspomniano wcześniej, przyjęcie odpowiedniej wartości czasu TPT nie gwarantuje, że w systemie zostaną dotrzymane ograniczenia czasowe. Jednym ze sposobów polepszenia pracy systemu jest wprowadzenie modelu odpytywania wielocyklowego. W poprzednim punkcie zdefiniowano pojęcie *zmiennej cyklicznej* jako określonych danych z ustalonym okresem aktualizacji wartości. Klasyczne podejście do problemu odczytania i przesłania wartości zmiennych cyklicznych w rozproszonym systemie akwizycji danych, zakłada użycie jednego cyklu do odpytania wszystkich zmiennych (ponieważ omawiane zagadnienie dotyczy systemów o architekturze master-slave, zamiast sformułowania *'odczytanie i przesłanie wartości zmiennej cyklicznej'* będzie używane

sformułowanie *'odpytanie zmiennej cyklicznej'*), tzn. każda zmienna jest odpytywana jeden raz w każdym cyklu (odpytywanie jednocyklowe). Taki schemat postępowania jest efektywny, jeżeli wszystkie zmienne posiadają ten sam okres aktualizacji. W miarę, jak okresy aktualizacji odpytywanych zmiennych będą się coraz bardziej różnić od siebie, taki schemat działania staje się coraz bardziej nieefektywny, ponieważ okres cyklu odpytywania musi być mniejszy od najkrótszego okresu aktualizacji zmiennych w systemie. Oznacza to, że zmienne o dłuższych okresach są odpytywane częściej niż to konieczne, a w związku z tym wykorzystanie magistrali jest bardzo nieefektywne. Poza tym, taki schemat odpytywania jednocyklowego może nie zapewniać spełnienia ograniczeń czasu rzeczywistego w systemie.

Alternatywnym rozwiązaniem może być odpytywanie wielocyklowe, w którym to rozwiązaniu każda zmienna jest odpytywana tylko jeden raz w ciągu swojego okresu aktualizacji. Poniżej przedstawiony zostanie przykład obrazujący różnice pomiędzy odpytywaniem jednocyklowym, a odpytywaniem wielocyklowym [6].

Załóżmy, że w naszym systemie występuje 6 zmiennych A, B, C, D, E, F o okresach aktualizacji: T1 dla zmiennych A, B, T2 = 2*T1 dla zmiennych C, D oraz T3 = 3*T1 dla zmiennych E, F. Jeżeli zrobimy założenie, że wszystkie zmienne posiadają taką samą długość, to każda zmienna zajmie jedną, taką samą szczelinę czasową podczas każdego pojedynczego cyklu odpytywania (długość szczeliny jest więc równa czasowi zapytania stacji master i czasowi przesłania wiadomości jednej zmiennej).



Rys. 9. Mechanizm odpytywania jednocyklowego i wielocyklowego

Fig. 9. The monocycle and multicycle polling

Na rys. 9 w części a) przedstawiono mechanizm odpytywania jednocyklowego, podczas którego, w każdym cyklu (o okresie T1) odpytywane są wszystkie zmienne. W części b) pokazano organizację odpytywania wielocyklowego, w którym każda zmienna jest odpytywana tylko jeden raz podczas swojego okresu aktualizacji. Dodatkowo, w każdym cyklu podstawowym, o okresie T1 zostaje wolne pasmo, o które można skrócić czas odpytywania, jeżeli w systemie nie były dotrzymane ograniczenia

czasowe dla istniejących zmiennych lub które można, jeżeli istnieje taka potrzeba, zapełnić innymi zmiennymi (rys. 9. c).

Oczywistą rzeczą jest, że jeżeli schemat postępowania zwany odpytywaniem jednocyklowym zapewni nam spełnienie warunków RT w systemie, to należy go zastosować – jest on bowiem prostszy i nie wymaga dodatkowych narzutów czasowych, które są niezbędne przy organizacji odpytywania wielocyklowego.

Natomiast w sytuacji, kiedy stosując odpytywanie jednocyklowe, w systemie nie zostaną dotrzymane ograniczenia RT należy zastosować odpytywanie wielocyklowe, które *może* przynieść zadawalające rezultaty w sensie spełnienia warunków RT (*może*, ponieważ możemy mieć do czynienia z systemem, w którym nie da się dotrzymać ograniczeń RT żadną z omawianych metod – system jest po prostu nieszeregowalny w sensie spełnienia warunków czasu rzeczywistego).

Odnosząc powyższe rozważania do przykładu z rys. 9, jeżeli przyjmiemy czas trwania szczeliny czasowej jako S = (1/6)*T1, to definiując wykorzystanie systemu jako

$$\frac{\sum_i S_i}{T1} * 100\% \tag{4}$$

gdzie: $\sum_i S_i$ oznacza sumaryczny czas szczelin czasowych odpytywanych w pojedynczym cyklu

dla odpytywania jednocyklowego (rys. 9a) otrzymamy wartość równą 100%, natomiast dla odpytywania wielocyklowego otrzymamy 66,6%.

Próba odpytania jakiejkolwiek, dodatkowej zmiennej w schemacie jednocyklowym spowoduje niedotrzymanie ograniczeń czasowych w systemie. Natomiast w przypadku odpytywania wielocyklowego, możemy odpytać kilka dodatkowych zmiennych (o okresach aktualizacji ≥ T1) – rys. 9c.

6. ORGANIZACJA ODPYTYWANIA WIELOCYKLOWEGO

6.1. TRANSFORMACJA OKRESÓW AKTUALIZACJI ZMIENNYCH

Jak wspomniano, cykl podstawowy odpytywania wielocyklowego jest równy najkrótszemu okresowi aktualizacji zmiennych w systemie. Pozostałe okresy aktualizacji zmiennych definiujemy jako wielokrotności okresu podstawowego tzn., jeżeli cykl podstawowy wynosi 15 ms, a dla pewnej zmiennej jej okres aktualizacji wynosi 160 ms, to dokonujemy zaokrąglenia w dół do największej wielokrotności cyklu podstawowego, ale mniejszej od okresu aktualizacji rozpatrywanej zmiennej - okres po transformacji będzie więc wynosił 150 ms = 10 * T1 (11*T1 = 165 ms > 160 ms).

Odpowiedni wzór do transformacji okresów może wyglądać następująco:

$$ki = \left\lfloor \frac{Ti}{T1} \right\rfloor \tag{5}$$

gdzie: $T1$ – najkrótszy okres aktualizacji zmiennej (długość cyklu podstawowego)
$\lfloor x \rfloor$ oznacza największą liczbę całkowitą mniejszą od x
Postępując w ten sposób dokonujemy transformacji wszystkich okresów aktualizacji zmiennych w systemie, a zmienne o tym samym okresie aktualizacji Ti będą tworzyć tzw. grupy Gri. Przyjmijmy następującą notację:
$T1$ – najkrótszy okres aktualizacji zmiennej w systemie (długość cyklu podstawowego), ni – liczba zmiennych w grupie Gri, Ti – okres aktualizacji grupy Gri (równy ograniczeniu czasowemu), $ki = \left\lfloor \frac{Ti}{T1} \right\rfloor$, N – maksymalna liczba grup w systemie.

6.2. WYZNACZENIE DŁUGOŚCI SZCZELINY CZASOWEJ

Załóżmy dla naszych dalszych rozważań, że czas odpytania każdej zmiennej, zwany szczeliną czasową, jest stały. To założenie jest do przyjęcia, jeżeli mamy do czynienia z systemami czasu rzeczywistego opartymi o cyfrowe interfejsy szeregowe, ponieważ transmitowane dane charakteryzują się generalnie niewielką długością.
Czas trwania szczeliny czasowej wyznaczamy ze wzoru:

$$t_{slot} = \frac{l_{Mast} + l_{zm}}{v} \tag{6}$$

gdzie: l_{Mast} – liczba bitów 'pytania' stacji master, l_{zm} – liczba bitów zmiennej o największej długości, v – szybkość transmisji [bit/s].
Tak więc maksymalna liczba szczelin czasowych L_{max}, wyznaczająca maksymalną długość cyklu podstawowego w schemacie odpytywania wielocyklowego, musi spełniać warunek:

$$L_{\max} \leqslant \left\lfloor \frac{T1}{t_{slot}} \right\rfloor \tag{7}$$

7. ALGORYTMY DOBORU DŁUGOŚCI CYKLU PODSTAWOWEGO

7.1. ALGORYTM JEDNORODNY

Na bazie tego algorytmu liczba L_{jedn}, określająca długość cyklu podstawowego, wyznaczana jest z zależności:

$$L_{jedn} = n1 + \left\lceil \sum_{i=2}^{N} \frac{ni}{ki} \right\rceil \tag{8}$$

gdzie: $\lceil x \rceil$ oznacza największą liczbę całkowitą większą od x.

Jak wynika ze wzoru (8) liczba zmiennych z grupy *Gri* odpytywana w cyklu podstawowym jest określona poprzez średnią wartość $\frac{ni}{Ti}$ zmiennych (Ti = ki*T1, T1 = const), dlatego też po okresie Ti zostanie odpytanych $\frac{ni}{Ti} * Ti = ni$ zmiennych z grupy *Gri*. Aby w systemie zostały spełnione ograniczenia czasu rzeczywistego, wystarczy spełnienie warunku:

$$L_{jedn} \leqslant L_{\max} \tag{9}$$

7.2. ALGORYTMY PRIORYTETOWE

Algorytm jednorodny traktuje wszystkie grupy w ten sam sposób. Jeżeli chcemy przydzielić priorytety do odpytywanych zmiennych, musimy zastosować odpowiednie, priorytetowe, algorytmy planowania i szeregowania zadań. Zaprezentowane zostaną dwa takie algorytmy: GRMS i EDF. Zastosowanie tych algorytmów w metodzie wielocyklowego odpytywania przedstawione zostanie w trzech krokach:

- wyznaczanie liczb L_{GRMS} i L_{EDF}, czyli liczby szczelin czasowych cyklu podstawowego,
- wyznaczenie priorytetów zmiennych,
- opis algorytmu odpytywania.

7.2.1. Wyznaczenie liczby szczelin czasowych cyklu podstawowego

Wyznaczanie liczb L_{GRMS} i L_{EDF} dokonywane jest w czasie inicjalizacji cyklu odpytywania.

METODA GRMS

Odnosząc założenia i twierdzenia metody GRMS do naszego systemu, zadanie *i* będzie odpowiadać grupie zmiennych *Gri*, dlatego też przyjmując $C_i^* = ni^*t_{slot}$, oraz zakładając maksymalne obciążenie systemu, możemy zapisać na podstawie zależności:

$$\sum_{i=1}^{N} \frac{C_i^*}{T_i} = 1 \tag{10}$$

W celu wyznaczenia L_{GRMS} zależność (10) zapisujemy w postaci:

$$\sum_{i=1}^{N}\left\lceil\frac{n_i * t_{slot}}{T_i}\right\rceil = \sum_{i=1}^{N}\left\lceil\frac{n_i * t_{slot}}{k_i * T_1}\right\rceil = \frac{L_{GRMS} * t_{slot}}{T_1} \tag{11}$$

Wzór (11) jest zapisem dyskretnym metody GRMS. Przekształcając go dostajemy:

$$L_{GRMS} = \sum_{i=1}^{N}\left\lceil\frac{n_i}{k_i}\right\rceil \tag{12}$$

Aby w systemie zostały dotrzymane ograniczenia czasu rzeczywistego, musi być spełniony warunek:

$$L_{GRMS} \leqslant L_{\max} \tag{13}$$

METODA EDF

Dla algorytmu EDF wielkość L_{EDF} jest wyznaczane poprzez użycie algorytmu jednorodnego. Można udowodnić, że zmienne odpytywane wg. algorytmu EDF w cyklu podstawowym o długości wyznaczonej poprzez algorytm jednorodny spełnią swoje ograniczenia czasowe.

Dowód:

Algorytm EDF działa wg. następującego sposobu (w przypadku metody odpytywania):

- zakładając, że Di = Ti < Dj = Tj dostajemy, że grupa *Gri* ma większy priorytet niż grupa *GRj*,
- wybierana jest grupa o największym priorytecie,
- EDF próbuje odpytać w każdym cyklu L_{EDF} zmiennych,
- zmienne z grupy *Gri* są odpytywane tylko wtedy, jeżeli aktualnie z tej grupy pozostały jakieś do odpytania.

Sekwencja odpytywania zmiennych wg. algorytmu EDF jest permutacją sekwencji algorytmu jednorodnego. Jeżeli zmienne mają przydzielone priorytety, to zmienna *zi* z grupy *Gri* zajmie szczelinę czasową zmiennej *zj* z grupy *GRj* i vice versa. Oznacza to, że mimo takiej zamiany, zmienna *zj* będzie odpytana przed upływem okresu Ti, a ponieważ Ti < Tj to zmienna *zj* dotrzyma swojego ograniczenia czasowego.

Możemy mieć też do czynienia z sytuacją, kiedy to zmienna z grupy *GRj* zajmie miejsce zmiennej z grupy *Gri*. Systuacja taka ma miejsce jedynie w cyklu, który jest ostatecznym cyklem dla zmiennych z grupy *GRj*, w którym mogą być one nadane bez przekroczenia swoich ograniczeń czasowych (tylko wtedy priorytet grupy *GRj* jest większy niż grupy *Gri*). Ponieważ zmiennych z grupy *GRj* jest tylko taka ilość, jaka została wygenerowana przez algorytm jednorodny w rozważanym cyklu, to zajmując miejsce zmiennych z grupy *Gri*, zwalniają jednocześnie miejsca zajmowane przez siebie, które zajmą zmienne z grupy *Gri*. W rozważanej sytuacji dochodzi więc tylko

do zamiany miejsc w obrębie jednego cyklu, nie powoduje to zatem żadnych dodatkowych komplikacji, mimo iż zmienne o dłuższym ograniczeniu czasowym zajęły miejsce zmiennych o krótszym ograniczeniu czasowym.

7.2.2. Przydzielanie priorytetów

METODA GRMS

W metodzie GRMS obowiązuje statyczny przydział priorytetów. Grupa *Gri* uzyskuje wyższy priorytet od grupy *GRj* jeżeli Ti < Tj.

METODA EDF

Tutaj przydział priorytetów odbywa się dynamicznie. Algorytm przyporządkowuje wyższy priorytet grupie z wcześniejszym ograniczeniem czasowym, czyli w naszym przypadku grupie, której do końca ograniczenia czasowego pozostaje najmniejsza liczba cykli. Obliczanie priorytetów może odbywać się z zależności:

$$priorytet(i) = k_i - Modulo\left(\frac{Numer - 1}{k_i}\right) \tag{14}$$

gdzie: $1 \leq i \leq N$
Grupie *Gri* dla której zmienna *priorytet(i)* osiąga najmniejszą wartość, przypisuje się najwyższy priorytet. Jeżeli dwie (lub więcej) grypy posiadają ten sam priorytet, grupa z mniejszą wartością indeksu odpytywana jest wcześniej.

8. PRZYKŁAD OBLICZENIOWY

Parametry przykładowego systemu, po dokonaniu transformacji okresów, przedstawiono w tabeli 1.

Tabela 1

Parametry przykładowego systemu

Parameters of an example system

Numer grupy *Gri*	Nazwa zmiennej	Liczba *ni*	Długość zmiennych w bajtach	Okres uaktualniania [ms]	Liczba *ki*
*GR*1	z11, z12, z13	3	4	T1 = 1,6	1
*GR*2	z21, z22, z23, z24	4	4	T2 = 3,2	2
*GR*3	z31, z32, z33	3	3	T3 = 4,8	3

Zakładamy, że czas trwania szczeliny czasowej (dla celów przykładu obliczeniowego) będzie wynosił:

$$t_{slot} = 0,264[ms]$$

Oznacza to, że dla odpytywania jednocyklowego w systemie nie zostaną dotrzymane ograniczenia czasowe (zapotrzebowanie czasowe dla tego typu odpytywania wynosi $(3 + 4 + 3)*t_{slot} = 2,64$ [ms], a okres T1 = 1,6 [ms]).
Zastosujmy więc schemat postępowania wielocyklowego przy użyciu algorytmu jednorodnego.
Na podstawie (7) wyznaczamy L_{max}:

$$L_{\max} = \left\lfloor \frac{1,6}{0,264} \right\rfloor = \lfloor 6,06 \rfloor = 6$$

Z zależności (8) wyznaczamy L_{jedn}:

$$L_{jedn} = 3 + \left\lceil \frac{4}{2} + \frac{3}{3} \right\rceil = 6$$

Ponieważ $L_{jedn} = L_{max}$, to w systemie zostaną dotrzymane ograniczenia RT. Odpowiednia sekwencja odpytywania podana jest na rys. 10.

z31	z32	z33	z31	z32	z33
z22	z24	z22	z24	z22	z24
z21	z23	z21	z23	z21	z23
z13	z13	z13	z13	z13	z13
z12	z12	z12	z12	z12	z12
z11	z11	z11	z11	z11	z11
T1	2*T1	3*T1	4*T1	5*T1	6*T1

Rys. 10. Sekwencja odpytywania wielocyklowego wygenerowana przez algorytm jednorodny

Fig. 10. The multicycle polling sequence generated by uniform algotithm

Załóżmy teraz, że rozpatrywany system został rozbudowany o kolejne zmienne (tab. 2). Dla metody odpytywania jednocyklowego zapotrzebowanie czasowe wynosi 3,7 [ms], a okres T1 nadal jest równy 1,6 ms.

Tabela 2

Parametry systemu po rozbudowie

System parameters after development

Numer grupy *Gri*	Nazwa zmiennej	Liczba *ni*	Długość zmiennych w bajtach	Okres uaktualniania [ms]	Liczba *ki*
*GR*1	z11, z12, z13, z14	4	4	T1 = 1,6	1
*GR*2	z21, z22, z23, z24	4	4	T2 = 3,2	2
*GR*3	z31, z32, z33, z34 z35, z36	6	3	T3 = 4,8	3

Z zależności (8) wyznaczamy L_{jedn}:

$$L_{jedn} = 4 + \left\lceil \frac{4}{2} + \frac{6}{3} \right\rceil = 8$$

Ponieważ $L_{jedn} > L_{max}$ warunki czasu rzeczywistego w systemie przedstawionym w tab. 2 nie zostaną dotrzymane.
Zastosujmy więc algorytmy priorytetowe, aby zagwarantować spełnienie warunków RT dla grup wysokopriorytetowych.

METODA GRMS

Z zależności (12) wyznaczamy L_{GRMS}:

$$L_{GRMS} = \lceil 4 \rceil + \left\lceil \frac{4}{2} \right\rceil + \left\lceil \frac{6}{3} \right\rceil = 8$$

Jak można było przewidzieć $L_{GRMS} > L_{max}$.

METODA EDF

Dla algorytmu EDF $L_{EDF} = L_{jedn}$, czyli $L_{EDF} > L_{max}$. Odpowiednie sekwencje odpytywania wielocyklowego, wygenerowane przez poszczególne algorytmy, przedstawiono na rys. 11. Dla algorytmów priorytetowych przyjęto założenie, że jeżeli zmienna przekroczy swoje ograniczenie czasowe, to nie jest odpytywana.

Jak wynika z poniższego rysunku, przy zastosowaniu algorytmu jednorodnego, przekroczenie ograniczeń czasowych dotyczy 50% zadań z *GR*2 i 50% zadań z *GR*3.

W przypadku algorytmu GRMS niespełnienie ograniczeń czasowych dotyczy tylko zadań z *GR*3 (ale 100%). Natomiast w przypadku algorytmu EDF odsetek zadań z niespełnionymi ograniczeniami RT wynosi odpowiednio 16,6% dla zadań z *GR*2 i 83,4% dla zadań z *GR*3.

z31	z32	z33	z31	z32	z33
z21	z22	z21	z22	z21	z22
z14	z14	z14	z14	z14	z14
z13	z13	z13	z13	z13	z13
z12	z12	z12	z12	z12	z12
z11	z11	z11	z11	z11	z11

T1 2*T1 3*T1 4*T1 5*T1 6*T1

Algorytm jednorodny

z22	z24	z22	z24	z22	z24
z21	z23	z21	z23	z21	z23
z14	z14	z14	z14	z14	z14
z13	z13	z13	z13	z13	z13
z12	z12	z12	z12	z12	z12
z11	z11	z11	z11	z11	z11

T1 2*T1 3*T1 4*T1 5*T1 6*T1

Algorytm GRMS

z22	z24	z32	z22	z22	z24
z21	z23	z31	z21	z21	z23
z14	z14	z14	z14	z14	z14
z13	z13	z13	z13	z13	z13
z12	z12	z12	z12	z12	z12
z11	z11	z11	z11	z11	z11

T1 2*T1 3*T1 4*T1 5*T1 6*T1

Algorytm EDF

Rys. 11. Sekwencje odpytywania wielocyklowego wygenerowane przez algorytm jednorodny oraz EDF i GRMS

Fig. 11. The multicycle polling sequence generated by uniform, GRMS and EDF algotithms

9. WNIOSKI

Interfejsy szeregowe I^2Cbus, SPI i 1-Wire, działające w oparciu o architekturę master-slave i stosujące tzw. metodę odpytań znalazły (i ciągle znajdują) szerokie zastosowanie we współczesnych urządzeniach elektronicznych. Dzięki nim w łatwy sposób może być zrealizowana koncepcja rozproszonego urządzenia czasu rzeczywistego. Istotnym zagadnieniem staje się analiza czasowa takiego systemu pod kątem spełnienia ograniczeń czasowych.

Spełnienie warunków przedstawionych w modelu analizy czasowej dla szeregowych interfejsów cyfrowych (pkt. 4) nie gwarantuje jeszcze, że w systemie zostaną dotrzymane ograniczenia czasowe, ale niespełnienie tego warunku powodowałoby, że w systemie na pewno nie byłyby dotrzymane ograniczenia dla wiadomości z ograniczeniem d_{min} – jest to więc warunek konieczny, ale niewystarczający.

Efektywność systemu (w sensie dotrzymania dotrzymane ograniczeń czasu rzeczywistego) można zwiększyć poprzez zastosowanie szeregowania wiadomości w węzłach systemu rozproszonego. W większości przypadków nie jest to możliwe, ponieważ układy slave to z reguły proste układy elektroniczne, a uruchomienie programu szeregującego wymaga przynajmniej obecności procesora z pamięcią.

Jednym ze sposobów polepszenia pracy podsystemu z interfejsem szeregowym opartym o architekturę *master-slave* jest wprowadzenie modelu odpytywania wielocyklowego i wybór jednego z algorytmów – jednorodnego lub priorytetowego.

Najprostszym algorytmem jest algorytm jednorodny. Jeżeli więc zastosujemy ten algorytm, tzn. wygenerujemy liczbę L_{jedn}, która spełni warunek (9), to stosowanie algorytmów priorytetowych GRMS lub EDF jest zbędne. Dodatkowo, liczba L_{jedn} równa jest liczbie L_{EDF}, a jest mniejsza lub równa liczbie L_{GRMS}, dlatego też zastosowanie algorytmu jednorodnego jest rozwiązaniem najbardziej optymalnym.

Sytuacja ulega zmianie, jeżeli $L_{jedn} > L_{max}$, co oznacza, że w systemie nie zostaną dotrzymane warunki RT. Oznacza to także, że warunki te nie zostaną spełnione przy

zastosowaniu algorytmów GRMS i EDF. Jednakże, w przypadku algorytmów priorytetowych możemy określić, dla których zmiennych ograniczenia zostaną dotrzymane (wysoki priorytet), a dla których nie (niskie priorytety) – właśnie w takim przypadku opłacalne jest stosowanie algorytmów priorytetowych.

10. BIBLIOGRAFIA

1. J. Bogusz: *Lokalne interfejsy szeregowe w systemach cyfrowych.* Wydawnictwo *btc*, 2004.
2. P. Hadam: *Projektowanie systemów mikroprocesorowych.* Wydawnictwo *btc*, 2004.
3. W. Mielczarek: *Szeregowe interfejsy cyfrowe.* Wydawnictwo Helion, 1993.
4. L. Sha, R. Rajkamur, S. Sathaye: *Generalised Rate-Monotonic Scheduling Theory: A Framework for Developing Real Time Systems.* Proc. of the IEEE. Vol. 82, No. 1, Jan. 1994, pp. 68-82.
5. J. Werewka, S. Żaba, A. Drwal: *Protokoły dostępu i charakterystyki czasowe magistral miejscowych.* POMIARY AUTOMATYKA ROBOTYKA – Miesięcznik Naukowo-Techniczny. Wrzesień 1997, ss. 4-11.
6. S. Żaba: *Zastosowanie metody odpytywania wielocyklowego w magistralach miejscowych opartych o architekturę master slave.* V Konferencja Systemy Czasu Rzeczywistego '98, 14-17 wrzesień Szklarska Poręba, 1998, ss. 325-336.
7. S. Żaba: *Message scheduling in distributed real-time system based on fieldbus.* 5th International Symposium on Methods and Models in Automation and Robotics MMAR'98, 25-29 August 1998, pp. 565-700.
8. S. Żaba: *Szeregowanie zadań ze statycznym przydziałem priorytetu w rozproszonych systemach czasu rzeczywistego.* XIII Krajowa Konferencja Automatyki, 21-24 wrzesień, Opole 1999, ss. 103-108.
9. S. Żaba: *Szeregowanie wiadomości z dynamicznym przydziałem priorytetu w rozproszonych systemach sterujących.* VI Konferencja Systemy Czasu Rzeczywistego, 27-30 wrzesień, Zakopane 1999, ss. 128-139.
10. S. Żaba, J. Werewka: *Szeregowanie wiadomości w rozproszonych systemach czasu rzeczywistego opartych o magistrale miejscowe.* Kwartalnik Elektroniki i Telekomunikacji, Wydawnictwo Naukowe PWN, 1999, T. 45, z. 1, ss. 25-50.
11. S. Żaba: *Analiza czasowa wybranych magistral miejscowych.* POMIARY AUTOMATYKA ROBOTYKA – Miesięcznik Naukowo-Techniczny, czerwiec 2003, ss. 12-15.
12. S. Żaba: *Badania eksperymentalne magistral miejscowych dla wybranych metod szeregowania wiadomości.* Kwartalnik Elektroniki i Telekomunikacji, Wydawnictwo Naukowe PWN, 2004, T. 50, z.2, ss. 261-286.

S. ŻABA

APPLICATION OF MULTICYCLE MECHANISM IN DIGITAL SERIAL INTERFACE BASED ON MASTER-SLAVE ARCHITECTURE

Summary

Modern electronic devices use digital serial interfaces to conecting microprocessor and other elements. It allows to use cheap and simple devices and to place them in most correct places. Two or three wires interfaces are used generally. The most popular architecture of that systems is master-slave. That architecture is easy to implement and there is lack of message collisions. In this paper three digital serial

interfaces: I^2Cbus, SPI and 1-Wire based on master-slave architecture are presented briefly. There are the most popular among that kind of digital serial interfaces.

Next a model of time analysis in sense of meeting time constrain is shown. Let the time of polling every slaves by a master be called *target polling time* (*TPT*). The necessary condition to meet time constrain in a system is the TPT value must meet condition $TPT \leq d_{min}$, where d_{min} is minimum value of messages deadline in a system. If a system still doesn't meet time constrain a multicycle mechanism as alternative manner to monocycle mechanism should be used in master-slave systems. The monocycle mechanism polls every variable during TPT time. The multicycle method is based on a rule that every variable in a system is polled only once during its period.

Three algorithms are shown which generate the polling sequences for multicycle mechanism: a uniform algorithm and two priority algorithms – based on Generalised Rate Monotonic Scheduling and Earliest Deadline First methods. In order to use the multicycle mechanism is necessary to transform periods of variables as follows: the base cycle of multicycle method equals the minimum period, the rest periods are calculated as integer multiple of the minimum period. Variables with the same period after transformation form groups of variable. Algorithms of multicycle mechanism poll variables from these groups as follows: the uniform algorithm from minimum period to maximum, priority algorithms according to assigned priority to groups.

The uniform algorithm is the simplest and additional time for performing algorithm is the smallest. If a system still doesn't meet time constrain the best way is to use priority algorithm, because it is possible to determine which variables will meet constrain (high priority) and which will not meet (low priority).

Keywords: distributed control system, real time system, digital serial interface, multicycle mechanism, GRMS, EDF methods

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 powinno charakteryzować oryginalne ujęcie zagadnienia, własna klasyfikacja, krytyczna ocena (teorii lub metod), omówienie aktualnego stanu, lub postępu danej gałęzi techniki oraz omówienie perspektyw rozwojowych. Artykuły publikowane w innych czasopismach nie mogą być kierowane do druku w Kwartalniku Elektroniki i Telekomunikacji w drugiej kolejności zgłoszenia.
Objętość artykułu nie powinna przekraczać 30 stron po około 1800 znaków na stronie, w tym rysunki i tabele.

Wymagania podstawowe.
Artykuły należy nadsyłać na wyraźnym, jednostronnym, czarno-białym wydruku komputerowym. Wydruk w formacie A4 powinien mieć znormalizowaną liczbę wierszy i znaków w wierszu (30 wierszy po 60 znaków w wierszu), w dwóch egzemplarzach, w języku polskim lub angielskim wybranym przez autora. Do wydruku powinna być dołączona dyskietka z elektronicznym tekstem artykułu. Preferowane edytory to WORD 6 lub 8.
Układ artykułu (w wersji podstawowej) musi być następujący:
- Tytuł.
- Autor (imię i nazwisko autora/ów).
- Miejsce pracy (nazwa instytucji, miejscowość, adres. + ew. adres elektroniczny (e-mail).
- Zwięzłe streszczenie powinno być w języku takim, w jakim jest pisany artykuł (wraz ze słowami kluczowymi).
- Tekst podstawowy powinien mieć następujący układ:
  1. WPROWADZENIE
  2. np. TEORIA
  3. np. WYNIKI NUMERYCZNE
  3.1. .....
  3.2. .....
  4. .....
  5. .....
  6. PODSUMOWANIE
  7. ew. PODZIĘKOWANIA
  8. BIBLIIOGRAFIA
- Układ streszczenia w dodatkowej wersji językowej powinien być następujący:
  AUTOR (inicjał imienia i nazwisko).
  TYTUŁ (w języku angielskim – o ile artykuł pisany jest w języku polskim i na odrwót).
Obszerne do 3600 znaków streszczenia (wraz z słowami kluczowymi) w języku:
  a. angielskim, gdy artykuł pisany jest w języku polskim.
  b. polskim, gdy artykuł pisany jest w języku angielskim.
Streszczenie to powinno pozwolić czytelnikowi na uzyskanie istotnych informacji zawartych w pracy. Z tego względu w streszczeniu tym mogą być cytowane numery istotnych wzorów, rysunków i tabel zawartych w podstawowej wersji językowej.
- Wszystkie strony muszą mieć numerację ciągłą.

Sposób pisania tekstu.
Tekst powinien być pisany bez używania wyróżnień, a w szczególności nie dopuszcza się spacjowania, 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 adjustacyjnych, np. podkreślenie linią przerywaną oznacza spacjowanie (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. Wielkość czcionki wydruku powinna być zbliżona co najmniej co wielkości czcionki maszyny do pisania (minimum 12 punktów). Przy podziale pracy na rozdziały i podrozdziały cyfrowe ich oznaczenia nie powinny być większe niż II stopnia (np. 4.1.1.).

Sposób pisania tabel.
Tabele powinny być pisane na oddzielnych stronach. Tytuły rubryk pionowych i poziomych powinny być napisane małymi literami z podwójnym odstępem między wierszami. Przypisy (notki) dotyczące tabel należy pisać bezpośrednio pod tabelami. Tabele należy numerować kolejno liczbami arabskimi, u góry każdej tabeli podać tytuł dwujęzyczny. W pierwszej kolejności w podstawowej wersji językowej, a później w dodatkowej wersji językowej. Tabele umieścić na końcu maszynopisu. Przyjmowane są tabele algorytmów i programy na wydrukach komputerowych. W tym przypadku zachowany jest ich oryginalny układ. Tabele powinny być cytowane w tekście.

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 tekście artykułu. Przykłady opisu publikacji:

- periodycznej — 1. F. Valdoni: A new milimetre wave satelite. E.T.T. 1990, vol. 2, no 5, pp. 141–148
- nieperiodyczne — 2. K. Andersen: A resource allocation framework. XVI International Symposium, Stockholm (Sweden), may 1991, paper A 2,4
- książki — 3. Y.P. Tvidis: Operation and modeling of the MOS transistors. New York, McGraw-Hill, 1987, p. 553

Materiały ilustracyjne.
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 (preferowany edytor Corel Draw). Fotografie lub diapozytywy przyjmowane są raczej 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 oraz numer rysunku. Spis podpisów pod rysunki i fotografie powinny być umieszczone na oddzielnej stronie. Podpisy pod rysunkami (fotografiami) powinny być dwujęzyczne: w pierwszej kolejności w podstawowej wersji językowej, a później w dodatkowej wersji językowej. Rysunki powinny być cytowane w tekście.

Uwagi końcowe.
Na odrębnej stronie powinny być podane następujące informacje:
- adres do korespondencji z kodem pocztowym (domowy lub do miejsca pracy),
- telefon domowy i/lub do miejsca pracy,
- adres e-mailowy (jeśli autor posiada).

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

Autora obowiązują korekta autorska, którą powinien wykonać w ciągu 3 dni od daty otrzymania tekstu z Redakcji oraz zwrócić osobiście lub listownie pod adres 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 AUTHORS OF K.E.T.

The editorial stuff will accept for publishing only original monographic and survey papers concerning widely understood electronics. Because of the fact that KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI is a journal of the Committee for Electronics and Telecommunications of Polish Academy of Scince, it presents scientific works concerning theoretical bases and applications from the field of electronics, telecommunications, microelectronics, optoelectronics, radioelectronics and medical electronics.
Articles should be characterised by original depiction of a problem, its own classification, critical opinion (concerning theories or methods), discussion of an actual state or a progress of a given branch of a technique and discussion of development perspectives.
An article published in other magazines can not be submitted for publishing in K.E.T. TThe size of an article can not exceed 30 pages, 1800 character each, including figures and tables.

Basic requirements

The article should be submitted to the editorial stuff as a one side, clear, black and white computer printout in two copies. The article should be prepared in English or Polish. Floppy disc with an electronic version of the article should be enclosed. Preferred wordprocessors: WORD 6 or 8.
Layout of the article.

- Title.
- Author (first name and surname of author/authors).
- Workplace (institution, adress and e-mail).
- Concise summary in a language article is prepared in (with keywords).
- Main text with following layout:
  - Introduction
  - Theory (if applicable)
  - Numerical results (if applicable)
  - Paragraph 1
  - Paragraph 2
  - .....
  - .....
  - Conclusions
  - Acknowledgements (if applicable)
  - References
- Summary in additional language:
  - Author (firs name initials and surname)
  - Title (in Polish, if article was prepared in English and vice versa)
  - Extenstive summary, hawever not exceeching 3600 characters (along with keywords) in Polish, if artide wos prepared in English and vice versa). The summary should be prepared in a way allowing a reader to obtaoin essential information contained in the artide. For that reason in the summary author can place numbers of essential formulas, figures and tables from the article.

Pages should have continues numbering.

Main text

Main text can not contain formatting such as spacing, underlining, words written in capital letters (except words that are commonly written in capital letters). Author can mark suggested formatting with pencil on the margin of the article using commonly accepted adjusting marks.
Text should be written with double line spacing with 35 mro left and right margin. Titles and subtitles should be written with small letters. Titles and subtitles should be numberd using no mor than 3 levels (i.e. 4.1.1.).

Tables

Tables with their titles should be places on separate page at the end of the article. Titles of rows and columns should be written in small letters with double line spacing. Annotations concerning tables

should be placed directly below the table. Tables should be numbered with Arabic numbers on the top of each table. Table can consist algorithm and program listings. In such cases original layout of the table will be preserved. Table should be cited in the text.

Mathematical formulas

Characters, numbers, letters and spacing of the formula should be adequate to layout of main text. Indexes should be properly lowered or raised above the basic line and clearly written. Special characters such as lines, arrows, dots should be placed exactly over symbols which they ar attributed to. Formulas should be numbered with Arabic numbers place in brackets on the right side of the page. Units of measure, letter and graphic symbols should be printed according to requirements of IEE (International Electronical Commission) and ISO (International Organisation of Standardisation).

References

References should be placed at the and of the main text with th subtitle „References". Refrences should be numbered (without brackets) adequately to references placed in the text. Examples of periodical [@], non-periodical [2] and book [3] references:

1. F. Valdoni: A new milimetre wave satelite. E.T.T. 1990, vol. 2, no 5, pp. 141–148
2. K. Anderson: A resource allocation framework. XVI International Symposium (Sweden). May 1991. paper A 2.4
3. Y.P. Tvidis: Operation and modeling of the MOS transistors. Nw York. McGraw-Hill. 1987. p. 553

Figures

Figures should be clearly drawn on plain or milimetre paper in the format not smaller than 9 × 12 cm. Figures can be also printed (preferred editor – CorelDRAW). Photos or diapositives will be accepted in black and whit format not gratd than 10 × 15 cm. On th margin of each drawing and on th back side of each ohoto authorp name and abbreviation of the title of article should be placed. Figure's captures should be given in two longuaqes (first in the language the article is writtes in and then in additional language). Figure's captures should be also listed on separate page. Figures should be cited in the text.

Additional information

On the separate page following information should be placed:

- mailing address (home or office),
- phone (home or/and office),
- e-mail.

Author is entiled to free of charg 20 copies of articl. Additional copies or the whole magazine can be ordered at publisher at th one's expense.

Author is obliges to perform the author's correction, which should be accomplished within 3 days starting from the date of receiving of the text from the ditorial stuff. Corrected text should be return to the editorial stuff personally or by mail. Correction marks should be placed on th margin of copies received from the editorial stuff or if needed on separate pages. In the case when the correction is not returned in said time limit, correction will be performed by technical editorial stuff of the publisher.

In case of changing of workplace or home address Authors are asked to inform the editorial stuff.