Bulletin of Physics and Applied Mathematics

Edited by:

Adam Chudecki

TUL Press, Lodz, Poland

Bulletin of Physics and Applied Mathematics
https://fizyka.p.lodz.pl/badania/biuletyn/

ISSN XXXX-YYYY

WYDAWNICTWO POLITECHNIKI ŁÓDZKIEJ
90-924 Łódź, ul. Wólczańska 223
tel. 42-631-20-87, 42-631-29-52
e-mail: zamowienia@info.p.lodz.pl
www.wydawnictwo.p.lodz.pl

Spis treści:

ADAM CHUDECKI*[1,2], JANUSZ TOMASZEWSKI†[1], PIOTR SŁOMA‡[1]

RÓWNANIA TRZECIEGO STOPNIA W DYDAKTYCE FIZYKI

[1] Centre of Mathematics and Physics, Lodz University of Technology, Al. Politechniki 11, 93-590 Łódź, Poland
[2] Institute of Physics, Lodz University of Technology, ul. Wólczańska 219, 90-924 Łódź, Poland

W artykule dyskutujemy przydatność równań trzeciego stopnia w izyce poziomu szkoły średniej i pierwszego roku studiów wyższych technicznych. Doświadczenia dydaktyczne wskazują, że metody rozwiązywania takich równań nie są szczegółowo przedstawiane uczniom i studentom. Ten stan rzeczy usprawiedliwia się twierdząc, że równania trzeciego stopnia pojawiają się raczej w zaawansowa-nych zagadnieniach. Polemizujemy z tą tezą analizując dwa, dość elementarne przykłady.

Keywords: równania stopnia trzeciego, równania sześcienne, wzory Cardano.

1. WSTĘP

1.1. Równania trzeciego stopnia

Ogólne równanie stopnia trzeciego ma postać,

$$ax^3 + bx^2 + cx + d = 0, \tag{1}$$

gdzie a, b, c i d to współczynniki rzeczywiste bądź zespolone, a x to zmienna (również rzeczywista, bądź zespolona). Zakłada się, że współczynnik $a \neq 0$, w przeciwnym razie równanie (1) staje się równaniem stopnia drugiego, którego sposoby na rozwiązanie znane były od starożytności.

Historia równań trzeciego stopnia sięga zamierzchłej historii starożytnego Babilonu, Egiptu i Grecji [1]. Jednak dopiero w XVI wieku udało się skompletować metodę ich rozwiązywania. Zawdzięczamy to wybitnym matematykom tamtych czasów, wśród których wymienić należy takich uczonych, jak: Scipio del Ferro, Antonio Mario Fiori, Niccolo Tartaglia i Girolamo Cardano. Fascynującej historii odkrycia sposobu na rozwiązywanie równań trzeciego stopnia, matematycznych „pojedynków na zadania", przekazywanych w zaufaniu formuł i - wreszcie - „zdrady", o jaką Cardano został przez Tartaglię oskarżony, nie będziemy tu przytaczać. Zainteresowanych odsyłamy do treściwego bloga P. Gładkiego [2].

Nie będziemy w szczegółach przedstwiać algorytmu, który pozwala na znalezienie wszystkich pierwiastków równania trzeciego stopnia (wiadomo, że każde równanie trzeciego stopnia ma przynajmniej jeden pierwiastek rzeczywisty). Przypomnijmy jedynie, że przez podstawienie:

$$x = y - \frac{b}{3a} \tag{2}$$

*adam.chudecki@p.lodz.pl
†janusz.tomaszewski@p.lodz.pl
‡piotr.sloma@p.lodz.pl

równanie (1) redukuje się do tzw. *postaci kanonicznej Cardano*:

$$y^3 = 3py + 2q, \quad -3p := \frac{c}{a} - \frac{b^2}{3a^2}, \quad -2q := \frac{2b^3}{27a^3} + \frac{d}{a} - \frac{bc}{3a^2}. \tag{3}$$

Równanie (3) może być rozwiązane przy pomocy tzw.: *wzoru Cardano*

$$y = \sqrt[3]{q + \sqrt{q^2 - p^3}} + \sqrt[3]{q - \sqrt{q^2 - p^3}} \tag{4}$$

Szczegółową dyskusję stosowalności wzoru (4) można znaleźć w wielu pozycjach, np. w [3].

Wzór Cardano (4) pozwolił Rafaelowi Bombelliemu odkryć liczby zespolone (*Algebra*, 1572), których obecna rola w matematyce i fizyce jest nie do przecenienia. Otóż Bombelli zauważył, że nawet gdy $q^2 - p^3 < 0$, można ze wzoru Cardano uzyskać rozwiązanie rzeczywiste. Bombelli posłużył się przykładem $x^3 = 15x + 4$. Łatwo odgadnąć, że liczba 4 jest rozwiązaniem tego równania. Jednak wzór Cardano daje

$$x = \sqrt[3]{2 + 11i} + \sqrt[3]{2 - 11i}, \tag{5}$$

gdzie $i^2 = -1$. Bombelli zaproponował algebraiczne reguły operacji na liczbach, które dziś nazywamy zespolonymi. Otrzymał

$$\sqrt[3]{2 + 11i} = 2 + i, \quad \sqrt[3]{2 - 11i} = 2 - i, \tag{6}$$

co w konsekwencji daje $x = 4$. W taki sposób odkryto liczby zespolone.

Zauważmy jeszcze, iż pośród wszystkich równań wielomianowych, ogólne metody ich rozwiązywania możliwe są do sformułowania jedynie dla stopni $\leqslant 4$. Rozwiązań równań stopnia piątego i wyższych nie da się już niestety określić w skończonej liczbie kroków, bazujących na wzorach zbudowanych ze współczynników równania i wykorzystujących cztery podstawowe działania arytmetyczne oraz pierwiastki stopni naturalnych. Dowodu tego niezwykłego faktu dostarcza tzw. *twierdzenie Abela-Ruffiniego*, szeroko omówione np. w [4].

1.2. Równania sześcienne w dydaktyce fizyki i matematyki

Doświadczenia zebrane przez autorów w trakcie pracy nauczyciela akademickiego dowodzą, że zagadnienie równań sześciennych (jak i równań czwartego stopnia), nie jest zagadnieniem szeroko poruszanym na ćwiczeniach z matematyki i fizyki. Równania wielomianowe stopni wyższych niż drugi rozwiązywane są już raczej na ćwiczeniach z metod numerycznych, gdzie odpowiednie pierwiastki znajduje się np. przy pomocy metody Newtona-Raphsona. Podejście takie usprawiedliwia się twierdząc, że równania sześcienne nie pojawiają się w zagadnieniach fizycznych poruszanych na lekcjach w szkole średniej i na wykładach na pierwszych latach studiów technicznych.

Nie jest to prawda. Faktycznie, problemy wymagające umiejętności rozwiązywania równań stopnia trzeciego nie są powszechne, ale z całą pewnością występują i to w zagadnieniach, których fizyczne podstawy nie wykraczają poza zakres szkoły średniej.
Jednym z najprostszych i najbardziej klasycznych problemów, w którym naturalnie pojawia się równanie sześcienne, jest zagadnienie dwóch, zawieszonych naprzeciwko siebie, jednoimiennie naładowanych kul. Odpowiednie zadania można znaleźć np. w kultowych już pozycjach pod redakcją M.S. Cedrika [5] i [6]. Ścisłe rozwiązanie takiego zagadnienia przedstawiamy w paragrafie 2.1. Bardziej wyrafinowanego przykładu dostarcza zagadnienie ruchu ciała w polu grawitacyjnym przy obecności stałych sił oporu powietrza – opisujemy je w paragrafie 2.2.

Wspomnijmy jeszcze, że z równaniem wielomianowym trzeciego stopnia spotkamy się także na gruncie termodynamiki. Omawiane na pierwszym roku studiów technicznych tzw. równanie Van der Waalsa stanu gazu rzeczywistego, zapisywane dla jednego mola tradycyjnie w następujący sposób:

$$\left(p + \frac{a}{V_m^2}\right)(V_m - b) = RT, \tag{7}$$

można przecież przekształcić do postaci:

$$pV_m^3 - (bp + RT)V_m^2 + aV_m - ab = 0. \tag{8}$$

Jest ono zatem równaniem wielomianowym trzeciego stopnia ze względu na objętość molową. Biorąc pod uwagę rząd wielkości empirycznych współczynników a i b [7] oraz realne wartości ciśnienia p i temperatury T, otrzymamy zawsze jedno rozwiązanie rzeczywiste takiego równania (z uwagi na dodatni wyróżnik postaci kanonicznej). Mamy zatem np. możliwość obliczenia analitycznie objętości molowej gazu dla innych niż normalne warunków ciśnienia i temperatury.

2. PRZYKŁADY

2.1. Zadanie 1: dwie naładowane kulki

Klasyczny problem, w którym pojawia się równanie trzeciego stopnia, to problem dwóch naładowanych ciał. Rozważmy dwie małe metalowe kulki o jednakowych masach m powieszone na niciach o długości l, zaczepionych w tym samym punkcie. Kulki te początkowo stykają się ze sobą. Następnie dostarczamy im ładunek elektryczny $2q$, który rozkłada się na nich po równo, a w efekcie odpychanie kulombowskie rozdziela kulki tworząc z ich środków i punktu zawieszenia trójkąt równoramienny o podstawie x (rysunek 1). Zadanie polega na znalezieniu wspomnianej odległości x, na jaką kulki oddalą się od siebie.

W uproszczonym wariancie opisywanym przez Cedrika [5] przyjmuje się, że kąt rozwarcia nitek jest bardzo mały, wskutek czego otrzymuje się trywialne równanie trzeciego stopnia o rozwiązaniu

$$x = \sqrt[3]{\frac{2klq^2}{mg}}, \tag{9}$$

gdzie k to „stała" elektrostatyczna, zaś g to przyspieszenie ziemskie. W wariancie „bezkompromisowym" na podstawie rysunku 1 znajdujemy związki:

$$\sin\theta = \frac{x}{2l}, \quad \operatorname{tg}\theta = \frac{kq^2}{mgx^2}. \tag{10}$$

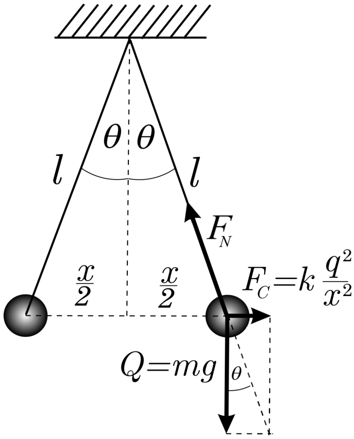
Eliminując z nich następnie kąt θ otrzymujemy równanie:

$$m^2g^2x^6 + k^2q^4(x^2 - 4l^2) = 0, \tag{11}$$

które można przekształcić do postaci:

$$ay^3 + y - 4 = 0, \quad a := \frac{m^2g^2l^4}{k^2q^4}, \quad y := \frac{x^2}{l^2}. \tag{12}$$

Równanie (12) ma jedno rzeczywiste rozwiązanie bez względu na wartość parametru a (jest to fizycznie jasne, jako że przedstawione zagadnienie może mieć tylko jedno rozwiązanie).



Rysunek 1. Dwie odpychające się kulki: Q – ciężar kulek, Fc – siła kulombowskiego odpychania między nimi, FN – siła naciągu nici. W stanie równowagi wypadkowa sił działających na kulkę wynosi zero
Źródło: opracowanie własne.

Używając do jego rozwiązania dostępnych online narzędzi (np. [8]), dostajemy:

$$y = \frac{\sqrt[3]{18a^2 + \sqrt{3}\sqrt{108a^4 + a^3}}}{a\sqrt[3]{9}} - \frac{1}{\sqrt[3]{3}\sqrt[3]{18a^2 + \sqrt{3}\sqrt{108a^4 + a^3}}}. \tag{13}$$

Nie mając dostępu do narzędzi bazujących na pakietach do obliczeń symbolicznych, uzyskanie rozwiązania (13) – bez znajomości wzoru Cardano – jest praktycznie niemożliwe. Jeśli jednak znamy wzór Cardano, porównując (3) z (12) łatwo zauważyć, że $3p = -\frac{1}{a}$ oraz $2q = \frac{4}{a}$. Z (4) dostajemy

$$y = \sqrt[3]{\frac{2}{a} + \sqrt{\frac{4}{a^2} + \frac{1}{27a^3}}} + \sqrt[3]{\frac{2}{a} - \sqrt{\frac{4}{a^2} + \frac{1}{27a^3}}} \tag{14}$$

Oczywiście rozwiązania (13) i (14) są tożsame, co łatwo sprawdzić. Wracając do podstawienia ze wzorów (12) otrzymujemy odległość między kulkami:

$$x = l\sqrt{\frac{\sqrt[3]{18a^2 + \sqrt{3}\sqrt{108a^4 + a^3}}}{a\sqrt[3]{9}} - \frac{1}{\sqrt[3]{3}\sqrt[3]{18a^2 + \sqrt{3}\sqrt{108a^4 + a^3}}}} \tag{15}$$

2.2. Zadanie 2: pionowy ruch małej kulki w polu grawitacyjnym ze stałymi oporami powietrza

W poprzednim przykładzie równanie trzeciego stopnia (12) zależało od jednego parametru rzeczywistego a, który mógł przyjmować dowolne wartości dodatnie. Na równanie trzeciego stopnia, zależne jedynie od ustalonych współczynników, natykamy się przy rozważaniu zagadnienia ciała wyrzuconego pionowo w górę w polu grawitacyjnym przy obecności stałych sił oporu (pomińmy dyskusję, czy stała siła oporu powietrza zdarza się w realnych sytuacjach).

Sformułujmy zagadnienie: napisz równania na prędkość i drogę podczas ruchu „w górę” i „w dół” dla ciała wystrzelonego pionowo do góry z prędkością początkową o wartości v_0; wyznacz czas wznoszenia i opadania, prędkość w chwili uderzenia o podłoże i maksymalną wysokość, na którą ciało się wzniosło. Zadanie rozwiąż w dwóch przypadkach:

- przy braku oporów powietrza,
- jeśli opory powietrza przyjmują stałą wartość.

W drugim przypadku przebadaj zależność całkowitego czasu ruchu oraz wartości prędkości od wartości siły oporu powietrza.

2.2.1. Ruch bez oporu

Rozważmy najpierw przypadek bez oporu (rysunek 2 z F=0).
W pierwszej fazie ruchu „w górę” równanie II zasady dynamiki Newtona ma postać $ma_g = -mg$ z warunkami początkowymi $s(0) = 0$, $v_g(0) = v_0$. W drugiej fazie ruchu "w dół" mamy $ma_d = mg$ oraz warunki początkowe $s(0) = 0$, $v_g(0) = 0$. Ruch jest ruchem jednostajnie zmiennym i łatwo znajdujemy zależności:[4]

$$v_g(t) = v_0 - gt; \qquad v_d(t) = gt\,; \tag{16}$$
$$s_g(t) = v_0 t - \frac{gt^2}{2}; \qquad s_d(t) = \frac{gt^2}{2};$$

Kiedy ciało osiąga maksymalną wysokość H_0, jego prędkość spada do zera. Oznaczając przez t_g czas wznoszenia, znajdujemy zależności:

$$t_g = \frac{v_0}{g}; \quad H_0 = \frac{v_0^2}{2g}\,; \tag{17}$$

W drugiej fazie ruchu, czyli w trakcie opadania w dół, ciało ma początkową prędkość równą zero. Spada z wysokości, na którą wcześniej się wzniosło, czyli z H_0. Czas opadania t_d znajdujemy zatem z warunku $s_d(t_d) = H_0$, otrzymując:

$$t_d = \sqrt{\frac{2H_0}{g}}. \tag{18}$$

Wykorzystując zależności (17) dostajemy

$$t_g = t_d. \tag{19}$$

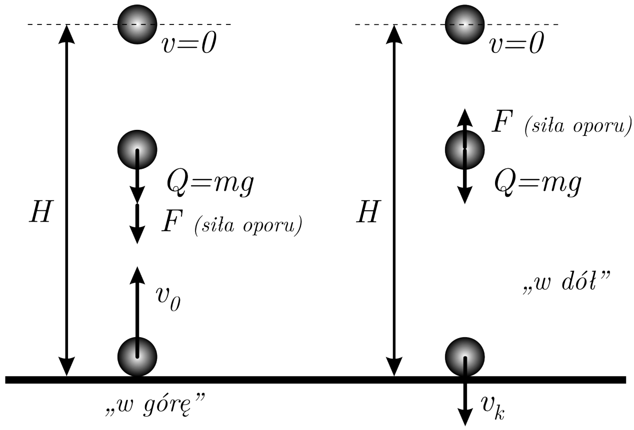
Prędkość uderzenia o podłoże $v_k = v_d(t_d)$, czyli

$$v_k = v_0. \tag{20}$$

Prędkość uderzenia o podłoże jest zatem równa prędkości, z jaką ciało zostało wystrzelone. To zrozumiałe, jako że brak oporów implikuje brak strat energii. Oznaczmy czas wznoszenia i czas opadania w przypadku braku oporów przez t_0. A zatem

$$t_0 := \frac{v_0}{g} = t_g = t_d. \tag{21}$$

[4] Będziemy używać indeksów „g” dla oznaczenia wielkości w trakcie ruchu „w górę” oraz „d” dla ruchu „w dół”.



Rysunek 2. Ruch ciała w górę i w dół
Źródło: opracowanie własne.

Całkowity czas ruchu

$$t_c = t_g + t_d = 2t_0. \tag{22}$$

2.2.2. Ruch ze stałym oporem

Rozważmy teraz sytuację, w której oprócz siły ciężkości działa siła oporu o stałej wartości $F \geqslant 0$, takiej samej podczas obu etapów ruchu (i zawsze przeciwnie skierowanej do chwilowej prędkości ciała). Równania ruchu mają teraz postać:

$$\begin{aligned} ma_g &= -mg - F \quad \Longrightarrow \quad a_g = -g(1+x) \\ ma_d &= mg - F \quad \Longrightarrow \quad a_d = g(1-x), \end{aligned} \tag{23}$$

gdzie oznaczyliśmy $x := \dfrac{F}{mg}$. Zauważmy, że siła oporu musi być mniejsza od ciężaru ciała. W przeciwnym razie w trakcie spadania ciała mogłoby dojść do niefizycznej sytuacji, w której ciało pozostałoby nieruchome (lub w najlepszym przypadku spadało ruchem jednostajnym). Zakładamy zatem, że $F < mg$, czyli $x < 1$. Oczywiście, $x \geqslant 0$.

Stosując równania (23) znajdujemy

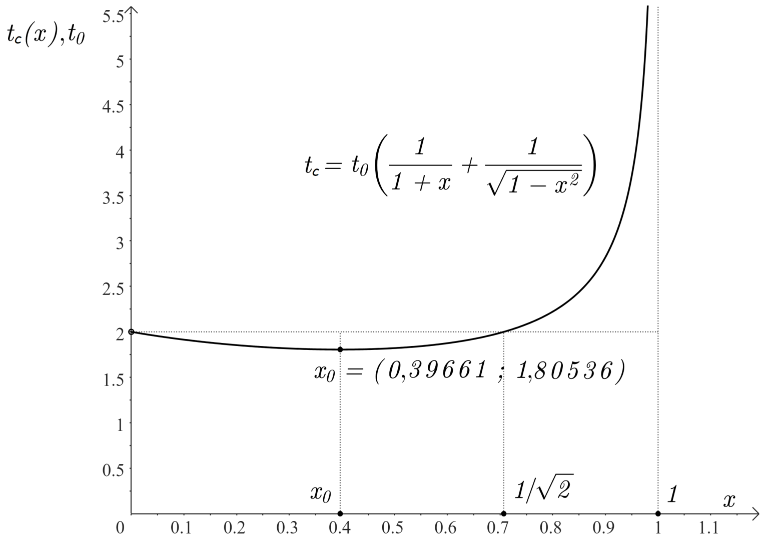
$$\begin{aligned} v_g(t) &= v_0 - (1+x)gt; \qquad & v_d(t) &= (1-x)gt; \\ s_g(t) &= v_0 t - (1+x)\frac{gt^2}{2}; \qquad & s_d(t) &= (1-x)\frac{gt^2}{2}; \end{aligned} \tag{24}$$

Podobnie jak w przypadku ruchu bez oporu, łatwo znajdujemy czas wznoszenia, wiedząc, że $v_g(t_g) = 0$, co daje:

$$t_g = \frac{v_0}{g(1+x)} = \frac{t_0}{1+x} \,. \tag{25}$$

Po czasie t_g ciało wznosi się na wysokość $H = s_g(t_g)$, czyli:

$$H = \frac{1}{1+x}\frac{v_0^2}{2g} = \frac{H_0}{1+x} \,. \tag{26}$$



Rysunek 3. Wykres funkcji $t_C(x)$

Źródło: opracowanie własne.

Aby wyznaczyć czas opadania, wykorzystajmy fakt, że $s_d(t_d) = H$, skąd otrzymujemy

$$t_d = \frac{v_0}{g\sqrt{1-x^2}} = \frac{t_0}{\sqrt{1-x^2}}. \tag{27}$$

Ciało uderza o podłoże z prędkością $v_k = v_d(t_d)$, co daje:

$$v_k = v_0\sqrt{\frac{1-x}{1+x}}. \tag{28}$$

Zauważmy, że $t_g < t_0 < t_d$. Dokładniej, zachodzi relacja

$$t_d = t_g\sqrt{\frac{1+x}{1-x}}. \tag{29}$$

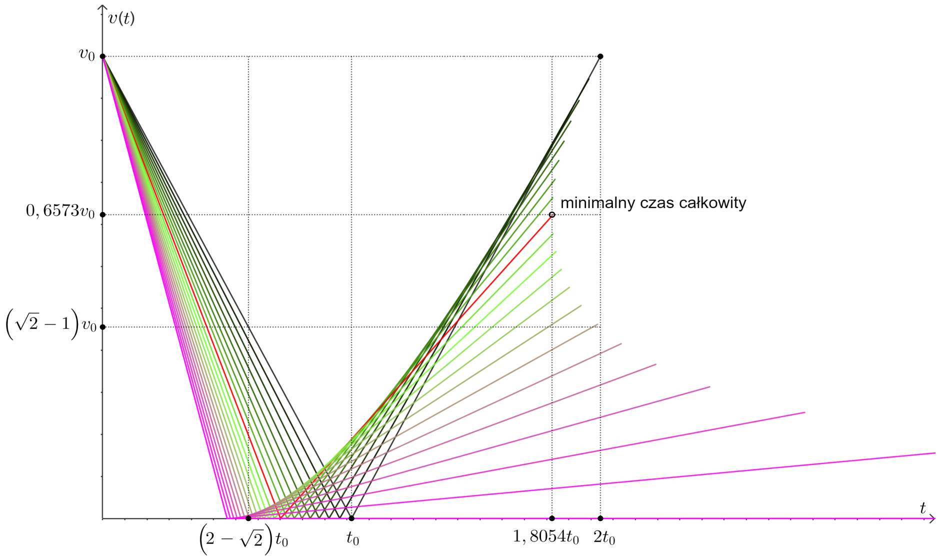
Czas całkowity wyraża się z kolei wzorem:

$$t_c(x) = t_0\left(\frac{1}{1+x} + \frac{1}{\sqrt{1-x^2}}\right). \tag{30}$$

Wykres funkcji $t_c(x)$ jest przedstawiony na Rysunku 3.

Z wykresu funkcji $t_c(x)$ widzimy, że dla pewnej wartości zmiennej x czas całkowity jest taki sam, jak w przypadku ruchu bez oporów. Aby znaleźć tę wartość, musimy rozwiązać równanie $t_c = 2t_0$, które daje:

$$x(1-2x^2) = 0 \tag{31}$$



Rysunek 4. Wykres prędkości w funkcji czasu v(t)
Źródło: opracowanie własne.

Równanie to ma trzy rozwiazania, z których pierwsze $x = -\dfrac{\sqrt{2}}{2}$ nie ma sensu fizycznego, drugie $x = 0$ odpowiada oczywistemu przypadkowi ruchu bez oporów, a trzecie, najbardziej nas interesujace, to $x = \dfrac{\sqrt{2}}{2}$. Zatem, jeśli siła oporu ma wartość $F = \dfrac{\sqrt{2}}{2}mg$, czas ruchu jest dokładnie taki sam, jak w przypadku braku oporów. Oczywiście w takim przypadku ciało wznosi się na wysokość niższą niż H_0, a dokładniej równą $H_0(2-\sqrt{2})$, oraz uderza w podłoże z prędkością $(\sqrt{2}-1)v_0$.

Teraz rozważmy zagadnienie, które doprowadzi nas do równania trzeciego stopnia. Zauważmy, że funkcja $t_c(x)$ ma minimum. Istnieje zatem taka wartość siły oporu, dla której całkowity czas ruchu jest najkrótszy. Minimum funkcji $t_c(x)$ znajdziemy przyrównując do zera jej pochodną. Dostajemy równanie:

$$2x^3 - 2x^2 + 3x - 1 = 0. \tag{32}$$

Równanie to ma jeden pierwiastek rzeczywisty, równy

$$x_0 = \frac{1}{3}\left(1 + \frac{\sqrt[3]{4+3\sqrt{78}}}{\sqrt[3]{4}} - \frac{7}{\sqrt[3]{2(4+3\sqrt{78})}}\right) \approx 0,39661. \tag{33}$$

Jest to interesujący wynik: jeśli siła oporu wynosi $F \approx 0,39661mg$, całkowity czas ruchu jest najkrótszy. Dokładniej, dla $x = x_0$ mamy:

$$t_g \approx 0,71602t_0, \quad t_d \approx 1,08934t_0, \quad t_c \approx 1,80534t_0, \quad v_k \approx 0,6573v_0, \quad H \approx 0,71602H_0 \tag{34}$$

Na rysunku 4 przedstawiony jest wykres wartości wektora prędkości w funkcji czasu dla kilku różnych wartości parametru x.

3. PODSUMOWANIE

Opisane w paragrafach 2.1 i 2.2 przykłady pokazują, że równania sześcienne pojawiają się nawet w prostych zagadnieniach fizycznych. Problem dwóch odpychających się kul to zadanie – na poziomie fizyki zagadnienia – stopniem trudności nie wykraczające poza ramy programu szkoły średniej. Jednak nawet takie zadanie stawia przed rozwiązującym je uczniem równanie stopnia trzeciego. Bez znajomości metody analitycznej lub bez dostępu do pakietów do obliczeń symbolicznych znalezienie rozwiązania (15) przez ucznia szkoły średniej lub nawet przez studenta pierwszego roku studiów technicznych jest, naszym zdaniem, niemożliwe. Podobnie jest z rozwiązaniem (33) bardziej wyrafinowanego problemu minimalnego czasu ruchu dla ciała poruszającego się w polu grawitacyjnym w obecności stałych sił oporu powietrza.

Podsumowując, warto podkreślić, że metodom rozwiązywania równań sześciennych powinno się poświęcać trochę więcej czasu, nawet jeśli nie na poziomie szkoły średniej, to przynajmniej na pierwszych latach studiów technicznych.

LITERATURA

[1] van der Waerden B.L., *Geometry and Algebra in Ancient Civilizations*, Springer-Verlag, Berlin 1983.

[2] Gładki P., *Uwagi historyczne o wzorach Cardano*, online http://www.math.us.edu.pl/pgladki/faq/node129.html (dostęp: 21.10.2024).

[3] Bronsztejn I.N., Siemiendajew K.A., *Matematyka. Poradnik encyklopedyczny*, PWN, Warszawa 2019.

[4] Ramond P., *Abel-Ruffini's Theorem: Complex but Not Complicated!*, "The American Mathematical Mononthly" 2022, vol. 129, no. 3.

[5] Cedrik M.S., *Ćwiczenia z fizyki dla kandydatów na wyższe uczelnie*, PWN, Warszawa 1974.

[6] Cedrik M.S., *Zadania z fizyki*, PWN, Warszawa 1986.

[7] Weast R.C., *Handbook of Chemistry and Physics (53rd Edn.)*, Chemical Rubber Company, Cleveland 1972.

[8] Wolfram Alpha Widgets, *Solutions of cubic equations*, online https://www.wolframalpha.com/widgets/view.jsp?id=3f4366aeb9c157cf9a30c90693eafc55 (dostęp: 21.10.2024).

JOANNA KUCNER[*1], BOŻENA ŚWIĄTEK[†1]

O PEWNYM RÓWNANIU W ZBIORZE LICZB ZESPOLONYCH

[1] óentre óf óathematics and Physics, óodz University óf Technology, Al. Politechniki ó1, ó3-590 Łód´z, Poland

Artykuł opisuje pewne równanie w zbiorze liczb zespolonych, które można rozwiązać kilkoma metodami odwołującymi się do różnych własności (zwłaszcza liczb zespolonych). Autorki wskazują także dydaktyczne zalety zapoznania studentów z tymi metodami.

Keywords: liczby zespolone, równania wielomianowe.

1. WSTĘP

Liczby zespolone zostały wprowadzone w XVI wieku na potrzeby algebry. Okazało się bowiem, że poprzez dopuszczenie pierwiastków z liczb ujemnych można wyprowadzać ogólne wzory do rozwiązywania równań trzeciego stopnia [1]. Liczby zespolone zostały następnie wykorzystane w innych zagadnieniach matematycznych, np. w rozwiązywaniu równań różniczkowych [2], w obliczaniu całek oznaczonych z funkcji rzeczywistych przy pomocy residuów [3]. Po wielu latach okazało się, że ten tak pozornie abstrakcyjny dział matematyki ma zastosowanie w wielu innych dziedzinach nauki. Przykładowo w elektrotechnice liczby zespolone przydają się one do analizy obwodów elektrycznych prądu przemiennego, wyrażając jego moc czynną oraz bierną [4]. Ponadto ułatwiają obliczanie prądów zwarciowych [4], co w praktyce przekłada się na płynną i nieprzerwaną dystrybucję prądu, np. do mieszkań.

Liczby zespolone pomagają w analizowaniu i modelowaniu złożonych zagadnień, które pojawiają się w grafice komputerowej i modelowaniu 3D [5], w funkcjach falowych w fizyce kwantowej [6], w dynamice płynów i aerodynamice [7], w teorii sygnałów i przetwarzania danych [8], w teorii chaosu i dynamiki nieliniowej [9]. Liczby zespolone są też kluczowe przy tworzeniu efektów filmowych, a ze względu na liczne zastosowania są wbudowane w niektórych językach programowania, np. w Fortran czy Python [10].

To wszystko sprawia, że liczby zespolone są tematem omawianym na wielu kierunkach na uczelniach wyższych. Jednak liczba godzin przeznaczona na to zagadnienie pozwala przedstawić słuchaczom tylko podstawy pozwalające posługiwać się liczbami zespolonymi, bez poruszania takich zagadnień, jak na przykład analiza zespolona.

Jednym z tematów, jakie wchodzą w zakres realizowanego materiału, jest rozwiązywanie równań w zbiorze liczb zespolonych [11]. W poniższym artykule chcemy rozważyć jedno z takich równań, które rozwiązać możemy na kilka sposobów z których każdy ma pewne walory dydaktyczne.

[*] joanna.kucner@p.lodz.pl
[†] bozena.swiatek@p.lodz.pl

2. ZESPOLONE RÓWNANIE WIELOMIANOWE

2.1. Metoda pierwsza

Rozważmy równanie:

$$z^3 - 8 = 0. \tag{1}$$

Jest ono oczywiście równoważne równaniu:

$$z^3 = 8, \tag{2}$$

a zatem, aby znaleźć rozwiązania równania (2), wystarczy policzyć pierwiastki trzeciego stopnia z liczby 8. Można to zrobić wykorzystując postać trygonometryczną liczby zespolonej. Łatwo policzyć, że $|8| = 8$, a argument tej liczby wynosi: $\phi = 0$. Obliczamy zatem pierwiastki z liczby 8 korzystając ze wzoru na pierwiastki n-tego stopnia z liczby zespolonej $z = |z|(\cos\phi + i\sin\phi)$, które możemy zapisać w postaci [11]

$$w_k = \sqrt[n]{|z|}(\cos\frac{\phi + 2k\pi}{n} + i\sin\frac{\phi + 2k\pi}{n}), \tag{3}$$

gdzie $k = 0, 1, ..., n-1$. A zatem

$$\begin{aligned}
w_0 &= \sqrt[3]{8}\left(\cos\frac{0+2\cdot 0\cdot\pi}{3} + i\sin\frac{0+2\cdot 0\cdot\pi}{3}\right) = 2(\cos 0 + i\sin 0) = 2(1+0i) \\
&= 2 \\
w_1 &= \sqrt[3]{8}\left(\cos\frac{0+2\cdot 1\cdot\pi}{3} + i\sin\frac{0+2\cdot 1\cdot\pi}{3}\right) = 2\left(\cos\frac{2}{3}\pi + i\sin\frac{2}{3}\pi\right) \\
&= -1 + \sqrt{3}i \\
w_2 &= \sqrt[3]{8}\left(\cos\frac{0+2\cdot 2\cdot\pi}{3} + i\sin\frac{0+2\cdot 2\cdot\pi}{3}\right) = 2\left(\cos\frac{4}{3}\pi + i\sin\frac{4}{3}\pi\right) \\
&= -1 - \sqrt{3}i.
\end{aligned} \tag{4}$$

2.2. Metoda druga

Drugi – i chyba najkrótszy – sposób rozwiązania równania (1) polega na skorzystaniu ze wzoru na różnicę sześcianów:

$$\begin{gathered}
z^3 - 8 = 0 \iff z^3 - 2^3 = 0 \iff (z-2)(z^2 + 2z + 4) = 0 \iff \\
\iff z = 2 \qquad \vee \qquad z^2 + 2z + 4 = 0 \iff \\
\iff z_1 = 2 \qquad \vee \qquad z_2 = -1 - \sqrt{3}i \qquad \vee \qquad z_3 = -1 + \sqrt{3}i.
\end{gathered}$$

2.3. Metoda trzecia

W trzecim sposobie rozwiązania równania (1) wykorzystamy postać kartezjańską liczby zespolonej z. Możemy zapisać, że

$$z = x + iy,$$

dla pewnych $x, y \in \mathbb{R}$. Wówczas równanie (1) przyjmie postać:

$$(x+iy)^3 - 8 = 0 \implies x^3 + 3x^2yi - 3xy^2 - y^3i = 8. \tag{5}$$

Liczby zespolone są sobie równe wtedy i tylko wtedy gdy mają równe części rzeczywiste i urojone, a zatem:

$$x^3 - 3xy^2 = 8, \tag{6a}$$
$$3x^2y - y^3 = 0. \tag{6b}$$

Zajmijmy się równaniem (6b):

$$y(3x^2 - y^2) = 0 \iff y = 0 \quad \vee \quad (\sqrt{3}x - y)(\sqrt{3}x + y) = 0 \iff$$

$$\iff y = 0 \ \vee \ y = \sqrt{3}x \ \vee \ y = -\sqrt{3}x.$$

Rozpatrzmy równanie (6a):

- dla $y = 0$ równanie (6a) przyjmuje postać $x^3 - 8 = 0$, zatem $x = 2$;
- dla $y = \sqrt{3}x$ równanie (6a) przyjmuje postać $x^3 - 3x \cdot 3x^2 - 8 = 0$. Doprowadzając lewą stronę do najprostszej postaci otrzymujemy $-8x^3 - 8 = 0$, a stąd $x = -1$;
- dla $y = -\sqrt{3}x$ równanie (6a) przyjmuje taką samą postać jak w poprzednim przypadku, zatem i tu również $x = -1$.

Podsumowując, mamy trzy rozwiązania układu równań (6a) i (6b), (a więc i równania (1)):

$$z_1 = 2, \qquad z_2 = -1 + \sqrt{3}i, \qquad z_3 = -1 - \sqrt{3}i.$$

2.4. Metoda czwarta

W czwartym sposobie rozwiązania równania (1) wykorzystamy fakt, że pierwiastki n-tego stopnia z liczby zespolonej z dzielą okrąg o promieniu $r = \sqrt[n]{|z|}$ na n równych łuków [11]. Jedno rozwiązanie równania (1) łatwo jest odgadnąć:

$$z_1 = 2$$

(argument główny tej liczby to $\phi = 0$). Zatem pozostałe pierwiastki z liczby 8 muszą być liczbami zespolonymi o argumentach $\frac{2}{3}\pi$ oraz $\frac{4}{3}\pi$ (bo $2\pi : 3 = \frac{2}{3}\pi$, $0 + \frac{2}{3}\pi = \frac{2}{3}\pi$ oraz $\frac{2}{3}\pi + \frac{2}{3}\pi = \frac{4}{3}\pi$). Stąd obliczamy (wykorzystując wzory redukcyjne) pozostałe pierwiastki równania:

$$z_2 = 2(\cos\frac{2}{3}\pi + i\sin\frac{2}{3}\pi) = 2[\cos(\pi - \frac{\pi}{3}) + i\sin(\pi - \frac{\pi}{3})] = 2(-\cos\frac{\pi}{3} + i\sin\frac{\pi}{3}) =$$

$$= 2(-\frac{1}{2} + \frac{\sqrt{3}}{2}i) = -1 + \sqrt{3}i,$$

oraz analogicznie

$$z_3 = 2(\cos\frac{4}{3}\pi + i\sin\frac{4}{3}\pi) = 2[\cos(\pi + \frac{\pi}{3}) + i\sin(\pi + \frac{\pi}{3})] = 2(-\cos\frac{\pi}{3} - i\sin\frac{\pi}{3}) =$$

$$= 2(-\frac{1}{2} - \frac{\sqrt{3}}{2}i) = -1 - \sqrt{3}i.$$

3. PODSUMOWANIE

Od wielu lat prowadzimy zajęcia dla studentów na różnych wydziałach Politechniki Łódzkiej, na których matematyka nie jest przedmiotem kierunkowym. Zazwyczaj jej nauczanie ogranicza się do dwóch semestrów, z czego łącznie kilkanaście godzin przeznaczone jest na wykład i ćwiczenia z liczb zespolonych. Ponadto wśród naszych studentów dość rzadko spotyka się osoby szczególnie interesujące się matematyką. Niech świadczy o tym fakt, że niewielu z nich zadaje proste pytanie: "Po co są liczby zespolone?", a jeszcze mniej z nich zna na nie odpowiedź. Dobrze więc, że zagadnienie to jest omawiane na zajęciach.

Musimy przyznać, że podstawy teorii liczb zespolonych (działania oraz równania w zbiorze liczb zespolonych) są często przyjmowane przez słuchaczy, którzy po raz pierwszy się z nimi spotkają, z pewną ciekawością, jako coś zaskakującego (np. fakt, że w zbiorze liczb zespolonych każde równanie kwadratowe ma przynajmniej jedno rozwiązanie czy też fakt, że pierwiastek n-tego stopnia z liczby zespolonej różnej od 0 ma n różnych wartości). Pewną trudność stanowi jednak konieczność posługiwania się w niektórych zadaniach trygonometrią (zwłaszcza wzorami redukcyjnymi).

Odnosząc się bezpośrednio do równania (1), uważamy, że warto poświęcić czas na rozwiązanie tego równania (bądź równania o podobnym stopniu trudności) na ćwiczeniach wszystkimi przedstawionymi metodami. Pierwszy sposób pozwala na nowo przypomnieć wprowadzoną na wykładzie definicję pierwiastka n-tego stopnia w zbiorze liczb zespolonych. Z naszych obserwacji wynika, że studenci rzadko potrafią skojarzyć rozwiązywanie tego równania z koniecznością obliczenia pierwiastków trzeciego stopnia z liczby 8. Oczywiście zaletą dydaktyczną tego sposobu jest również utrwalenie pojęcia postaci trygonometrycznej liczby zespolonej i wzoru Moivre'a.

Drugi sposób jest okazją do przypomnienia, że w zbiorze liczb zespolonych każde równanie kwadratowe ma rozwiązanie oraz powtórzenie sposobu rozwiązywania takich równań. W dalszym toku nauczania matematyki będzie to dla studentów niezbędną umiejętnością, choćby przy zagadnieniach związanych z równaniami różniczkowymi liniowymi rzędu drugiego. Drugi sposób daje również sposobność poznania lub przypomnienia wzoru na różnicę sześcianów, który pozwala uniknąć procedury dzielenia wielomianów.

Trzeci z kolei sposób jest okazją do przypomnienia warunku na równość dwóch liczb zespolonych zapisanych w postaci kartezjańskiej. Ponadto zaletą pod względem dydaktycznym jest tu konieczność rozwiązywania otrzymanego nieliniowego układu równań i właściwego zinterpretowania otrzymanych wyników, z czym nierzadko studenci spotykają się po raz pierwszy.

Czwarty sposób pozwala przypomnieć geometryczną interpretację pierwiastków z liczby zespolonej. Z naszego doświadczenia wynika, że przez te wszystkie lata tylko jeden student połączył taką interpretację z rozwiązaniem równania (1), mimo że było to zagadnienie omawiane na wcześniejszych zajęciach.

Reasumując: naszym zdaniem – mimo małej ilości czasu, jaką można na zajęciach ze studentami przeznaczyć na kwestie związane z liczbami zespolonymi – warto w ramach podsumowania tej tematyki rozwiązać zaproponowane przez nas równanie (1) każdą z przedstawionych metod. Uważamy, że warto to zrobić nie tylko w celu uświadomienia studentom, że istnieją różne sposoby podejścia do tego samego zadania, ale że jest to również okazja do utrwalenia pewnych zagadnień, pojęć i technik z teorii liczb zespolonych.

LITERATURA

[1] Hołubowski W., Pleszczyński M., Różański M., Słowik R., Smuda A., Wituła R., *Liczby zespolone. Tom 1. Podstawowe operacje na liczbach zespolonych*, Wydawnictwo Politechniki Śląskiej, Gliwice 2023.

[2] Krysicki W., Włodarski L., *Analiza matematyczna w zadaniach, cz. 2*, Wydawnictwo PWN, Warszawa 2011.

[3] Kącki E., Siewierski L., *Wybrane działy matematyki wyższej z ćwiczeniami*, Wydawnictwo Wyższej Szkoły Informatyki w Łodzi, Łódź 2002.

[4] Bolkowski S., *Teoria obwodów elektrycznych*, Wydawnictwo WNT Warszawa 2016.

[5] Vince J., *Mathematics for Computer Graphics*, Springer-Verlag, London 2006.

[6] Kryszewski S., *Mechanika kwantowa*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2018.

[7] Jeżowiecka K., Szewczyk H., *Mechanika płynów*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2001.

[8] Oppenheim A.V., Schafer R.W., *Cyfrowe przetwarzanie sygnałów*, Wydawnictwa Komunikacji i Łaczności, Sulejówek 1979.

[9] Awrejcewicz J., *Tajemnice nieliniowej dynamiki*, Wydawnictwo Politechniki Łódzkiej, Łódź 1997.

[10] Saha A., *Matematyka w Pythonie*, Wydawnictwo Helion, Gliwice 2021.

[11] Dobrowolska K., Dyczka W., Jakuszenkow H., *Matematyka dla studentów studiów technicznych cz. 1*, Wydawnictwo Helpmath, Łódź 1996.

MICHAŁ WASIAK[1]

AN ANALYSIS OF THE THEORY OF TRANSFER LENGTH METHOD

[1] I nstitute of Physics, Lodz University of Technology, ul. Wólczańska 227/221, ó3-005 Łódź, Poland

In this paper we present derivation of the equations describing the resistance between the electrodes used in The Transfer Length Method, both for the linear and circular configurations. For the circular configuration, we have obtained an approximated formula which is more accurate than the most widely used formula. Conditions allowing for verification if the approximations can be applied are also presented.

Keywords: Transfer Length Method, modified Bessel functions.

1. INTRODUCTION

Since the 1960s the Transmission Line Method (TLM) has been a basic method for deremination of the contact resistivity of metalic electrodes deposited on a semiconductor [1–4]. It allows also for determination of the sheet resistance of the semiconductor material on which the electrode is deposited. Usually, the measurement of the resistivity of the semiconductor is performed if the semiconductor layer is thin relative to the lateral dimensions of the electrodes because in this case the theory and hence the interpretation of the experimental data is much simpler than in a more general case [2].

TLM measurements are performed using a set of either rectangular electrodes (usually called simply a TLM measurement), or using electrodes in form of concentric rings (circular TLM or cTLM). In this paper, a detailed derivation of the formulas used for interpretation of the experimental data and a discussion of the important approximations often used in practical applications is presented.

The solutions for equations describing the current flow between the TLM electrodes are distinctly different in the case of linear and cylindrical configurations. First, a simpler case will be analysed.

2. LINEAR TLM

A linear TLM structure consists of a series of rectangular electrodes separated by a distance which is different for each to adjacent electrodes. In Figure 1 a top view of a part of such a structure is presented.

For each of the electrode pair, a current-voltage (I–V) curve is measured. Assuming that this curve is linear, the electrical resistance R for this pair can be determined. This resistance is a result of the resistance caused by the contact between the metal electrode and the semiconductor it is deposited on, and by the resistance of the semiconductor itself. The resistance of the metal can be neglected. In what follows, we will derive a formula for R. We will use the following approximation (see Figure 1 for the definitions of the symbols):

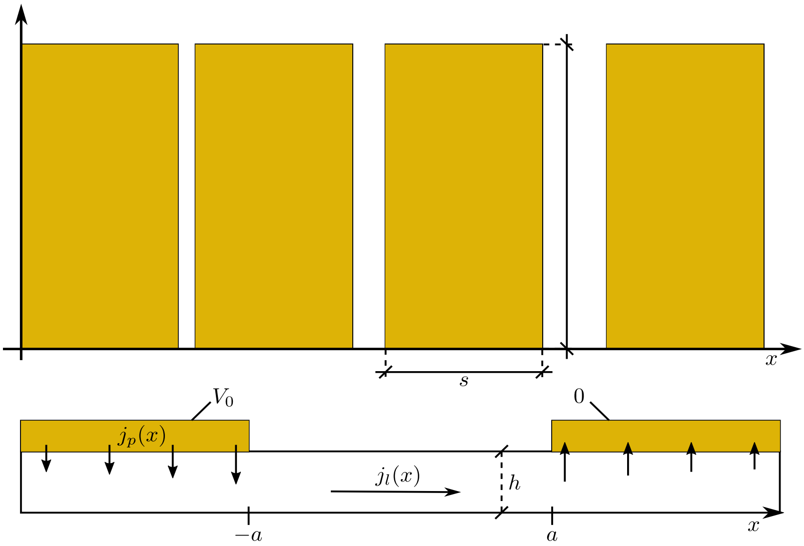


Figure 1. On the top: Top view of part of a linear TLM structure. On the bottom: a side cross-section of an area between two neighbouring electrodes
Source: own work.

- The thickness h of the semiconductor layer is very small, so that we can assume that current flows uniformly in the vertical direction and hence we can neglect this dimension in our calculations;
- The current density does not depend on y. This can be true if the distances between the electrodes are much smaller than their width, or if the TLM structure is fabricated in such a way that the current cannot spread beyond the area determined by the electrodes' width;
- The electrodes are semi-infinite in the x direction (i.e. in Figure 1 $s = \infty$); the potential of the left electrode if V_0, and 0 on the right electrode. Later, we will discuss the case in which $s < \infty$.

The density of the vertical current which flows from the electrode into the semiconductor (or the other way round) depends on x and can be expressed as:

$$j_p(x) = \varsigma_c\big(V_c - u(x)\big) \tag{1}$$

where $u(x)$ is the local electric potential; V_c is the electrode potential (equal to either V_0 or 0); and ς_c is the surface conductivity of the electrode–semiconductor interface (the unit of ς_c is $\mathrm{S/m^2}$). The above formula is valid if x refers to a point covered by the electrode, otherwise $j_p = 0$. The horizontal current I_l and its density j_l for any x are given by the following formula:

$$j_l(x) = -\sigma_s u'(x) \tag{2}$$

$$I_l(x) = hwj_l(x) \tag{3}$$

where σ_s is the electrical conductivity of the semiconductor material. On the other hand, the horizontal current in the area below an electrode (i.e. for $|x| > a$) is related to j_p in the following manner:

$$I_l(x) = \begin{cases} w \displaystyle\int_{-\infty}^{x} j_p(\xi)\, d\xi & \text{for } x < -a \\ I_l(a) + w \displaystyle\int_{a}^{x} j_p(\xi)\, d\xi & \text{for } x > a \end{cases} \tag{4}$$

We differentiate both sides of the above equation with respect to x, substitute (3) and we get the following equation (valid both for $x > a$ and $x < -a$):

$$hwj_l'(x) = wj_p(x) \tag{5}$$

By reducing w and substituting (1) and (2) we get:

$$u''(x) - \gamma^2 u(x) + \gamma^2 V_c = 0 \tag{6}$$

where

$$\gamma^2 = \frac{\varsigma_c}{h\sigma_s} \tag{7}$$

We could assume that the conductivities of the left and the right electrode are different ($\varsigma_c^{(1)}$ and $\varsigma_c^{(2)}$ respectively). Then we would have coefficients γ_1 for $x < -a$ and γ_2 for $x > a$. The general solutions in these two areas are as follows:

$$u_1(x) = C_1 \exp(\gamma_1 x) + \tilde{C}_1 \exp(-\gamma_1 x) + V_0 \qquad \text{for } x < -a \tag{8}$$

$$u_2(x) = \tilde{C}_2 \exp(\gamma_2 x) + C_2 \exp(-\gamma_2 x) \qquad \text{for } x > a \tag{9}$$

For $x \to -\infty$ potential u has to tend to V_0, while for $x \to \infty$ to 0 (which are the potential of the respective electrodes). Otherwise, an infinite current would flow through the structure. It means that both coefficients $\tilde{C}$ must vanish and hence:

$$u_1(x) = C_1 \exp(\gamma_1 x) + V_0 \qquad \text{for } x < -a \tag{10}$$

$$u_2(x) = C_2 \exp(-\gamma_2 x) \qquad \text{for } x > a \tag{11}$$

Between the electrodes (for $|x| < a$), current $I_l(x)$ and hence its density $j_l(x)$ must be constant and equal, respectively, to I (which is the total current flowing between the electrodes) and j. Using Eq. (2) we conclude that in this area $u''(x) = 0$ which means that u is a linear function:

$$u(x) = -\frac{j}{\sigma_s}\, x + b \quad \text{for } |x| < a \tag{12}$$

Continuity of $j_l(x)$ end Eqs. (2), (10) and (11) give us the following relation:

$$-C_1 \sigma_s \gamma_1 \exp(-\gamma_1 a) = j = C_2 \sigma_s \gamma_2 \exp(-\gamma_2 a) \tag{13}$$

and hence:

$$\frac{C_1}{C_2} = -\frac{\gamma_2}{\gamma_1} \exp\big(- (\gamma_2 - \gamma_1) a\big) \tag{14}$$

Continuity of $u(x)$ at $x = -a$ and $x = a$ provides further two equations which can be used to determine parameters C_1, C_2, b, j. The parameter we really need is j, since it gives us the total current I. After some elementary calculations we get:

$$\frac{V_0}{j} = \frac{1}{\sigma_s} \left(2a + \gamma_1^{-1} + \gamma_2^{-1}\right) \tag{15}$$

Let $L = 2a$ denote the distance between the electrodes. The resistance between them is equal to V_0/I, so after substituting (7) we obtain:

$$R(L) = \frac{L}{hw\sigma_s} + \frac{1}{w\sqrt{h\sigma_s}}\left(\frac{1}{\sqrt{\varsigma_{c1}}} + \frac{1}{\sqrt{\varsigma_{c2}}}\right) = \frac{L}{hw\sigma_s} + \frac{2\sqrt{\varrho_c}}{w\sqrt{h\sigma_s}} \tag{16}$$

where

$$2\sqrt{\varrho_c} = \frac{1}{\sqrt{\varsigma_{c1}}} + \frac{1}{\sqrt{\varsigma_{c2}}} \tag{17}$$

Since Eq. (15) γ_1^{-1} and γ_2^{-1} are simply added, in the linear TLM experiment it is impossible to determine separately the resistivities of the individual electrodes. Instead, we can use a single γ. However, in principle these resistivities should be the same, so this is not considered a problem.

In the TLM experiment $R(L)$ is measured for several values of L. This relation should be linear, as predicted by Eq. (16), and the experimental data provide parameters α and R_0 in the following relation:

$$R(L) = \alpha L + R_0 \tag{18}$$

Combining the above experimental parameters and the theoretical relation we obtain two relations for the semiconductor electrical conductivity (in the horizontal direction) and the contact resistivity:

$$\sigma_s = \frac{1}{hw\alpha} \tag{19a}$$

$$\varrho_c = \frac{1}{4}R_0^2 w^2 h\sigma_s = \frac{wR_0^2}{4\alpha} \tag{19b}$$

The simple and practical formula (19) was derived under the assumption that the electrodes can be considered semi-infinite along x direction. Now we will provide a criterion to verify if this assumption is valid in individual cases.

2.1. Verification if the semi-infinite electrodes assumption is valid

The relations derived so far assume that the current flowing from an electrode to the semiconductor uses the whole infinite length of the electrode, while in fact the available length is equal to s. Let us calculate the current which flows through the part of an electrode which in fact does not exist. Integrating formula (1) with the proper limits and using Eq. (13) and (2) we obtain the following formulas for the total current I and the current injected to the actually existing part of the second electrode (equal to the current injected by the actually existing part of the first electrode):

$$I = \int_a^\infty \varsigma_c u(x) = \frac{wj}{\sigma_s\gamma^2} \tag{20}$$

$$I_s = \int_a^{a+s} \varsigma_c u(x) = \frac{wj}{\sigma_s\gamma^2}(1 - \exp(-\gamma s)) \tag{21}$$

If the difference between those two currents is negligibly small (relative to the total current), the simplification we used can be considered valid. This gives the following criterion which can be used for such a validation:

$$\exp(-\gamma s) \ll 1 \iff \exp\left(-\frac{s}{\sqrt{h\varrho_c\sigma_c}}\right) = \exp\left(-\frac{2s\alpha}{R_0}\right) \ll 1 \tag{22}$$

What if one cannot apply the considered simplification?

2.2. Solution for finite electrodes

The problem with finite electrodes can be solved. The differential equations derived above are still valid. What is different is the boundary conditions. Instead of

$$\lim_{x\to\pm\infty} j_l(x) = 0 \tag{23}$$

now we have to use

$$\lim_{x\to\pm(a+s)} j_l(x) = 0 \tag{24}$$

In this case, we will use a more convenient form of the general solution (6)

$$u_1(x) = A_1 \cosh\big(\gamma_1(x-x_1)\big) + \tilde{A}_1 \sinh\big(-\gamma_1(x-x_1)\big) + V_0 \qquad \text{for } x < -a \tag{25}$$

$$u_2(x) = \tilde{A}_2 \cosh\big(\gamma_2(x-x_2)\big) + A_2 \sinh\big(-\gamma_2(x-x_2)\big) \qquad \text{for } x > a \tag{26}$$

where $x_1 = -a - s$ and $x_2 = a + s$ are the outer boundaries of the electrodes. Then, using (24) one can see that $\tilde{A}_1 = \tilde{A}_2 = 0$. The properties of the hiperbolic functions allow us to write (13) in the following way:

$$-A_1\sigma_s\gamma_1 \sinh(\gamma_1 s) = j = A_2\sigma_s\gamma_2 \sinh(\gamma_2 s) \tag{27}$$

Relation (12) remains true and combining it with the above formula we obtain a more general version of Eq. (15):

$$\frac{V_0}{j} = \frac{1}{\sigma_s}\left(2a + \gamma_1^{-1}\coth(\gamma_1 s) + \gamma_2^{-1}\coth(\gamma_2 s)\right) \tag{28}$$

Because $\coth(x) \xrightarrow[x\to\infty]{} 1$, under condition (22) the above formula reduces to the form presented in Eq. (15).

The exact solution does not change the formula for σ_s, however ϱ_c is given in an implicit form. The non-simplified version of relations (19) can be expressed in the following form, assuming that $\gamma_1 = \gamma_2 = \gamma$:

$$\sigma_s = \frac{1}{hw\alpha} \tag{29a}$$

$$R_0 = \frac{2\coth(\gamma s)}{hw\sigma_s\gamma} = 2\sqrt{\frac{\alpha\varrho_c}{w}}\coth\left(s\sqrt{\frac{w\alpha}{\varrho_c}}\right) \tag{29b}$$

In the analysis of linear TLM presented above, we always assumed that the current flow is bounded to the area of the width equal to the electrodes' width. This is a sound assumption only if the distance between the electrodes is much smaller than the electrode width. However, if the conductivity of the semiconductor material is high, the resistance between the electrodes can be vary small relative to the contact resistance. This can severely deteriorate the accuracy of the determination of the semiconductor conductivity. In such a case, a cTLM measurement may be a better choice.

3. CIRCULAR TLM

A single cTLM electrode pair has a form of two concentric rings, where the inner ring is usually a full disk, as shown in Figure 2. In a cTLM experiment the resitivities for several such pairs, with different dimensions, are measured. In this section, we will derive a theoretical formula for the resistivity of such a pair. In our derivation we will use polar co-ordinates. We will also use the first of the three approximations used in section 2.

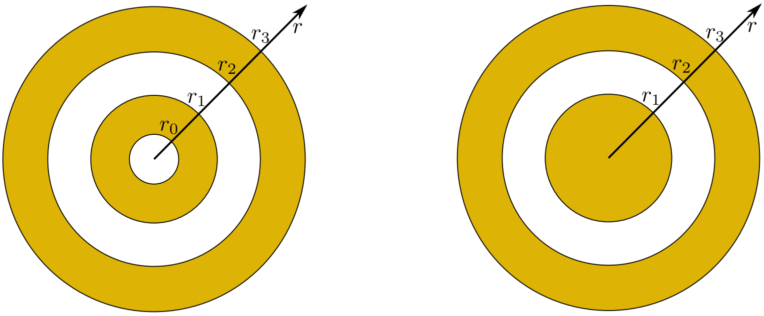


Figure 2. A top view of two posible configurations of an electrode pair used in the cTLM experiment
Source: own work.

The density of the (horizontal) current flowing between the electrodes through the semiconductor has only the radial component j_l, which is independent from the angle, because of the rotational symmetry of the system. Similarly, the electrical potential u depends only on the radial co-ordinate r. In polar co-ordinates, the analogues to formulas (1) and (2) have the following form:

$$j_p(r) = \varsigma_c\big(V_c - u(r)\big) \tag{30a}$$

$$j_l(r) = -\sigma_s u'(r) \tag{30b}$$

where the symbols have the same meaning as in section 2. We will consider a circular sector with its central angle φ. For $\varphi = 2\pi$ we will get the whole plane. The formula that describes the relation between the current flowing through this sector and its density looks as follows:

$$I_l(r) = hr\varphi j_l(r) \tag{31}$$

In the area beneath the electrodes we additionally have the following relations:

$$I_l(r) = \begin{cases} \varphi \displaystyle\int_{r_0}^{r} \xi j_p(\xi)\, d\xi & \text{for } r_0 < r < r_1 \\ I_l(r_2) + \varphi \displaystyle\int_{r_2}^{r} \xi j_p(\xi)\, d\xi & \text{for } r_2 < r < r_3 \end{cases} \tag{32}$$

By differentiating formulas (31) and (32), substituting Eq. (30) and (7) we obtain the following equation (valid below the electrodes):

$$r^2 u''(r) + r u'(r) - \gamma^2 r^2 \big(u(r) - V_c\big) = 0 \tag{33}$$

The general solution for this equation has the following form

$$u(r) = B_I I_0(\gamma r) + B_K K_0(\gamma r) + V_c \tag{34}$$

where I_0 i K_0 are the modified Bessel functions. The equation for the potential between the electrodes is much simpler:

$$u'(r) + r u''(r) = 0 \tag{35}$$

Its solution, for $\varphi = 2\pi$ can be expressed using the total current I in the following way:

$$u_m(r) = -\frac{I}{2\pi h\sigma_s} \ln\left(\frac{r}{p}\right) \tag{36}$$

where p is a parameter whose value we are yet to determine.

In what follows we will simplify our considerations assuming that $r_0 = 0$ (which is how the inner electrode is usually fabricated) and that $r_3 \to \infty$. The latter condition is very similar to the third approximation used in section 2. Then, using the following properties of the Bessel function:

$$\lim_{\xi\to 0} K_0(\xi) = \lim_{\xi\to\infty} I_0(\xi) = \infty \tag{37}$$

$$\lim_{\xi\to 0} I_0(\xi) = \lim_{\xi\to\infty} K_0(\xi) = 0 \tag{38}$$

we obtain

$$u_1(r) = B_1 I_0(\gamma_1 r) + V_1 \tag{39a}$$

$$u_2(r) = B_2 K_0(\gamma_2 r) + V_2 \tag{39b}$$

where u_1, u_2 are the electrical potentials in the semiconductor below, respectively, the inner and the outer electrode, while V_1, V_2 are the respective electrode potentials. Continuity of the current at r_1 and r_2 provides the following equations:

$$-h\sigma_s B_1\gamma_1 2\pi r_1 I_0'(\gamma_1 r_1) = I = -h\sigma_s B_2\gamma_2 2\pi r_2 K_0'(\gamma_2 r_2) \tag{40}$$

Using the following properties of the modified Bessel function [5]:

$$I_n' = \frac{1}{2}(I_{n-1} + I_{n+1}) \tag{41a}$$

$$K_n' = -\frac{1}{2}(K_{n-1} + K_{n+1}) \tag{41b}$$

$$I_{-1} = I_1 \tag{41c}$$

$$K_{-1} = K_1 \tag{41d}$$

we obtain the following formula:

$$-\sigma_s B_1\gamma_1 2\pi r_1 I_1(\gamma_1 r_1) = I = \sigma_s B_2\gamma_2 2\pi r_2 K_1(\gamma_2 r_2) \tag{42}$$

The resistance between the electrodes $(V_1 - V_2)/I$ can be expressed, using Eqs. (36), (39) and (40) in the following way:

$$R(r_1, r_2) = \frac{1}{2\pi h\sigma_s}\left(\ln\left(\frac{r_2}{r_1}\right) + \frac{1}{r_1\gamma_1}\frac{I_0(r_1\gamma_1)}{I_0'(r_1\gamma_1)} - \frac{1}{r_2\gamma_2}\frac{K_0(r_2\gamma_2)}{K_0'(r_2\gamma_2)}\right) \tag{43}$$

or, using Eq. (41):

$$R(r_1, r_2) = \frac{1}{2\pi h\sigma_s}\left(\ln\left(\frac{r_2}{r_1}\right) + \frac{1}{r_1\gamma_1}\frac{I_0(r_1\gamma_1)}{I_1(r_1\gamma_1)} + \frac{1}{r_2\gamma_2}\frac{K_0(r_2\gamma_2)}{K_1(r_2\gamma_2)}\right) \tag{44}$$

Neither of the two relations is very convenient. Let us try to get rid of the modified Bessel functions using an approximation which is valid if $|x|$ is sufficiently large [6]:

$$I_0(x) \approx \frac{e^x}{\sqrt{2\pi x}} \tag{45}$$

$$K_0(x) \approx \sqrt{\frac{\pi}{2\pi x}} e^{-x} \tag{46}$$

It allows us to find a convenient estimation of the Bessel function to their derivatives:

$$\frac{I_0(x)}{I_0'(x)} \approx \frac{2x}{2x-1} \tag{47}$$

$$\frac{K_0(x)}{K_0'(x)} \approx -\frac{2x}{2x+1} \tag{48}$$

In the literature, often the above two ratios is approximated simply by 1 [7, 8]. Below we present three formulas for the resistance, starting from the most accurate and ending on the least accurate of them:

$$R(r_1, r_2) = \frac{1}{2\pi h\sigma_s}\left(\ln\left(\frac{r_2}{r_1}\right) + \frac{1}{r_1\gamma_1}\frac{I_0(r_1\gamma_1)}{I_1(r_1\gamma_1)} + \frac{1}{r_2\gamma_2}\frac{K_0(r_2\gamma_2)}{K_1(r_2\gamma_2)}\right) \tag{49}$$

$$\approx \frac{1}{2\pi h\sigma_s}\left(\ln\left(\frac{r_2}{r_1}\right) + \frac{2}{2r_1\gamma_1 - 1} + \frac{2}{2r_2\gamma_2 + 1}\right) \tag{50}$$

$$\approx \frac{1}{2\pi h\sigma_s}\left(\ln\left(\frac{r_2}{r_1}\right) + \frac{1}{r_1\gamma_1} + \frac{1}{r_2\gamma_2}\right) \tag{51}$$

Unlike in the linear TLM case, in the above formulas the contributions from the individual electrodes (described by γ_1 and γ_2) is not the same. If, hypothetically, the electrode resistivity depended on the direction of the current flow (but otherwise were perfectly linear), a measurement of the IV curve for a cTLM structure, would show a difference in slope between the negative and positive current side. If this in not the case, we can say that $\gamma_1 = \gamma_2 = \gamma$, and the above formulas can be written in the following form:

$$R(r_1, r_2) = \frac{1}{2\pi h\sigma_s}\left(\ln\left(\frac{r_2}{r_1}\right) + \frac{1}{r_1\gamma}\frac{I_0(r_1\gamma)}{I_1(r_1\gamma)} + \frac{1}{r_2\gamma}\frac{K_0(r_2\gamma)}{K_1(r_2\gamma)}\right) \tag{52}$$

$$\approx \frac{1}{2\pi h\sigma_s}\left(\ln\left(\frac{r_2}{r_1}\right) + \frac{2}{2r_1\gamma - 1} + \frac{2}{2r_2\gamma + 1}\right) \tag{53}$$

$$\approx \frac{1}{2\pi h\sigma_s}\left(\ln\left(\frac{r_2}{r_1}\right) + \frac{1}{\gamma}\left(\frac{1}{r_1} + \frac{1}{r_2}\right)\right) \tag{54}$$

3.1. Validity of the Bessel functions approximations

Let us analyze when the approximations considered above are valid. The exact solution Eq. (49) (however assuming that the outer ring is semi-infinite) contains terms in the form $B(x)/B'(x)$ where B is a modified Bessel function: either I_0 or K_0. In the approximation (50) and (51) we used the following sequence of simplifications:

$$\frac{I_0(x)}{I_0'(x)} = \frac{I_0(x)}{I_1(x)} \approx \frac{2x}{2x-1} \approx 1 \tag{55}$$

$$-\frac{K_0(x)}{K_0'(x)} = \frac{K_0(x)}{K_1(x)} \approx \frac{2x}{2x+1} \approx 1 \tag{56}$$

The accuracy for both approximations improves with $|x| \to \infty$. However, we do not know yet when x is sufficiently large. Now, we will try to answer this question. Let us denote:

$$Q_I(x) = \frac{I_0(x)}{I_1(x)} \qquad q_I(x) = \frac{2x}{2x-1} \qquad \tilde{q}(x) = 1$$
$$Q_K(x) = \frac{K_0(x)}{K_1(x)} \qquad q_K(x) = \frac{2x}{2x+1}$$
$$r_I(x) = 1 - \frac{q_I(x)}{Q_I(x)} \qquad \tilde{r}_I(x) = 1 - \frac{\tilde{q}(x)}{Q_I(x)}$$
$$r_K(x) = 1 - \frac{q_K(x)}{Q_K(x)} \qquad \tilde{r}_K(x) = 1 - \frac{\tilde{q}(x)}{Q_K(x)}$$

Functions r and $\tilde{r}$ are simply the relative error of approximation q and $\tilde{q}$ respectively. Figure 3 presents a comparison of the approximations used in this paper. If we assume that an error below 1% is acceptable, we can see that approximation q, and hence formula (50) is valid if $\gamma r_1 > 4.5$, while formula (51) should not be used be used unless $\gamma r_1 > 50$. Since formula (50) is only slightly more complicated than formula (51), there is probably no reason to use the latter at all for fitting to experimental data. However, it seems that Eq. (50) does not appear in the literature and only Eqs. (49) and (51) are reported.

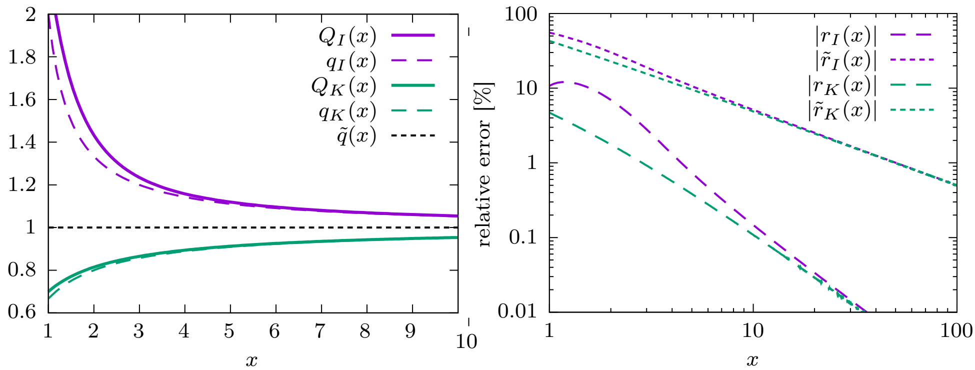


Figure 3. Left: Comparison of the considered approximations of the ratios of the Bessel functions; Right: The relative errors of the approximations
Source: own work.

The above estimations are important not only to validate if the formula used for determination of parameters γ and σ_s was correct. If Eq. (50) is to be used, then one has to accept that for the radii of the outer rings greater than $\sim 1.4 r_0$ the combined resistance of both electrodes is less than 10% of the total resistance. This greatly (and negatively) impacts the accuracy of determination of the contact resistance in such an experiment.
Of course, whether or not the contact resistance can be reliably determined in such an experiment depends on the system geometry and the material parameters. If $\gamma r_0 > 50$, the cTLM configuration itself is not ideal to determine the contact resistance. On the other hand, when the material conductivity is high, that it is hard to measure in the linear TLM configuration, cTLM can be a suitable method. Whatever the situation, since Eq. (51) compared with Eq. (50) does not seem to provide any significant simplification, there is probably no reason to use Eq. (51).

4. SUMMARY

No we will summarize the paper presenting the procedure of experimental data analysis. We assume that a set of I-V curves have been measured—one curve for each pair of electrodes, and provided that they are linear, for each of them the resistance was determined. If the I-V curves are not linear, then the electrode deposition was faulty, and the TLM theory presented here cannot be applied. If we want to determine σ_s we also need to know the thickness of the semiconductor layer.

4.1. Linear TLM

In the case of linear TLM, the experimental data give us a relation $R(L)$, where R is the resistance between electrodes separated by a distance L. A linear function $f(L) = \alpha L + R_0$ is fitted to the experimental data. Then, the condition given in Eq. (22) should be verified. If this condition is fulfilled, formulas (19) can be used to determine the investigated material parameters. If not, formulas (29) should be used, however in order to find ϱ_c a numerical root-finding will be necessary.

4.2. Circular TLM

When cTLM is used, we measure a function $R(r_1, r_2)$. Often, r_1 is identical in all the measured electrode pairs, and the we can assume that we have a function $R(r_2)$. Then we fit relation (53) to the experimental data, obtaining the values for σ_s and γ. If $r_1\gamma > 4.5$, formula (53) can be considered valid. If not, we have to perform least-square fitting again, using this time Eq. (52). Using Eq. (54) should be avoided.

REFERENCES

[1] Berger H., *Contact resistance on diffused resistors*, 'IEEE International Solid-State Circuits Conference. Digest of Technical Papers' 1969, vol. XII, pp. 160–161.

[2] Eidelloth S., Brendel R., *Analytical theory for extracting specific contact resistances of thick samples from the transmission line method*, 'IEEE Electron Device Letters' 2014, vol. 35, no. 1, pp. 9–11.

[3] Reeves G.K., Harrison H.B., *Obtaining the specific contact resistance from transmission line model measurements*, 'IEEE Electron Device Letters' 1982, vol. 3, no. 5, pp. 111–113.

[4] Liu T., Huang R., Li F., Huang Z., Zhang J., Liu J., Zhang L., Zhang S., Dingsun A., Yang H., *Study on the measurement accuracy of circular transmission line model for low-resistance ohmic contacts on III-V wide band-gap semiconductors*, 'Current Applied Physics' 2018, vol. 18, no. 7, pp. 853–858.

[5] Kącki E., Siewierski L., *Wybrane działy matematyki wyższej z ć wiczeniami,* Wydawnictwo PWN, Warszawa 1975, pp. 133–134.

[6] Temme N.M., *A double integral containing the modified Bessel function: Asymptotics and computation*, 'Mathematics of Computation' 1986, vol. 47, no. 176.

[7] Ho J-K., Jong C-S., Chiu C.C., Huang C-N., Shih K-K., Chen L-C., Chen F-R., Kai J-J., *Low-resistance ohmic contacts to p-type GaN achieved by the oxidation of Ni/Au films*, 'Journal of Applied Physics' 1999, vol. 86, no. 8, pp. 4491–4497.

[8] Klootwijk J.H., Timmering C.E., *Merits and limitations of circular TLM structures for contact resistance determination for novel III-V HBTs* [in:] Hagiwara Y. (ed.), *Proceedings of the 2004 International Conference on Microelectronic Test Structures, Japan 2004, pp. 247–252.*