# The Smith Chart

Aleksandar Marinčić
Faculty of Electrical Engineering, University of Belgrade

## Historical background

Electrical transmission can be traced back to 1729 when Stephen Gray discovered that the electrostatic phenomenon of attraction of small bits of mater could occur at one end of a damp string several hundreds feet long when the other end is touched with an electrostatically charged body [1]. This was disclosed sixty years after Otto von Guericke had noted that short treads connected to a primitive electrostatic machine become charged throughout their length, but it was not associated with electrical fluid transmission until Gray's time. Gray also discovered that his damp string should be supported by dry silk treads but not by fine brass wires. In 1735 Charles DuFay was the first to make distinction between electrical conductors and electrical insulators and reported the existence of two different kinds of electricity, labelled later in 1747 as positive and negative by Benjamin Franklin. On the basis of Gray's discovery between 1770 and 1830 several electrostatic telegraph systems were constructed in various parts of the world, which could transmit signals over distances up to a few miles.

Following Volta's discovery of the chemical pile in 1800 and Oersted's discovery of the magnetic effect of a current in 1820 resulted in magnetic telegraphs of which practical success achieved Wheatstone and Cook in 1839 and Morse in 1844. Because of problems with insulation, buried transmission lines were quickly replaced by open wire lines on poles. However, soon a need to span rivers, seas and ocean urged again development of buried cable transmission. This required further development of transmission line theory and the promoters of transatlantic submarine cables turned to William Thomson (later Sir William and ultimately Lord Kelvin)for advice. In 1855 Thomson published the first distributed circuit analysis of a uniform transmission line. Based on his specifications in 1858 an underwater cable was made and laid across the Atlantic. It carried messages for few weeks before insulation failed.

The importance of cable development and technological problems associated with its design resulted in the development of a new class of technical personnel which formed the world's first professional association of electrical engineers. From these emerged the American Institute of Electrical Engineers in the United States in 1884 and the Institution of Electrical Engineers in England five years later.

Further important mathematical development of signal transmission calculation over transmission lines was made by Oliver Heaviside, beginning in 1880. Heaviside noted that on most practical lines, voice signals should travel with reduced loss if the distributed inductance of the line could be increased without adversely changing the other distributed circuit coefficients. Heaviside's work relayed strongly on Maxwell's theory and he wrote a series of 47 papers between 1885 to 1887, all under the tittle „Electromagnetic Induction and its Propagation" which is the foundation of modern transmission-line theory [2]. The problem of telephone signal transmission was basically solved by Michael Pupin who developed mathematical theory of loaded lines with periodically inserted inductance coil. He also made artificial transmission lines and experimentally proved his theory and technique known as „pupinization". This technique was the only one suitable for the long distance signal transmission until the development of vacuum tube amplifiers around 1915. The first long distance carrier-frequency telephone systems, in which several voice frequency channels were transmitted over a single wire pair was made feasible in 1919. This technique, known as the „frequency division multiplexing" was subsequently used for the transmission of many telephone channels over one coaxial cable, reaching few years ago the limit of around 10000 analogue telephone channels.

Coaxial cable was another wire structure which was originally investigated by Hertz in view of its shielding property and ability to propagate high frequency electromagnetic waves. He even experimented with hollow metallic tube transmission but his early death in 1894 prevented him from completing this research. It was Rayleigh who in 1897 showed mathematically that Hertz's ideas of guiding electromagnetic waves by proper wire structures were correct. He did this by solving the boundary value problem for Maxwell's equations for metallic tubes of circular and rectangular cross-sections. He showed that such waves could exist in a set of normal modes that must have a longitudinal electric or magnetic field component, and that „propagation starts at some high frequency set by the diameter of a circular pipe, or width of a rectangular pipe, and the mode number" [2].

# Theoretical background of the Smith chart

The differential equations for a uniform transmission line are well-known Telegraph's equations of the form

$$\frac{\partial v(z,t)}{\partial z} = -Ri(z,t) - L\frac{\partial i(z,t)}{\partial t} \qquad (1)$$

$$\frac{\partial i(z,t)}{\partial z} = -Gv(z,t) - C\frac{\partial v(z,t)}{\partial t} \qquad (2)$$

A complete solution of equations (1) and (2) would find expressions for $v$ and $i$ as functions of $z$ and $t$, subject to boundary conditions determined by the nature of the source generator at $z=0$ and the device connected to the end of line $z=l$, where $l$ is the length of the line.

The first step to find solution of the above simultaneous equations is to eliminate one of the variables. By taking the first derivative of equations (1) and (2) with respect to $z$, after simple calculations one obtains the following equations:

$$\frac{\partial^2 v(z,t)}{\partial z^2} = LC\frac{\partial^2 v(z,t)}{\partial t^2} + (LG+RC)\frac{\partial v(z,t)}{\partial t} + RGv(z,t) \qquad (3)$$

$$\frac{\partial^2 i(z,t)}{\partial z^2} = LC\frac{\partial^2 i(z,t)}{\partial t^2} + (LG+RC)\frac{\partial i(z,t)}{\partial t} + RGi(z,t) \qquad (4)$$

Although both $v(z,t)$ and $i(z,t)$ obey the same differential equation, their solutions are different because of the boundary conditions which are not the same for the two variables.

Even when $R, L, C$ and $G$ are postulated to be constants for all values of current and voltage and their derivatives, equations (3) and (4) are second order linear partial differential equations in the time coordinate and one space coordinate. Although similar to „standard" partial differential equations, they have some additional terms which do not allow us to write any simple complete general solution for $v(z,t)$ and $i(z,t)$.

In special case when a sinusoidal time dependence of voltage and current are assumed, i.e.

$$v(z,t) = \mathrm{Re}\{V(z)\exp(j\omega t)\} \qquad (5)$$

$$i(z,t) = \mathrm{Re}\{I(z)\exp(j\omega t)\} \qquad (6)$$

equations (3) and (4) become

$$\frac{\partial^2 V(z)}{\partial z^2} - [-\omega^2 LC + j\omega(LG+RC) + RG]V(z) = 0 \qquad (7)$$

$$\frac{\partial^2 I(z)}{\partial z^2} - [-\omega^2 LC + j\omega(LG+RC) + RG]I(z) = 0 \qquad (8)$$

These equations are simple to solve and their solutions in terms of $V$ and $I$ as phasor functions of $z$ can be written in the form

$$V(z) = V_1\,\mathrm{e}^{-\gamma z} + V_2\,\mathrm{e}^{+\gamma z} \qquad (9)$$

$$I(z) = I_1\,\mathrm{e}^{-\gamma z} + I_2\,\mathrm{e}^{+\gamma z} \qquad (10)$$

where

$$\gamma = \sqrt{(R+j\omega L)(G+j\omega C)} = \alpha + j\beta \qquad (11)$$

or

$$\alpha = \left\{\tfrac{1}{2}\left[\sqrt{(R^2+\omega^2 L^2)(G^2+\omega^2 C^2)} - \omega^2 LC + RG\right]\right\}^{1/2} \qquad (12)$$

$$\beta = \left\{\tfrac{1}{2}\left[\sqrt{(R^2+\omega^2 L^2)(G^2+\omega^2 C^2)} + \omega^2 LC - RG\right]\right\}^{1/2} \qquad (13)$$

In equations (9) and (10) $V_1, V_2, I_1, I_2$ are also phasors and are the sets of two arbitrary coefficients that are related through the boundary conditions. Returning now to equation (5) we can write

$$v(z,t) = |V_{1p}|\mathrm{e}^{-\alpha z}\,\mathrm{Re}\{\mathrm{e}^{j(\omega t - \beta z + \phi_1)}\} + |V_{2p}|\mathrm{e}^{-\alpha z}\,\mathrm{Re}\{\mathrm{e}^{j(\omega t - \beta z + \phi_2)}\} \qquad (14)$$

where $|V_{1p}|$ and $|V_{2p}|$ are the *peak* amplitudes of the arbitrary phasor coefficients $V_1$ and $V_2$ which are assumed to stand for *rms.* values. In a similar way we can derive expression for $i(z,t)$. In equation (11) $\alpha$ is the attenuation factor and $\beta$ is the phase factor. Both factors are function of frequency.

The first term on the right hand side of equation (14) is the direct wave, i.e. the wave travelling in the $+z$ direction, while the second term is the reflected wave travelling in the opposite direction. Both waves are harmonic waves. At every coordinate $z$ on the line the voltage varies harmonically with time, with constant amplitudes.

General solution when an arbitrary time dependence of the generator is assumed cannot be obtained in a closed form similar to that shown in equation (14). However, in many cases of practical interest, steady state sinusoidal time solution given by equation (14) is of great value. While it cannot give answers to a general telephone signal transmission it can be of direct use in radio technique where a modulated carrier transmission is studied.

The physical meaning of equations (9) and (10) can be explained in terms of the direct and reflected voltage and current waves. The terms with $e^{-\gamma z}$ represent direct waves travelling in the $+z$ direction, while the terms with $e^{+\gamma z}$ represent reflected waves travelling in the $-z$ direction.

For a line with no reflected wave $V_2 = I_2 = 0$, the ratio of direct voltage and direct current is given by

$$\frac{V}{I} = \frac{V_1}{I_1} = \sqrt{\frac{R+j\omega L}{G+j\omega C}} = Z_0 = R_0 + jX_0 \qquad (15)$$

and is independent of the position on the line. This ratio is „characteristic" of the line itself and is appropriately named „characteristic impedance" of the line. In general $Z_0$ is a complex number and its components are given by

$$R_0 = \frac{1}{\sqrt{G^2+\omega^2 C^2}} \left\{ \tfrac{1}{2} \left[ \sqrt{(R^2+\omega^2 L^2)(G^2+\omega^2 C^2)} + \omega^2 LC + RG \right] \right\}^{1/2} \qquad (16)$$

$$X_0 = \frac{\pm 1}{\sqrt{G^2+\omega^2 C^2}} \left\{ \tfrac{1}{2} \left[ \sqrt{(R^2+\omega^2 L^2)(G^2+\omega^2 C^2)} - \omega^2 LC - RG \right] \right\}^{1/2} \qquad (17)$$

The sign in equation (17) is positive if $\omega L / R > \omega C / G$, which is true for most practical cases.

For high frequency operation of transmission lines if $\omega L / R$ and $\omega C / G$ are sufficiently large compared to unity, equations (12,13,16,17) give

$$\alpha_{hf} = \tfrac{1}{2} R / Z_0 + \tfrac{1}{2} G Z_0 \qquad (18a)$$

$$\beta_{hf} = \omega \sqrt{LC} \qquad (18b)$$

$$Z_{0hf} = R_{0hf} = \sqrt{L/C} \qquad (19a)$$

$$X_{0hf} = 0 \qquad (19b)$$

The minimum frequency that will qualify as a „high" frequency depends on the actual values of $R, L, C, G$. For a high accuracy results it is also important to remember that the distributed circuit coefficients may be functions of frequency, particularly $R$ and $G$ increase steadily that the values of ratios $\omega L / R$ and $\omega C / G$ are large compared to unity.

## Impedance relations

When a uniform transmission line is not terminated in its characteristic impedance $Z_0$, but with any arbitrary impedance $Z_L$, there are always reflected waves on the line, and the impedance at any point of the line differs from $Z_0$. The input impedance at a distance $z$ measured from the generator is given by

$$Z_Z = \frac{V}{I} = \frac{V_1 e^{-\gamma z} + V_2 e^{\gamma z}}{I_1 e^{-\gamma z} + I_2 e^{\gamma z}} = Z_0 \frac{V_1 e^{-\gamma z} + V_2 e^{\gamma z}}{V_1 e^{-\gamma z} - V_2 e^{\gamma z}} \qquad (20)$$

If we put $z = l$, where $l$ is the line length, the impedance $Z_Z$ is equal to the load impedance $Z_L$ and we have

$$Z_L = Z_0 \frac{V_1 e^{-\gamma l} + V_2 e^{\gamma l}}{V_1 e^{-\gamma l} - V_2 e^{\gamma l}} = Z_0 \frac{1+\Gamma_L}{1-\Gamma_L} \qquad (21)$$

where we introduced a phasor, called the reflection coefficient, representing the ratio of phasor voltages at the load

$$\Gamma_L = \frac{V_2}{V_1} e^{2\gamma l} = \frac{Z_L - Z_0}{Z_L + Z_0} = |\Gamma_L| e^{j\phi_L} \qquad (22)$$

By substituting equation (22) into (20), after simple calculations one obtains the input impedance of a line of length $l$ by putting $z = 0$ in the form:

$$Z_{z=0} = Z_{in} = Z_0 \frac{1+\Gamma_L e^{-2\gamma l}}{1-\Gamma_L e^{-2\gamma l}} \qquad (23a)$$

or

$$\frac{Z_{in}}{Z_0} = \frac{Z_L / Z_0 + \tanh(\gamma l)}{1 + (Z_L / Z_0)\tanh(\gamma l)} = \tanh(\gamma l) + \tanh^{-1}(Z_L / Z_0) \qquad (23b)$$

from which we see that for the terminating impedance equal to the characteristic impedance, the input impedance of any line length is equal to the characteristic impedance.

On a line terminated in an arbitrary terminating impedance, voltage and current waves vary with distance in a form of standing wave pattern. For a lossless line, the ratio of maximum to minimum voltage is constant, for a lossy line it decreases towards the generator. In various measurements we often use a short line that can be considered lossless. In that case $\gamma = j\beta$, and from equation (9) we can write

$$V(z) = V_1 e^{-j\beta z}\left(1+\frac{V_2}{V_1}e^{j2\beta z}\right) = V_1 e^{-j\beta z}\left(1+|\Gamma_L|e^{j\phi_L}e^{-j2\beta(l-z)}\right) \quad (24)$$

When the exponential term in equation (24) is a multiple of $2\pi$, the magnitude of $V(z)$ is a maximum, while when it is an odd multiple of $\pi$, it is a minimum. The ratio of the two magnitudes is the voltage standing wave ratio

$$\mathrm{VSWR} = \frac{1+|\Gamma_L|}{1-|\Gamma_L|} \quad (25a)$$

at any point on the line, and conversely

$$|\Gamma_L| = \frac{\mathrm{VSWR}-1}{\mathrm{VSWR}+1} \quad (25b)$$

In deriving equation (24) it was assumed that the line length is $l$, so that $(l-z)$ is the distance measured from the load.

## Graphical representation of transmission line quantities

Arithmetical calculation of complex exponential and hyperbolic functions appearing in the transmission line equations are time-consuming and the use of mathematical tables is less effective than for the corresponding functions of real variables. This is the main reason for long use of graphical aids in transmission line calculations.

A *Chart Atlas of Complex Hyperbolic Functions,* published by A.E.Kennelly in 1914 was widely used for decades for the solution of equation (23). The charts presented a loci of the real and imaginary parts of the complex hyperbolic tangent and other functions over the complex variable or neper-radian plane [1]. Good significant-figure precision was achieved by the large size graphs of about 50 $\times$ 50 cm$^2$, and used on systems as cable pairs and open-wire lines at voice frequencies and low carrier frequencies. For high frequency systems that appeared some decades later, these charts were cumbersome and another approach finally evolved.

As explained earlier, at high frequencies the characteristic impedance of the line can be assumed to be real with a high accuracy. The complex impedance $Z_Z$ is normalised with respect to the characteristic impedance and is given by

$$\mathbf{z}_Z = \frac{Z_Z}{Z_0} = r_Z + jx_Z = \frac{1+\Gamma_Z}{1-\Gamma_Z} \quad (26)$$

where index $z$ refers to the distance at which the input impedance of the line is $Z_Z$, and the complex reflection coefficient is $\Gamma_Z$. For simplicity we will further omit the index $z$ and write

$$\mathbf{z} = \frac{Z}{Z_0} = r + jx = \frac{1+\Gamma}{1-\Gamma} = |\mathbf{z}|e^{j\theta} \quad (27)$$

Equation (27) is the starting equation for the hemisphere charts that appeared in 1930's. The first was made by Smith at Bell Telephone Laboratories in 1931 but not published until January 1939 [4]. It was a hemisphere chart with $r$ and $x$ circles, radial scales. At the same time Carter in RCA Review published the hemisphere chart with $|\mathbf{z}|$ and $\theta$ circles [5]. In 1944 Smith published another paper about an improved transmission line calculator [6]. The present day use of the Smith chart is in many respect founded on the latter paper which will be reviewed in the following part of our paper.

In Smith's words „the calculator is, fundamentally a special kind of impedance coordinate system, mechanically arranged with respect to a set of movable scales to portray the relationship of impedance at any point along a uniform open wire or coaxial transmission line to the impedance at any other point and to the several other parameters" [6]. The parameters which are plotted include:

1. Impedance, or admittance at any point along the line with the corresponding complex reflection coefficient at the same point;
2. Length of the line between any two points in wavelengths;
3. Attenuation between any two points in decibels;
4. Voltage or current standing wave ratio.

The impedance at any point along a transmission line is normally considered to be that impedance which would be measured at the input of a line section connected to the load. For a steady state, the generator impedance, as well as the impedance looking towards the generator from any point at which we observe the above defined line impedance, do not affect the relative distribution of current or voltage along the transmission line. The generator impedance can affect only the power delivered to the transmission line system, and not the reflection coefficient or transformed load impedance.

The Smith chart is plotted on the voltage reflection coefficient plane which can be considered as plotted on polar coordinates as $\Gamma(|\Gamma|,\phi) = |\Gamma|e^{j\phi}$, or rectangular coordinates of the real and imaginary components as $\Gamma(\Gamma_r,\Gamma_i) = \Gamma_r + j\Gamma_i$. Each component of the reflection coefficient is a function of the normalised impedance at that cross-section, and they are related through the complex equation (27) which we now rewrite in terms of the real and imaginary components of **z** and $\Gamma$:

$$r + jx = \frac{1 - \Gamma_r - j\Gamma_i}{1 + \Gamma_r + j\Gamma_i} \qquad (28)$$

or

$$\Gamma_r + j\Gamma_i = \frac{r + jx - 1}{r + jx + 1} \qquad (29)$$

From equation (28), after separating the real and imaginary parts we obtain

$$\left[\Gamma_r - \frac{r}{1+r}\right]^2 + \Gamma_i^2 = \frac{1}{(1+r)^2} \qquad (30)$$

$$\left[\Gamma_r - 1\right]^2 + \left[\Gamma_i - \frac{1}{x}\right]^2 = \frac{1}{x^2} \qquad (31)$$

On rectangular coordinates (30) is the equation of a circle whose center, for any value of $r$, is located at $\Gamma_r = r/(r+1), \Gamma_i = 0$, and whose radius is $1/(r+1)$. Several of these circles for various values of $r$ are plotted in Fig.1. All the circles pass the point 1, 0.

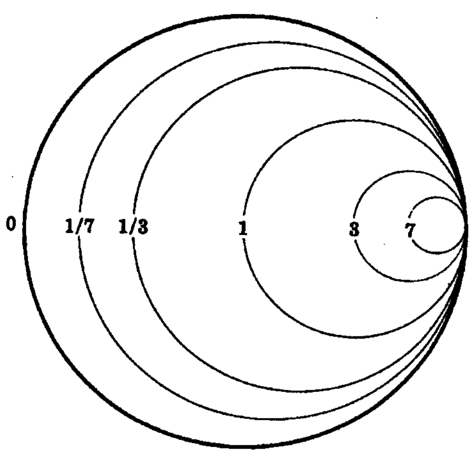


Fig. 1. Coordinate circles for constant normalized resistance on the Smith chart. The radii of the particular circles shown are related by simple fractions.

Equation (31) is another set of circles whose center is located at $\Gamma_r = 1, \Gamma_i = 1/x$, and whose radius is $1/x$. Several of these circles are plotter in Fig.2, this time for positive and negative values of $x$. The most obvious symmetry seen here is the mirror-image symmetry about the horizontal central axis of the chart, for the circles corresponding to the same absolute values of $x$. Only part of the circles inside the central circle of radius one is shown, as the maximum values of $|\Gamma|$, for any passive loaded line, is one. For negative resistance, as for example in the case of some amplifiers, the magnitude of the reflection coefficient $|\Gamma|$ can be greater than unity.

The circles given by equations (30) and (31) are orthogonal circles that make a conform mapping chart. To each $\Gamma_r, \Gamma_i$ there is a single corresponding $r, x$ pair, and vice versa.

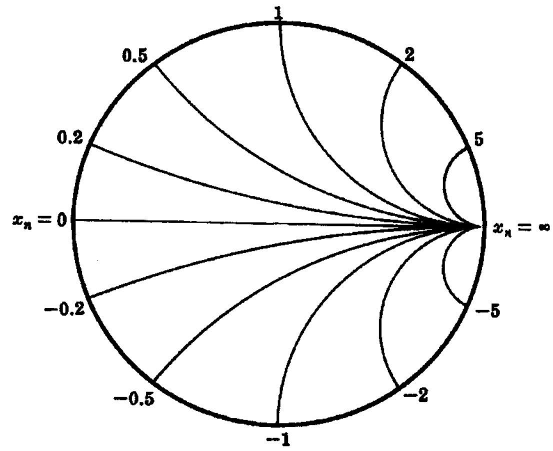


Fig. 2. Coordinate circles for constant normalized reactance on the Smith chart.

Another type of conform mapping chart was made by Carter [5], in which case to each $\Gamma_r, \Gamma_i$ there is a single corresponding $|\mathbf{z}|, \theta$ pair, and vice versa. On the $\Gamma_r, \Gamma_i$ coordinate system, constant $|\mathbf{z}|$ curves, as well as constant $\theta$ curves are circles given by equations:

$$\left[\Gamma_r + \frac{1+|\mathbf{z}|^2}{1-|\mathbf{z}|^2}\right]^2 + \Gamma_i^2 = \frac{4|\mathbf{z}|^2}{\left(1-|\mathbf{z}|^2\right)^2} \qquad (32)$$

$$\Gamma_r^2 + \left[\Gamma_i + \cot\theta\right]^2 = 1/(\sin\theta)^2 \qquad (33)$$

The Carter chart of normalised impedance magnitude and phase angle, for several typical values of $|\mathbf{z}|$ and $\theta$ is plotted in Fig.3. The vertical central diameter of the chart is for $|\mathbf{z}| = 1$. Two impedances of equal phase angle but with reciprocally normalised magnitude lie at mirror-image points relative to the $\Gamma_i$ axis. Carter chart is another form of Smith chart as each pair of $|\mathbf{z}|, \theta$ has corresponding pair $r, x$ related simply as $r = |\mathbf{z}|\cos\theta$, and $x = |\mathbf{z}|\sin\theta$

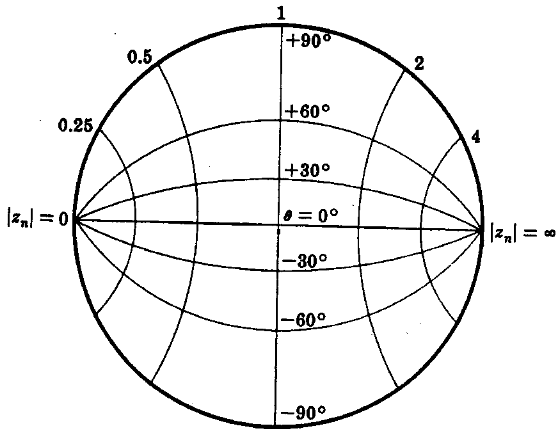


Fig. 3. The Carter chart. Coordinates of normalized impedance magnitude and phase angle on the reflection coefficient plane.

## Concluding remarks

The construction of the detailed Smith chart in its standard published form is shown in Fig.4. This chart radically simplified transmission line analysis, particularly in the ultrahigh frequency field. It played an important role in developing microwave radar systems.

During the Second World War, the inventor of the Smith chart worked on antennas with great success, so that in 1952 he was elected a Fellow of the Institute of Radio Engineers for his contribution to antennas and graphical analysis. In 1969 he published book *Electronic Applications of the Smith Chart in Waveguide, Circuit, and Component Analysis.* After retiring from the Bell Laboratories, Smith founded Analog Instruments, which sold

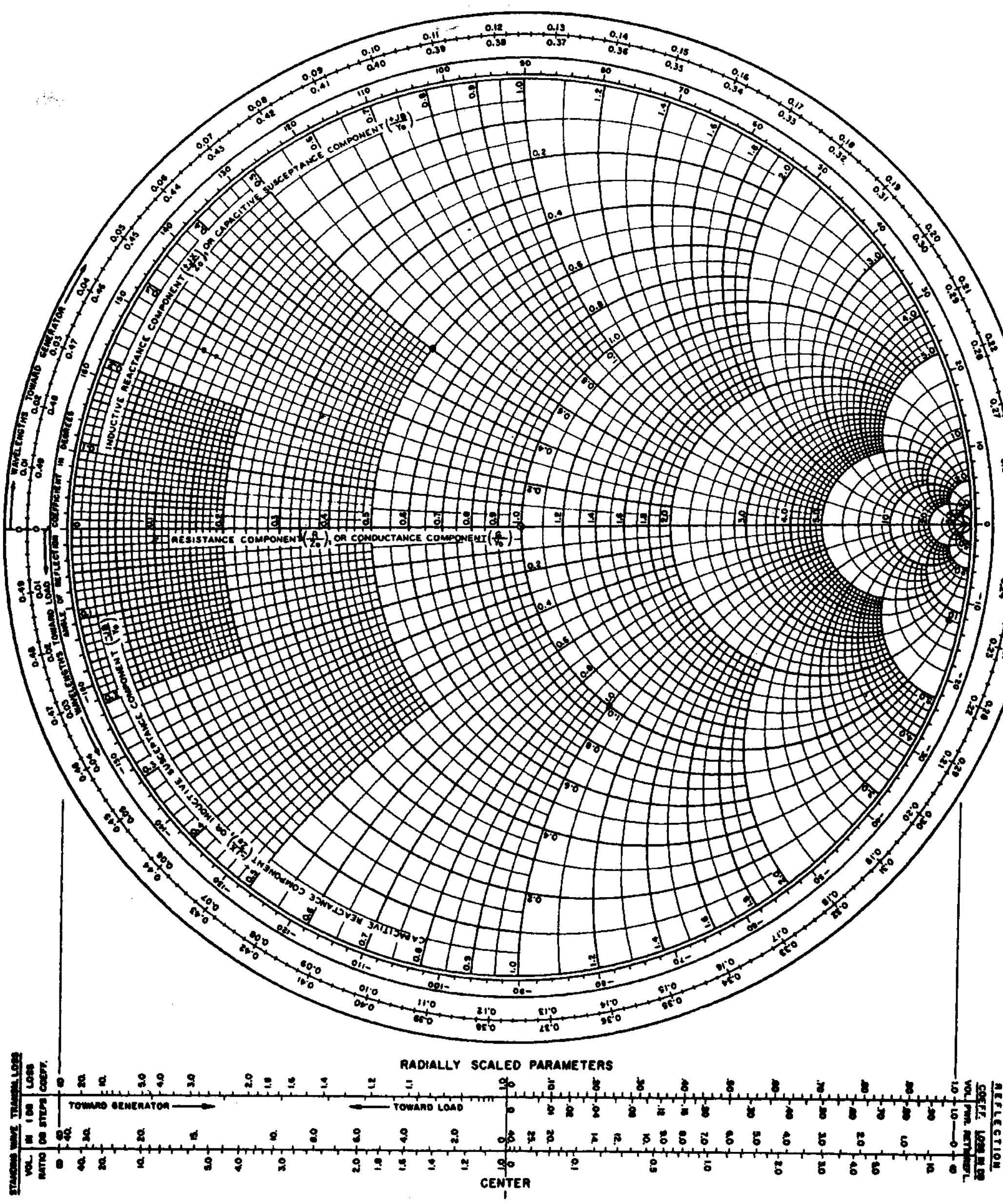


Fig. 4. A standard commercially available form of Smith chart graph paper. Copyrighted 1949 by Kay Electric Company, Pine Brook, N. J.

navigational instruments for the light aircraft and later added a line of Smith chart and related items. At the time he died in August 1987, it was reported that more than a nine million copies of the chart had been sold [7]. The Smith chart appears in at least 12 different types of the trademarked type, including a „negative Smith chart" for the analysis of negative resistance devices. Today, the most modern, computer based automatic network analysers rely on the Smith chart for data display, and the Smith chart use is common in most modern textbook and courses in electrical engineering.

The Smith chart is no doubt a unique diagram which have been widely used for nearly seventy years and is still one of the most common every-day mean for presenting data. It has many advantages due to its simplicity to interpret and give an overall, quick insight over complex behaviour of many microwave circuit components, both passive and active. We believe that it will be in use for many years to come not only as a pedagogically perfect analogue data display, but also as an aid to professionals in obtaining quick answers to many transmission line problems which they meet.

## Literature

[1] Chipman, R.A.: „Theory and Problems of Transmission Lines", Schaum's outline series, McGraw-Hill Book Company, p.3-8.

[2] Bryant, J.H.: „Coaxial Transmission Lines, Related Two-Conductor Transmission Lines, Connectors, and Components: A U.S. Historical Perspective, IEEE Trans. Micr. Theory and Techniques, Vol.MTT-32, No.9, Sept.1984, pp.970-983.

[3] Wheeler, H.A.: „Reflection Charts Relating to Impedance Matching", IEEE Trans. Micr. Theory and Techniques, Vol.MTT-32, No.9, Sept.1984, pp.1008-1021.

[4] Smith, P.H.: „Transmission Line Calculator", Electronics, vol.12, no.1, pp.29-31, January 1939.

[5] Carter, P.S.: „Charts for transmission-line measurements and computations", RCA Rev., vol.3,pp.355-368, January 1939.

[6] Smith, P.H.: „An improved transmission line calculator", Electronics, vol.17, no.1, pp.130-133, 318-325, January 1944.

[7] Brittain, J.E.: „The Smith Chart", IEEE Spectrum, August 1992, pp.65.