DATA TRANSMISSION 2A

T. Bruins

Lectures given in the
Technical Training Programme of CERN 1970-1971

GENEVA

1971
Propriété littéraire et scientifique réservée pour tous les pays du monde. Ce document ne peut être reproduit ou traduit en tout ou en partie sans l'autorisation écrite du Directeur général du CERN, titulaire du droit d'auteur. Dans les cas appropriés, et s'il s'agit d'utiliser le document à des fins non commerciales, cette autorisation sera volontiers accordée.

Le CERN ne revendique pas la propriété des inventions brevétables et dessins ou modèles susceptibles de dépôt qui pourraient être décrits dans le présent document; ceux-ci peuvent être librement utilisés par les instituts de recherche, les industriels et autres intéressés. Cependant, le CERN se réserve le droit de s'opposer à toute revendication qu'un usager pourrait faire de la propriété scientifique ou industrielle de toute invention et tout dessin ou modèle décrits dans le présent document.

© Copyright CERN, Genève, 1971

Literary and scientific copyrights reserved in all countries of the world. This report, or any part of it, may not be reprinted or translated without written permission of the copyright holder, the Director-General of CERN. However, permission will be freely granted for appropriate non-commercial use. If any patentable invention or registrable design is described in the report, CERN makes no claim to property rights in it but offers it for the free use of research institutions, manufacturers and others. CERN, however, may oppose any attempt by a user to claim any proprietary or patent rights in such inventions or designs as may be described in the present document.
ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

DATA TRANSMISSION 2A

T. Bruins

Lectures given in the
Technical Training Programme of CERN 1970-1971

GENEVA
1971
FOREWORD

Data transmission at speeds such that both the length and the quality of the line make little effect is relatively simple, and analyses of such conditions can be made with very simple d.c.-type calculations. The advent of computer networks with computer-terminal or computer-computer transmissions over a wide speed range, invoke techniques which can no longer be considered in the previous category, both for economic reasons and for reasons of distances involved. One obvious case which comes to mind is the use of lines, originally installed as simple voice or low-speed telegraph lines, as high-speed data transmission media. This course presents some selected techniques, which permit the appreciation of problems involved when a line is used other than in the simple "d.c. condition". A complete coverage of the field in such a short work is, of course, impossible, and no apology is made for having neglected certain techniques. Rather, the course is presented as a selection of associated topics and as a stimulus for more advanced reading.

Before we are able to see how a time-dependent signal reacts in a transmission line, a short introduction to network analyses will be necessary.

A certain familiarity with mathematical analyses and electronic network theory has been assumed.
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FOREWORD</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>INTRODUCTION TO NETWORK ANALYSIS</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>SOME PROPERTIES OF A TRANSMISSION LINE</td>
<td>5</td>
</tr>
<tr>
<td>III</td>
<td>INTRODUCTION TO FOURIER ANALYSES</td>
<td>11</td>
</tr>
<tr>
<td>IV</td>
<td>MODULATION</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. INTRODUCTION</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>1.1 Amplitude modulation (AM)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.1.1 Frequency modulation</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>1.1.2 Phase modulation</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>2. DIGITAL MODULATION TECHNIQUES</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>2.1 Pulse duration modulation (PDM)</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>2.2 Pulse position modulation (PPM)</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>2.3 Delay modulation [delay shift key (DSK)]</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>2.4 Pulse code modulation (PCM)</td>
<td>24</td>
</tr>
<tr>
<td>V</td>
<td>TRANSMISSION TECHNIQUES</td>
<td>25</td>
</tr>
<tr>
<td>VI</td>
<td>ERROR DETECTION AND ERROR CORRECTION</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>BIBLIOGRAPHY</td>
<td>37</td>
</tr>
</tbody>
</table>
A combination of resistances, capacitances, and inductances is generally called a passive network. We may define two different intersection points of a network as input points, and two other intersection points as output points. A voltage signal $V_i(t)$ on the input points will generally give rise to a voltage signal on the output points, $V_{out}(t)$. The fraction $V_{out}(t)/V_{in}(t) = F(t)$ is called the transfer function and is to be found with techniques based on Kirchhoff laws. For details, one is referred to more specific books on the subject. In the following example, capacitances and inductances are written as operators. So the capacitive impedance of a capacitor of C farads is defined by the relationship between charge and voltage:

$$Q(t) = C \cdot V(t)$$

$$\frac{dQ(t)}{dt} = i(t) = C \frac{d}{dt} V(t)$$

$$Z_{cap} = \frac{1}{C} \frac{d}{dt} = \frac{1}{C_p} \left( p = \frac{d}{dt} \right) .$$

The inductive impedance of a coil of L henrys is found from the relationship between flux and current:

$$\phi(t) = L \cdot i(t)$$

$$\frac{d\phi(t)}{dt} = V(t) = L \frac{d}{dt} i(t)$$

$$Z_{ind} = L \frac{d}{dt} = L \cdot p \left( p = \frac{d}{dt} \right) .$$

Example:

With $a$, $b$, and $c$ linear impedance one gets

$$V_{out} = \frac{bc + ac}{ab + ac + bc} \cdot V_i .$$

If we take time-dependent functions instead, and general non-linear impedances, the transfer function will be an operator and the equation will be a differential equation.

The above example with $a = C$ farads, $b = L$ henrys and $c = R$ ohms will become

$$\left( \frac{1}{C_p} \cdot Lp + \frac{1}{C_p} \cdot R + Lp \cdot R \right) V_{out}(t) = \left( Lp \cdot R + \frac{1}{C_p} \cdot R \right) V_i(t)$$

$$\left( Lp C_p^2 + Lp + R \right) V_{out}(t) = \left( Lp C_p^2 + 1 \right) R \cdot V_i(t) .$$
If we now let $V_i(t) = V_i \cdot e^{st}$, and replace $p$ again by $d/dt$

$$LRC \left( \frac{d}{dt} \right)^2 V_{out}(t) + L \left( \frac{d}{dt} \right) V_{out}(t) + RV_{out}(t) =
LRC \left( \frac{d}{dt} \right)^2 V_i \cdot e^{st} + R \cdot V_i \cdot e^{st} = V_i \cdot e^{st} \cdot (LRC \cdot s^2 + R)$$

The solution of this equation is

$$V_{out}(t) = V_i \cdot e^{st} \cdot \frac{LRCs^2 + R}{LRCs^2 + Ls + R} \equiv V_i(t) \cdot \frac{P(s)}{Q(s)} .$$

The solutions of the homogeneous equation

$$LRC \left( \frac{d}{dt} \right)^2 V_{out}(t) + L \left( \frac{d}{dt} \right) V_{out}(t) + RV_{out} = 0$$

give the so-called eigensolutions of the equation. They represent the theoretical solutions of $V_{out}(t)$, where $V_i(t) = 0$ (right-hand side of the equation "zero"). The solution is found by looking for roots of the corresponding equation $LRCs^2 + Ls + R = 0$.

With $s_1$ and $s_2$ as roots, the basic solution will be of the form $C_1 \cdot e^{s_1 t} + C_2 \cdot e^{s_2 t}$.

More details of the above-mentioned differential equation can be found in any textbook on differential equations.

It should be mentioned that solutions such as $s_1$ and $s_2$ may be complex, but need a negative real part.

Thus we arrive at a solution $e^{st} = e^{(a+ib)t} = e^{at} \cdot e^{ibt}$, $e^{ibt} = e^{at} \cdot (\cos bt + i \sin bt)$.

The physical value has $e^{at} \cos bt$ as its real part. Only damped or stable solutions are acceptable, so $a < 0$. With our solution

$$V_{out}(t) = \frac{P(s)}{Q(s)} \cdot V_i \cdot e^{st} = \frac{P(s)}{Q(s)} \cdot V_i \cdot e^{st} ,$$

we have a powerful equation.

The method of representing a signal as $e^{st}$ is a mathematical trick. For its physical meaning one is interested in the real part of it. So

$$\text{Re} (e^{st}) = \text{Re} (e^{(a+ib)t}) = e^{at} \cos bt .$$

Giving an harmonic movement with angular frequency

$$\omega = 2\pi f = \frac{2\pi}{T} ,$$

the function $F(s)$ is a function of the complex variable $s$. It is the magnitude and not the phase of $F$ which is of interest to us, so we have $|F(s)|$.

Like any real function of a single real variable, it is possible to look at the real function $F(s)$ of a complex variable. It is interesting where $|F(s)| = 0$, and where $|F(s)| = \infty$:

$$|F(s)| = 0 \text{ when } P(s) = 0 .$$
This equation gives the zero points of the network, \( s_{x_1}, s_{x_2}, \ldots, s_{x_n} \). \(|F(s)| = \infty\) is found when \( Q(s) = 0 \) [and \( |P(s)| \neq 0 \)]

The solutions of \( Q(s) = 0 \) were called the eigensolutions or poles of the network, normally written as \( s_{x_1}, s_{x_2}, \ldots, s_{x_n} \). \( Q(s) \) is called the Kirchhoff discriminant.

Example:

\[
\frac{V_{i_2}(t)}{V_i(t)} = \frac{1}{\frac{1}{Cs} + \frac{1}{Ls + R + \frac{1}{Cs}}} = \frac{1}{LCRs^2 + CLs + 1}.
\]

Zero points: \( s_0 = \infty \).

Poles: \( RLCs^2 + CLs + 1 = 0 \)

\[
s_{x_{1,2}} = \frac{-CR \pm \sqrt{(CR)^2 - 4LC}}{2LC}
\]

\[
s_{x_1} = \frac{-R}{2L} + \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC}}
\]

\[
s_{x_2} = \frac{-R}{2L} - \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC}}
\]

with

\[
\left(\frac{R}{2L}\right)^2 - \frac{1}{LC} < 0\quad \frac{L}{R^2C} > \frac{1}{4}
\]

\[
s_{x_1} = -\frac{R}{2L} + i \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}
\]

and

\[
s_{x_2} = -\frac{R}{2L} - i \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}
\]

So an input signal \( V_1(t) = e^{s_1x^t} \) gives \(|F(s)| = \infty\).

If we take an harmonic input signal \( \cos \omega t \), it is easy to calculate the values of \(|F(\omega)|^{*}\).

\(|F(\omega)|\) will have maxima at

\[
\omega = \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}
\]

and

\[
\omega = -\sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}
\]

*) A signal \( V_1(t) = e^{(s_1+i\omega)t} \) \( a > 0 \) is a damped signal and goes to zero. It is more interesting to calculate the properties of the network for undamped input voltages such as \( V_1(t) = V_1 e^{i\omega t} \) or \( V_1 \cdot \cos \omega t \).
The negative value will not have any physical significance.

For $\omega \to \infty$, $|F(i\omega)| \to 0$. Hence

$$|F(i\omega)|$$

The value of $\omega$ where $|F| = e^{-1}|F|_{\text{max}}$ defines the so-called cut-off frequency. Signals of frequency $> \omega_c$ do not get through the network.

In the case where no normal resistance is present, $R = 0$, $|F|_{\text{max}}$ is found at $\omega = \sqrt{1/LC}$, a higher value.

If an R-C-L network is connected to a current source or a voltage source, the transfer functions in both cases will be:

$$i(t) \quad \text{R-C-L} \quad V_{\text{out}}(t)$$

$$V_i(t) \quad \text{R-C-L} \quad V_{\text{out}}(t)$$

$$V_{\text{out}}(t) \left[ \frac{P(s)}{Q(s)} \right]_{\text{open network}}$$

$$V_{\text{out}}(t) \left[ \frac{P(s)}{Q(s)} \right]_{\text{closed network}}$$

$$P(s)_{\text{closed}} = P(s)_{\text{open}}$$

$$\frac{V_i}{i} = Z_{\text{in}} = \frac{\left[ \frac{Q(s)}{P(s)} \right]_{\text{closed}}}{\left[ \frac{Q(s)}{P(s)} \right]_{\text{open}}} = \frac{Q(s)_{\text{closed}}}{Q(s)_{\text{open}}}.$$  

The input impedance of a network is the Kirchhoff discriminant of the short-circuited network divided by the Kirchhoff discriminant of the open circuit.
CHAPTER II

SOME PROPERTIES OF A TRANSMISSION LINE

The impedance of a length of coaxial cable is to be calculated by means of its eigenmovements.

A line of length "L", open at the end, is taken:

As was stated
\[ Z = \frac{\text{discr. closed network}}{\text{discr. open network}}. \]

The discriminant network = 0 gives solutions \( s \), these being the eigenvalues. So the eigenvalues of the short-circuited line are the solutions of the discriminant of the closed line = 0.

Eigenmovements with \( \xi = v \cdot \tau \):

\[
\begin{align*}
\tau &= \frac{1}{4} T = \frac{1}{4} \cdot \frac{2\pi}{\omega} + \omega = \frac{1}{2} \frac{\pi}{\tau} \\
\tau &= \frac{3}{4} T = \frac{3}{4} \cdot \frac{2\pi}{\omega} + \omega = \frac{3}{2} \frac{\pi}{\tau} \\
\tau &= \frac{5}{4} T = \frac{5}{4} \cdot \frac{2\pi}{\omega} + \omega = \frac{5}{2} \frac{\pi}{\tau}
\end{align*}
\]

extrapolated: \( \omega = \frac{(2n + 1) \pi}{2\tau} \)

Here \( \cos \omega \tau = 0 \) has this solution, and \( \cos \omega \tau = \cosh i\omega t = \cosh s \tau \) is the numerator or a part of the numerator.
To find the denominator, the same method is followed with the open cable:

\[ \tau = \frac{1}{2} \quad T = \frac{1}{2} \frac{2\pi}{\omega} \rightarrow \omega = \frac{\pi}{\tau} \]
\[ \tau = T = \frac{2\pi}{\omega} \rightarrow \omega = \frac{2\pi}{T} \]

or in general: \( \omega = \frac{n \cdot \pi}{\tau} \).

The equation with this solution is \( \sinh \, st = 0 \), thus giving \( \sinh \, st \) as the denominator or part of the denominator:

\[ Z(s) = \frac{\cosh \, st}{\sinh \, st}. \]

The resemblance to an L-C network is obvious if we take a simple example:

\[ Z_{\text{in}} = Ls + \frac{1}{Cs} = \frac{LCs^2 + 1}{Cs} \]

\[ Z_{\text{in}} = 0 \quad \text{when} \quad LCs^2 + 1 = 0 \quad , \quad s_{x_{1,2}} = \pm i \sqrt{\frac{1}{LC}} \]

\[ Z_{\text{in}} = \infty \quad \text{when} \quad Cs = 0 \quad , \quad s_0 = 0. \]

In fact, a cable can be considered as a continuous series of L-C elements, thus
With inductance and capacitance continuously distributed over the line, there is no preference for a representation by means of network 1 or network 2:

**Network 1**

![Network 1 Diagram]

**Network 2**

![Network 2 Diagram]

A very special impedance for a network is the so-called "characteristic impedance" $Z_{ch}$. If an L-C-R network is terminated with this impedance, its input impedance will become equal to this characteristic impedance.

So for an observer measuring $Z_{in}$, it is as if the filter were nothing more than two resistanceless wires:

![L-C-R Circuit Diagram]

In the case of network 1, $Z_{ch}$ is to be calculated in this way ($a = c$):

\[
Z_{in} = a + \frac{b(c + Z_{ch})}{b + c + Z_{ch}}
\]

\[
Z_{in} = Z_{ch} = \frac{a(b + c + Z_{ch}) + b(c + Z_{ch})}{b + c + Z_{ch}}
\]

\[
Z_{ch} = \sqrt{a^2 + 2ab} = \sqrt{\frac{1}{c} \cdot \sqrt{\frac{1}{c} \cdot \sqrt{\frac{L}{c} \cdot \frac{S}{c} + 4}}}
\]

In example 2 ($a = c$):

\[
Z_{in} = Z_{ch} = \frac{a\left(b + \frac{cZ_{ch}}{c + Z_{ch}}\right)}{a + b + \frac{cZ_{ch}}{c + Z_{ch}}}
\]

\[
Z_{ch} = \sqrt{\frac{ba^2}{b + 2a}} = \sqrt{\frac{L}{c} \cdot \frac{2}{(cS)^2}}
\]

\[
Z_{ch} = \sqrt{\frac{L}{c} \cdot \sqrt{\frac{4}{LcS^2 + 4}}}
\]

There are other methods of calculating the characteristic impedance, such as: calculating the impedance of the circuit without any termination resistance, and calculating the impedance of the circuit with a termination resistance zero (short-circuited).
The geometric mean of these two values will give the characteristic impedance.

Example:

\[ z_{\text{open}} = a + b = \frac{LCs^2 + 2}{2Cs} \]

\[ z_{\text{short-circuit}} = a + \frac{bc}{b + c} = \frac{Ls}{2} + \frac{Ls}{LCs^2 + 2} \]

\[ z_{ch} = \sqrt{z_{op} \cdot z_{sh}} = \sqrt{\frac{L}{C} \cdot \frac{LCs^2 + 4}{4}}. \]

The characteristic impedance of a line is obtained by taking the geometric mean of the characteristic impedances of the two treated networks:

\[ z_{ch(\text{cable})} = \sqrt{z_{ch1} \cdot z_{ch2}} = \sqrt{\frac{L}{C}}. \]

This value could have been derived in a shorter way. If a fixed voltage \( V \) is applied over a line with a capacity of \( C \) F/m and \( L \) H/m, one can write:

\[ Q(t) = C \cdot \xi(t) \cdot V \]

\[ \frac{dQ}{dt} = C \cdot \frac{d\xi}{dt} \cdot V = C \cdot v_{\text{propagation}} \cdot V; \]

\[ i = C \cdot v \cdot V \quad (\text{II.1}) \]

\( v \) is the velocity of the waveform, and similarly

\[ \phi(t) = L \cdot \xi(t) \cdot i \]

\[ \frac{d\phi}{dt} = L \cdot v \cdot i; \]

\[ V = L \cdot v \cdot i. \quad (\text{II.2}) \]

Elimination of \( V/i \) gives

\[ v = \frac{1}{\sqrt{LC}} \quad (\text{II.3}) \]

for the velocity of the waveform in the line and with Eqs. (II.1) and (II.3):

\[ i = C \cdot \frac{1}{\sqrt{LC}} \cdot V = \sqrt{\frac{C}{L}} \cdot V = \frac{1}{z_{ch}} \cdot V. \]

\( z_{ch} = \sqrt{LC} \) is the characteristic impedance of the line.

Normally a transmission line is not to be represented by \( L \) and \( C \) only. At higher frequencies the skin effect will play an important role. Furthermore, capacitance between
two conductors or between a conductor and its shielding will not be negligible. Instead of the approach with only L and C, it is usually better to work with additional effective resistance R and the capacitive admittance G:

For an idealized cable, R = 0 and G = 0. With the eigenmovements of a length of transmission line, we have derived

\[ Z_{in} = \frac{\cosh i\omega t}{\sinh i\omega t} \]

\[ i\omega t = i\omega \cdot \frac{\text{length}}{v_{\text{propag}}} = i\omega \cdot \sqrt{L \cdot C} \cdot \xi = \sqrt{(i\omega L)(i\omega C)} \cdot \xi = P \cdot \xi. \]

With P one obtains another powerful cable constant, the propagation constant. In the non-idealized case, effective resistance and admittance are added to L and C. The general expression for the impedance becomes

\[ Z_{in} = Z_{ch} \cdot \frac{Z_T \cdot \cosh P\xi + Z_{ch} \cdot \sinh P\xi}{Z_{ch} \cdot \cosh P\xi + Z_T \cdot \sinh P\xi}, \]

where, with \( Z_T = Z_{ch} \), the input impedance \( Z_{in} = Z_{ch} \). The generalized terms for P and \( Z_{ch} \) will be

\[ P = \sqrt{(R + i\omega L)(G + i\omega C)} = \alpha + i\beta \]

\[ Z_{ch} = \sqrt{\frac{(R + i\omega L)}{(G + i\omega C)}}. \]

The attenuation of a signal in a transmission line is given with the equation

\[ V(x) = e^{-\alpha x} \cdot V(0) \]

\[ = V(0) \cdot e^{-\alpha x} \cdot e^{-i\beta x}, \]

where \( V(x) \) is the amplitude at a distance \( x \) from the origin 0; \( \alpha \) is the attenuation constant:

\[ \alpha = \sqrt{\frac{1}{2} \sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2) + \frac{1}{2}(RG - \omega^2 LC)}}, \]

and \( \beta \) is the phase constant:

\[ \beta = \sqrt{\frac{1}{2} \sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2) - \frac{1}{2}(RG - \omega^2 LC)}}. \]
An input signal consisting of a wide spectrum of harmonics encounters different delays for its components. The deformation of a signal due to this phase delay is called "phase shift or delay distortion".

Very often $\omega L \gg R$ and $\omega C \gg G$. In this case

$$\alpha \approx \frac{R}{2} \sqrt{\frac{C}{L}} + \frac{G}{2} \sqrt{\frac{L}{C}} \tag{II.4}$$

and

$$\beta \approx \omega \sqrt{LC} \tag{II.5}$$

$$V(x) = V(0) \cdot e^{-\alpha x} \cdot e^{-i\omega \sqrt{LC} \cdot x}$$

or better

$$V(x, t) = V(0, t) \cdot e^{-\alpha x} \cdot e^{-i\omega \sqrt{LC} \cdot x}$$

It is clear that in this special case the attenuation is a constant factor, so frequency-independent. The phase shift, however, seems to depend on $\omega$, hence on the frequency $f$. In this particular case the phase velocity $v_p$ equals the group velocity $v_g$:

$$v_g = v_p = \frac{1}{\beta}$$

In the more general case $\beta = \omega / v_p$ and, with $v_p = f \cdot \lambda$, one gets $\beta = 2\pi / \lambda$. The group speed $v_g = d\omega / d\beta$ and $v_g = x/t$:

$$v_g = \frac{d\beta \cdot v_p}{d\beta} = v_p + \beta \frac{dv_p}{d\beta} = v_p - \lambda \frac{dv_p}{d\lambda}$$

This relationship between phase and group velocity is known from vibration theory, and can be derived in many ways. In the special case of $\omega L \gg R$ and $\omega C \gg G$, the normal attenuation was frequency-independent. The fact that in this case $v_p = v_g$ makes the phase-shift term $\exp[-i\omega \sqrt{LC} \cdot x]$ look like $\exp[-i\omega (1/v_g) \cdot x] = e^{-i\omega t}$. Every harmonic component $e^{i\omega t}$ in the input signal will be compensated for, and there will be no phase distortion. In the general case, where neither $RGL$ nor $C$ is to be neglected, there will not be any attenuation distortion when $\alpha$ is not dependent on $\omega$, so $d\alpha / d\omega = 0$:

$$\frac{d\alpha}{d\omega} = \sqrt{\frac{\frac{1}{2} [R^2 + \omega^2 L^2] (C^2 + \omega^2 C^2) + \frac{1}{2} (R G - \omega^2 L C) / \omega} = 0$$

and this is true when $\omega = 1 / \omega G$. In this case $\alpha = \sqrt{R G}$ and $\beta = \omega / \sqrt{LC}$. With $\beta = \omega / v_p$, we again see that $v_p = v_g$, and we recognize the case where $\omega L \gg R$ and $\omega C \gg G$. So in the case where $\omega L \ll R$, there will be no attenuation distortion and no phase-shift distortion.
CHAPTER III

INTRODUCTION TO FOURIER ANALYSES

A periodic function \( f(x) \) with \( |f(x)| \leq M \), having only a finite number of discontinuities, is to be developed into a series of sine and cosine terms.

Suppose \( f(x) \) is periodic with period \( 2\pi \):

\[
f(x) = a_0 \cos 0x + a_1 \cos 1x + a_2 \cos 2x + ... \\
b_0 \sin 0x + b_1 \sin 1x + b_2 \sin 2x + ...
\]

with

\[
\int_{-\pi}^{+\pi} \cos nx \cdot \sin mx \, dx = 0 \quad n \neq m \\
\int_{-\pi}^{+\pi} \sin nx \cdot \sin mx \, dx = 0 \quad n \neq m \\
\int_{-\pi}^{+\pi} \cos nx \cdot \sin mx \, dx = 0,
\]

it is possible to write the coefficients as functions of \( f(x) \).

Multiply \( f(x) \) by \( \cos nx \), \( n = 0, 1, 2, 3, \ldots \), and integrate over the period \( 2\pi \):

\[
\int_{-\pi}^{+\pi} f(x) \cdot \cos 0x \cdot dx = \int_{-\pi}^{+\pi} a_0 \, dx + 0 + 0 + ...
\]

\[
2\pi a_0 = \int_{-\pi}^{+\pi} f(x) \, dx, \quad a_0 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(x) \, dx
\]

\[
\int_{-\pi}^{+\pi} f(x) \cos nx \, dx = 0 + \int a_n (\cos nx)^3 \, dx + 0 + ... \quad (n \neq 0)
\]

\[
a_n = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(x) \cos nx \, dx
\]

\[
b_n = 0.
\]

Multiply instead with \( \sin (nx) \)

\[
b_n = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(x) \cdot \sin nx \, dx
\]
Example:

\[
a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \, dx = \frac{\pi}{2\pi} = \frac{1}{2}
\]

\[
a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) (\cos nx) \, dx = \frac{1}{\pi} \int_{0}^{\pi} (\cos nx) \, dx = 0
\]

\[
b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) (\sin nx) \, dx = \frac{1}{\pi} \int_{0}^{\pi} (\sin nx) \, dx
\]

\[
= \frac{1}{\pi} \cdot \frac{1}{n} \int_{0}^{\pi} \sin nx \, dn \, x = \frac{1}{n\pi} (1 - \cos nx)
\]

\[n = 1, \quad b_1 = \frac{1}{1\pi} (1 - \cos 1\pi) = \frac{2}{1\pi}
\]

\[n = 2, \quad b_2 = \frac{1}{2\pi} (1 - \cos 2\pi) = 0
\]

\[b_{n(\text{even})} = 0, \quad b_{n(\text{odd})} = \frac{2}{3\pi}
\]

So

\[f(x) = \frac{1}{2} + \frac{2}{\pi} \sin x + \frac{2}{3\pi} \cdot \sin 3x + \ldots.
\]

The constant term \( \frac{1}{2} \) is obviously the effective value of the block signal \( f(x) \):
Generally, digital data signals are block functions or step functions of the time. Normally these functions do not have the period $2\pi$, but generally periods $T$.

So $f(x)$ is periodic (modulo $2\pi$) and is written as $f(x) = f(x + 2\pi)$; $f(t)$ is periodic (modulo $T$), $f(t) = f(t + T)$. Change the variable $Z = at$, such that when $t$ increments by $T$, $Z$ increments by $2\pi$:

$$a(t + T) = Z + 2\pi$$
$$a = \frac{2\pi}{T}.$$ 

Hence

$$Z = \frac{2\pi}{T} \cdot t$$
$$f(t) = f\left(\frac{Z}{a}\right) = f\left(\frac{T}{2\pi} \cdot Z\right) = F(Z).$$

The newly obtained function $F(Z)$ is modulo $2\pi$:

$$F(Z + 2\pi) = f \left[ \frac{T}{2\pi} \left( Z + 2\pi \right) \right] = f(t + T) = f(t) = F(Z).$$

So

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(Z) \, dZ = \frac{1}{2\pi} \int_{-T/2}^{T/2} f(t) \, dt \cdot \frac{2\pi}{T} \cdot t = \frac{1}{T} \int_{-T/2}^{T/2} f(t) \, dt$$

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cdot \cos n \cdot \frac{2\pi}{T} \cdot t \, dt \quad \text{\(2\pi/T = \omega_0\)},$$

etc.

A more straightforward development uses the complex Fourier coefficients:

$$f(t) = a_0 + \sum_{n=1}^{\infty} \left[ a_n \cos(nt) + b_n \sin(nt) \right]$$

$$= a_0 + \sum_{n=1}^{\infty} \left[ a_n \frac{e^{int} + e^{-int}}{2} + b_n \frac{e^{int} - e^{-int}}{2i} \right],$$

$$f(t) = \sum_{n=-\infty}^{n=\infty} C_n \cdot e^{int}, \quad \text{(III.1)}$$
in which $C_n$ is a complex number

$$C_n = \frac{1}{2}(a_n + ib_n),$$

$$C_{-n} = \frac{1}{2}(a_n - ib_n).$$

Multiply Eq. (III.1) by $e^{-int}$ and integrate from $-\pi$ to $+\pi$:

$$\int_{-\pi}^{+\pi} C_n \cdot e^{int} \, dt = 0 \quad n \neq 0.$$

So

$$\int_{-\pi}^{+\pi} f(t) \cdot e^{-int} \, dt = C_n \cdot 2\pi,$$

$$C_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(t) \cdot e^{-int} \, dt \quad (III.2)$$

when $f(t)$ is modulo $2\pi$, or

$$C_n = \frac{1}{T} \int_{-T/2}^{+T/2} f(t) \cdot e^{-jn\omega_0 t} \, dt \quad (III.3)$$

when $f(t)$ is modulo $T$.

Thus Eq. (III.1) becomes

$$f(t) = F(z) = \sum_{n=-\infty}^{+\infty} C_n \cdot e^{inz} = \sum_{n=-\infty}^{+\infty} C_n \cdot e^{iunt}.$$

In Eq. (III.3) $n\omega_0$ is the angular frequency of the $n^{th}$ harmonic of $f(t)$.

Let us now consider as an example an infinite pulse train symmetrical about $t = 0$: $f_0 = 1/T_0$ is the pulse repetition frequency, $\tau$ is the pulse width:

It is interesting to see the influence of the pulse width on the frequency spectrum, especially on the amplitude of the higher harmonics:

$$C_n = \frac{1}{T_0} \int_{-T_0/2}^{+T_0/2} A \cdot e^{-jn\omega_0 t} \, dt = \frac{1}{T_0} \int_{-\tau/2}^{+\tau/2} A \cdot \cos(-n\omega_0 t) \, dt = \frac{A\tau}{T_0} \sin\left(\frac{n\omega_0 \tau}{2}\right) \left(\frac{1}{n\omega_0 \tau/2}\right)$$
\[ \Delta f_0 = \frac{1}{T_0}, \quad \frac{n\omega_0}{2\pi} = nf_0 \]

\[ x(t) = \frac{\Delta t}{T_0} \sum_{n=-\infty}^{\infty} \frac{\sin(nf_0\tau)}{[nf_0\tau]} \cdot e^{in\omega_0 t}. \]

Thus the spectrum depends in a special way on the values of pulse width \( \tau \) and repetition frequency \( T_0 \). A smaller pulse, for example, with width \( \tau^* < \tau \), will have lower amplitudes of the major spectral contributors (with frequency \( f < 1/\tau^* \)). With \( 1/\tau^* > 1/\tau \), however, more frequencies of the higher harmonics will play a role. With the distance between two harmonics being \( 1/T_0 \), the consequence of increasing \( T_0 \) is a greater number of higher harmonics at values \( f < 1/\tau \). The amplitudes of the lower harmonics will decrease.

When \( T \to \infty \), we no longer have a periodic, infinite pulse train, but we are dealing with a single pulse:

\[ x(t) = \lim_{T \to \infty} \sum_{n=-\infty}^{n=+\infty} C_n \cdot e^{in\omega_0 t} \]

\[ \Delta f_n = f_n - f_{n-1} = \frac{1}{T}, \]

\[ x(t) = \lim_{T \to \infty} \sum_{n=-\infty}^{n=+\infty} \frac{1}{T} \cdot T \cdot C_n \cdot e^{in\omega_0 t} \]

\[ = \lim_{\Delta f_n \to 0} \sum_{n=-\infty}^{n=+\infty} \Delta f_n \cdot T \cdot C_n \cdot e^{in\omega_0 t} \]

\[ = \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{+\infty} x(t) \cdot e^{-i\omega t} \, dt \right] e^{i\omega t} \, d\omega. \]

The spectrum obtained is a continuous spectrum instead of a discrete one. When writing \( i\omega = s \), or more generally \( s = a + i\omega \),

\[ x(t) = \frac{1}{2\pi I} \int_{-\infty}^{+\infty} x^*(s) \cdot e^{st} \, ds \]  \hspace{1cm} (III.4)

\[ x^*(s) = \int_{-\infty}^{+\infty} x(t) \cdot e^{-st} \, dt. \]  \hspace{1cm} (III.5)
Equation (III.5) is known as the Laplace image of \( x(t) \), sometimes written as \( Lx(t) \). It facilitates calculus of spectral distributions, i.e.

\[
x(t) = A \left[ e(t + \frac{1}{2}T) - e(t - \frac{1}{2}T) \right].
\]

\[
x^*(s) = \frac{Lx(t)}{s} = \frac{A^2}{s} \cdot \frac{\sin \frac{\pi fT}{fT}}{\frac{\pi fT}{fT}}.
\]
CHAPTER IV

MODULATION

1. INTRODUCTION

It is obvious that the spectrum of a pure harmonic signal such as a cosine consists of one frequency component:

\[ y(t) = \cos \omega_c t \]

\[ x(t) = A \left[ \varepsilon \left( t + \frac{1}{2} \right) - \varepsilon \left( t - \frac{1}{2} \right) \right] . \]

If we want to know the spectrum of the product of the two signals \( x(t) \cdot y(t) = z(t) \), we will find the product of the spectra. So:

\[ z(t) \]

\[ L_z(t)(f) \]

\( y(t) \) is called the carrier, \( x(t) \) is the message signal, and \( z(t) \) is called the modulated carrier. The required bandwidth of \( z(t) \) is twice the bandwidth of the original signal \( x(t) \). The above-mentioned modulation is called amplitude modulation (AM). If a cable or network passes only the frequencies in the range \((f_1, f_u)\), modulation of the signal is essential. The carrier frequency could be \( f_c = \frac{1}{2}(f_1 + f_u) \).

The range \((f_1, f_u)\) is called the bandwidth of a cable or a network. Amplitude modulation is just one of the many modulation techniques. In general, modulation could be defined as follows: "Modulation is a systematic alteration of a carrier wave in accordance with a message".
1.1 Amplitude modulation (AM)

Message: \[ \text{Message: } \quad x_{\text{mess}}(t) = x_m(t); \]

Carrier: \[ y_c(t) = A_c e^{i\omega_c t}. \]

The modulated carrier is the sum of the carrier and a fraction of the product of the message and the carrier:

\[ x_{\text{mc}}(t) = A_c e^{i\omega_c t} + m \cdot x_m(t) \cdot A_c e^{i\omega_c t} \]

\[ x_{\text{mc}}(t) = A_c [1 + m \cdot x_m(t)] e^{i\omega_c t}. \]

**Example 1**

\[ x_m(t) = \sin \omega_m t, \quad y_c(t) = \sin \omega_c t, \quad \omega_c > \omega_m. \]

The modulation factor \( m = \frac{1}{2}(50\%). \)

**Example 2**

\( x_m(t) \) is a block signal,

\[ x_c(t) = \sin \omega_c t, \]

\( m = \frac{1}{2}, \omega_c > \text{block repetition frequency}. \)
Amplitude modulation is an example of a linear modulation technique [in another group of modulation techniques (the exponential modulation), two well-known examples will be considered: frequency modulation and phase modulation]:

Carrier $x_c(t) = A_c e^{j\theta_c(t)}$.

With amplitude modulation, $\theta_c = \omega_c t$ with exponential modulation, $\theta_c \neq \omega_c t$. In general: $\theta_c = \omega_c t + \phi(t)$.

1.1.1 Frequency modulation

As there is no longer "a frequency", we define the so-called "instantaneous frequency"

$$\omega_i = \frac{d\theta}{dt} = \omega_c + \frac{d\phi(t)}{dt}$$

$$f_i = \frac{1}{2\pi} \frac{d\theta}{dt} = \frac{\omega_c}{2\pi} + \frac{1}{2\pi} \frac{d\phi}{dt}.$$  

In frequency modulation

$$\frac{1}{2\pi} \frac{d\phi}{dt} = F \cdot x_m(t) \quad F = \text{modulation factor}.$$  

So the modulated carrier is:

$$x_{mc}(t) = A_c e^{j\theta(t)} = A_c e^{j\omega_c t + j2\pi f_i t} dt.$$  

Example 1

$$x_m(t) = \sin \omega_m t, \quad x_c(t) = \sin \omega_c t.$$  

The instantaneous frequency

$$f_i = f_c + \frac{1}{2\pi} \frac{d\phi(t)}{dt}$$

$$f_i = f_c + F \cdot x_m(t)$$

$$f_i = f_c + F \cdot \sin \omega_m t$$
Example 2

\[ x_m(t) \text{ is a block pulse, } x_c(t) = \sin \omega_c(t) = \sin \omega_c t; \quad f_1 = f_c + F \cdot x_m(t); \]

\[ x_m(t) \]

\[ x_c(t) \]

\[ x_{mc}(t) \]

Example 3

\[ x_m(t) \text{ is a block pulse, } x_c(t) \text{ is also a block pulse. Take, for example, } F = f_c \]

\[ x_m(t) \]

\[ x_c(t) \]

\[ x_{mc}(t) \]

The modulated carrier could be obtained in a rather easy way \( x_c^*(t) \) having twice the frequency of \( x_c(t) \):

\[ x_{mc}(t) = x_m(t) \cdot x_c^*(t) + \overline{x_m(t)} \cdot x_c(t). \]

So:

1.1.2 Phase modulation

A second example of an exponential modulation technique is phase modulation:

Carrier: \( x_c(t) = A_c e^{j\omega_c t} \).
In phase modulation

\[ \theta(t) = \omega_c t + \phi(t) \]

with

\[ \phi(t) = P \cdot x_m(t) \quad P = \text{modulation factor} \]

So the phase modulated carrier is

\[ x_{mc}(t) = A_c e^{j\theta(t)} = A_c e^{j\omega_c t + jP \cdot x_m(t)} \]

**Example 1**

\[ x_m(t) = \sin \omega_m t, \quad x_c(t) = \sin \omega_c t \]

The instantaneous frequency

\[ f_i = f_c + \frac{1}{2\pi} \frac{d}{dt} \theta(t) \]
\[ = f_c + \frac{P}{2\pi} \frac{d}{dt} \cdot \sin \omega_m t \]
\[ = f_c + \frac{1}{2\pi \omega_m} \cdot \cos \omega_m t \]

**Example 2**

\( x_m \) is a digital pulse train

\[ x_c(t) = \sin \omega_c t \]
\[ P = \pi \]
\[ \phi_i(t) = \omega_c t + \pi x_m(t) \]
Example 3

$x_m(t)$ is a digital pulse train, $x_C(t)$ is a digital signal as well (square pulse or digital clock). Again $P = \pi$:

The modulated carrier $x_{mc}(t)$ could be obtained in a very easy way. An EXCLUSIVE OR operation on $x_m(t)$ and $x_C(t)$ gives

$$x_m(t) \cdot \overline{x_C(t)} + \overline{x_m(t)} \cdot x_C(t) = x_{mc}(t).$$

In the digital cases, amplitude modulation is called "pulse amplitude modulation (PAM)". Frequency modulation is called "frequency shift keying (FSK)", and phase modulation is called "phase shift keying (PSK)". When private local transmission lines are used, modulation is generally not essential. Digital data can be transmitted as two, three, or four level signals. They might have a constant or a variable bit-length interval.

Some examples are given below; data word or bit-pattern (binary) 101101:

a) 0 1 0 1 1 0 1

b) 0

Advantage: strictly bit-length independent.
c) 

Advantage: no effective d.c. component (easy amplification).

d) 

Advantage: large pulse width (see Fourier analyses, \( \tau \) value).
(In data transmission the "1" signal is called "Mark" and "0" signal "Space".)

\[
\begin{array}{cccc}
1 & 0 & 1 & 1 & 0 & 1 \\
\end{array}
\]

The pulse shapes given above were all examples of the so-called basic signal shapes or "base band signals".

2. **DIGITAL MODULATION TECHNIQUES**

2.1 **Pulse duration modulation (PDM)**

Here the logical "one" and the logical "zero" are pulses of a different length (Morse).

2.2 **Pulse position modulation (PPM)**

Every bit is represented by a pulse, but "ones" are shifted for half a period

\[
\begin{array}{cccc}
1 & 0 & 1 & 1 & 0 & 1 \\
\end{array}
\]

\[
\frac{1}{2} T_0 \quad T_0
\]

2.3 **Delay modulation [delay shift key (DSK)]**

This is defined in the following way:

i) level change in the middle of every one bit;
ii) no level change for a "zero" bit followed by a "one" bit;
iii) level change at the end of every "zero" followed by another "zero":

\[
\begin{array}{cccc}
1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
\end{array}
\]

Advantage: the minimal pulse width is one bit-length.
2.4 Pulse code modulation (PCM)

An analogue signal is translated into a series of digital pulses by analysing the analogue signal shape at fixed intervals. In PCM only discontinuous message variations are allowed, so signal variations due to noise could be rejected as noise when the variation is somewhere in between the distinct message values.

Example:

![Diagram showing pulse code modulation](image)

On non-private data channels (i.e. the PTT network), dissipation outside a certain frequency range is often restricted or even forbidden. Consequently, data structures have to be analysed on higher harmonics outside the legitimate frequency band. For instance, an unmodulated data source of 500 kilobit/sec could be analysed. Let us consider the ASCII character "U", consisting of a series of alternate square pulses. In defining the opening or start bit as a logical one, the character appears as "1-0-1-0-1-0-1-0". The data rate of 500 kb/sec gives a bit width of \( \tau = 2 \mu \text{sec} \) and a block-repetition time \( T = 4 \mu \text{sec} \). With a signal amplitude of \( A = 2 \text{ V} = 2000 \text{ mV} \), one gets an effective d.c. value of

\[
\frac{\tau}{T} \cdot A = \frac{2}{4} \cdot 2000 \text{ mV} = 1000 \text{ mV}.
\]

The most important of the higher harmonics will be situated at 250 kHz with an amplitude of \( 2/\pi \cdot 1000 \text{ mV} = 636 \text{ mV} \). Other significant components are found at 750 kHz with an amplitude of \( 2/3\pi \cdot 1000 \text{ mV} = 212 \text{ mV} \) and at 1.25 MHz with an amplitude of \( 2/5\pi \cdot 1000 \text{ mV} = 127 \text{ mV} \). The usefulness of this little analysis will become clear if one takes into account the Swiss PTT restrictions for the frequency range from 150 kHz to 310 kHz. This band is used for the distribution of radio programs (télédiffusion), and our frequency component at 250 kHz of 635 mV will be strictly forbidden.

Clearly, an alternative transmission technique should be applied, such as data modulation on a 500 kHz clock carrier. Generally, the use of band-pass filters will be highly recommended in this type of situation.
CHAPTER V

TRANSMISSION TECHNIQUES

The transmission technique, where data bits are transmitted sequentially, is called "serial transmission". A transmission facility for data is called a "data channel". If only data are transmitted, the receiving system has to do the character synchronization. This is generally done by means of a series of frame or status bits. Every character or word has a fixed length, starting with a "zero" or START bit ending with some "one" or STOP bits.

This technique is called "asynchronous transmission". With only one, two, or three status bits it is a noise-sensitive technique, normally used in noise-free environment or at lower bit rates.

Another technique is block synchronization instead of character synchronization (synchronous transmission). There are not one or two single status bits per character, but a whole string of special status message or control characters enclosing a block of data. Special characters have been defined such as "SYN", "SOH", or "ACK", in addition to the normal alphabet (ASCII -- American Standard Code for Information Interchange -- 7 information bits and one parity bit).

Sometimes a timing, sampling, or clock signal is transmitted together with the data. The same transmission line could be used, in which case the clock signal has to be added to -- or modulated with -- the message or data stream.

If the messages use one channel for two-directional traffic, but no simultaneous bi-directional traffic is possible, the system is called "half duplex". If two separate and independent transmission channels are used for the two directions of the message flow, and simultaneous use is possible, the transmission system is called "full duplex". When only one-direction traffic is possible, the transmission system is called "simplex".

In parallel data channels, data are normally accompanied by a control bit, using a special line in a parallel direction to the data lines. For slow and average speed applications a serial data transmission system is preferred, because of some obvious advantages such as standard interface techniques, easy switching and multiplex facilities, use of PTT telephone networks, cable cost, etc.

An example of serial circuit switching in time is symbolized in the next figure:

![Diagram of serial circuit switching in time]

a: rotating switch or logics (electronic) equivalent
The period of the rotating switch is:

\[ T = \frac{2\pi}{\omega} \text{ sec} < t. \]

This technique is normally referred to as "time division multiplexing" or "time division switching (TDI)". It is essentially a "bit"-oriented transmission technique, contrary to the character or message-oriented transmission techniques based on a character or message as being the smallest transmission unit.

The PTT normally uses serial transmission techniques, so it is tempting to use this available transmission network for data transmission. There are, however, a number of problems to be encountered. Normally, a human voice has a frequency range of 300 Hz to 3000 Hz. A channel that can handle this spectrum is normally called a voice-grade channel. If the bandwidth is any bigger, it is called "wide band". In normal telephone traffic, the voice is frequency-modulated with a device called a "modem" (Modulator-Demodulator). The modem has in general an approved standard interface, following the so-called 'CCITT' specifications. Special higher quality lines or channels are sometimes used with a bandwidth of \( 12 \times 4000 \text{ Hz} = 48 \text{ kHz} \). These channels are equivalent to 12 voice-grade lines, and can consequently handle 12 independent voice messages, using in principle 12 different carrier frequencies. A group of \( 12 \times 4 \text{ kHz} = 48 \text{ kHz} \) is called a "primary group", and a unit of five primary groups with a range of \( 5 \times 48 \text{ kHz} = 240 \text{ kHz} \) is called a "secondary group".

For low-speed digital data transmission, it will in general be sufficient to use ordinary voice-grade lines.

For low-speed applications, voice-grade telephone lines are to be used in combination with PTT approved modems. With teletypes for example, a modem of up to 200 bauds\(^1\) could be used. The normal dial-up facility of a telephone set may be used. When contact has been established, the modem, instead of the telephone, will be connected to the lines\(^2\).

A teletype connected to a modem will send "one" and "zero" signals to the modem. The modem modulates this digital signal into two frequencies about an average frequency within the voice-grade bandwidth. Reception of two different frequencies in this band at the modem is demodulated into digital "zero" and "one" bits, to be interpreted by the teletype as such. The four frequencies are basically audible, and can be heard on the telephone set. This property is applied in the so-called acoustic coupler. This device translates the two tones received from the telephone set into binary signals, e.g. teletype. It also translates binary signals into two audible frequencies, destined for the microphone in the telephone set.

The normal voice-grade telephone lines pass by a telephone exchange. The switching and the organization of the transmission is done here (i.e. remodulation to higher frequencies for the higher bands: 4000-8000 Hz, 8000-12,000 Hz, etc.).

\(^1\) In analogue transmission, the baud rate is defined as the maximum number of information units a channel can handle per second. In digital transmission, the baud rate normally stands for the number of information elements per second.

\(^2\) An example of this is the frequency-modulated 200-baud modem, I.T.T. GH1101. This modem is used at CERN for external teletype connections: transmitter frequencies 1180 Hz for the logical 1 and 980 Hz for the logical 0; receiver frequencies 1850 Hz for logical 1 and 1650 Hz for a 0 (or vice versa).
Telephone traffic normally does not use the lowest bands of a transmission line. The bandwidth from zero to 300 Hz is given to telegraph or telex-users. Normal telegraph transmissions require a bandwidth of about 70 Hz. Telephone lines, made and adapted for optimal telephone traffic, obviously do not make ideal high-speed data channels. This is especially because of some of the artificial changes to lines in the form of adding lumped inductance at uniform intervals (loaded lines). For low frequencies \( R \gg \omega L \), so the attenuation constant \( \alpha = \sqrt{\frac{2}{R}} \) (curve 1). If however, \( L \) is made greater, \( \omega L \gg R \), and \( \alpha \approx \frac{R}{2} \sqrt{\frac{R}{L}} \) (curve 2), and is independent of the frequency and small when \( L \) is big (see page 9):

Added inductance, however, does not improve the phase distortion of a line. This is in general not disadvantageous for telephone messages, because the human ear is not very sensitive to this type of distortion. In digital data transmission, however, phase distortion is a serious cause of erroneous data interpretation. Especially at higher frequencies \( \omega L \gg R \), a phase distortion constant is found: \( \delta = \omega \sqrt{LC} \). Replacement of all the old loaded cables will be necessary, and this has sometimes already been done. Meanwhile, special phase distortion compensators are developed. Distortion compensators or equalizers have been designed for fixed and flexible line length.

On private lines one may use the simplified filters for a fixed cable length; but on switched network lines, compensators with an automatic adjustment are needed. The basic idea of the compensation technique is as follows: a pulse \( p(t) = e(t) - e(t - \Delta t) \) is sent to a delay line with "tap-off" facilities at every \( \Delta t \) sec:

In this way one obtains pulses which have each to be amplified by an amplifier \( C_n \). The added result will give a step function \( g_{\Delta t}(t) \). The amplification coefficients \( C_n \) should be positive or negative and also variable, in order to give an adjustable filter.

*) pupinized.
The corresponding spectrum when $\Delta t$ is small enough is:

$$A(\omega) = C_0 + 2 \sum_{n=1}^{N} C_n \cdot \cos(n\omega) .$$

The filters can be used at the receiver side, as well as on the transmitter side. In the latter case, a low pass-filter is used with a cut-off frequency of $1/2\pi$. There are two ways of automatically adjusting these filters. In one case, there is a defined training time after a line change has taken place. In the other case the message is used directly. The latter method needs faster adjustment techniques, and is basically only used in synchronous data transmission where status characters such as "SYNC" are repeated several times before the actual message. For private lines, filters could be used at the transmitter and at the receiver side. Here the adjustment of the amplification factor $C$ is made only once. Instead of a delay line, shift registers of flip-flops could be used in the filter. In this case it is called a "digital filter".

Noise

Until now, data transmission was considered ignoring any external conditions. Only the quality of "characteristic properties" of the transmission line or "channel" were taken into account. External noise is a very important error source in data communication. In principle, noise is defined as all signal interference causing errors in the message.

First of all, there is the thermal motion of the electrons in the conductors and semiconductors. In general, this natural noise has frequencies with a constant power spectrum and a Gaussian probability distribution (white noise).
Other types of noise are variations of the above-mentioned white noise. They are generally called "coloured noise". At CERN, unnatural noise is the main reason for message errors. The already known mains interference of 50 Hz can usually be traced back, but ground noise in the different CERN locations, or high-voltage "impulsive noise" in the experimental areas is sometimes difficult to combat. Small potential differences between a data transmitter and a data receiver will almost certainly cause erroneous data interpretation at the receiver side, when data are sent over a single wire. For this reason, it is quite common to use two lines instead of one. A transmitter drives a current into one of the lines and absorbs it out of the other. At the receiver, the line is terminated with the characteristic cable resistance, and the voltage built up over this resistance is to be amplified by a differential amplifier. When the bit value changes to its complement, an inverted current causes an inverted voltage drop over the termination resistance.

When transistor T₁ is conducting, transistor T₂ is not conducting, and vice versa.

To avoid the wasted power dissipation in R₁ and R₂, these resistances could be replaced by transistors. The transistor replacing R₁ should be conducting when transistor T₂ is conducting, etc.

At the receiver end, there is no ground connection and the differential amplifier detects only voltage differences. If there is an introduction of noise on the data lines, it will in general be of common polarity on the two conductors (the twisted pair). It is com-
sequently not noticed by a circuit which is only sensitive to changes in the polarity of its inputs. There is, however, a certain limit above which this common voltage on the two lines may not rise. This so-called common mode rejection is normally of the order of several volts, while differential sensitivity is of the order of some millivolts. With sufficient current for the differential signals, resistance bridges will improve the common mode rejection.

When unidirectional lines are used, reamplification is easy. However, for bidirectional lines or so-called "half duplex" connections, more complicated methods are used.

For analogue data, the PTT uses amplifiers with so-called line simulators in order to avoid oscillation. Half duplex digital data transmission often uses two different voltage levels for the two traffic directions. Amplification offers no special problem in this case.
CHAPTER VI

ERROR DETECTION AND ERROR CORRECTION

It is naturally essential for a receiver to know if the received message is a good or an erroneous message. Messages can, of course, be repeated many times if necessary. This is, however, a most inefficient way. At low data rate requirements the data message is often reflected at the receiver side. In this way the transmitter is able to compare the reflected message with the original, and to detect any errors as such. The technique is called "echo-plexing".

Instead of repeating a message completely, only a part of the message or a function of the message could be used. This will not give the "redundancy" of the repeat technique; but by sending the additional message bits in a more sophisticated way, fairly good error detection is still possible. The additional error check bits or redundancy bits are transmitted at the end of the corresponding message bits. In most cases it is sufficient to take as a redundancy function the binary addition of the message bits -- modulo 2.

A binary number N, modulo 2, means \( N + 2_{10} = N \). So

\[
\begin{align*}
0 + 2_{10} &= 0 + 10_2 = 0 \\
1 + 10_2 &= 1 \\
10_2 + 10_2 &= 10_2 = 0 .
\end{align*}
\]

(VI.1)

In this case it is the least significant bit of the addition. This function is called the parity function or parity bit.

**Example:**

<table>
<thead>
<tr>
<th>Message</th>
<th>1101</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parity bit is</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Message</th>
<th>1111</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parity bit is</td>
<td>0</td>
</tr>
</tbody>
</table>

There exists a parallel (lateral) parity check (one parity bit/word, character), and a longitudinal parity check on entire blocks of data. When more redundancy is required, more complicated error codes could be introduced. The extra parity bit makes single and odd error detection possible.

Additional bits allow for higher-order bit detection, but also allow for error correction.

**Example**

Suppose a three-bit message 110 is transmitted and received as 111. If the third bit stands for the parity bit, it is possible to check for single errors. The received 111 could have been 011, 101, or 110.
If the patterns 011 and 101 had been forbidden pattern combinations, the restoration to 110 would have been easy. We can say that in general, checking and correcting of bit errors are related to the so-called "distance" between officially defined messages.

The "hamming distance" between two messages is the number of bits that must be changed to make the messages the same.

Example

The distance between 101001 and 011011 is 3. In the case of the parity bit, 011 and 101 were not allowed. The hamming distance between the tolerated message 110 and the undefined words 011 and 101 is "two". The first well-defined word should have been 100.

With the ordinary parity bit, a distance of "two" was created.

Example

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000, 0011, 0101, 0110, 1001, 1010, 1100, etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The creation of a distance of "three" is not as easy as the creation of a distance "two". A well-known code, with a distance of three, is the hamming code. Here a message of 4 bits needs three additional error code bits. Bits 3, 5, 6, and 7 are message bits. Bit 1 (2^0) is the parity sum of bits 3, 5, and 7. (So the total of 1, 3, 5, and 7 is zero.) Bit 2 (2^1) is the parity sum of bits 3, 6, and 7, and bit 4 (2^2) is the sum of bits 5, 6, and 7.

Example 1

The check made on bit 1 says "Error", hence bit 1, 3, 5, or 7 is wrongly received.

The check made on bit 2 says "No error", so bits 2, 3, 6, and 7 are correct.

The check made on bit 4 says "No error", so bits 4, 5, 6, and 7 are correct. Obviously bit 1 is the guilty one.

In general, the erroneously received bit is found by ordering the parity checks. If a check on parity bit 2^n gives "Error", a corresponding bit 2^n is set to "one". So in our example: 2^2 + 2^1 + 2^0 = 0 + 0 + 1 = 1, and consequently bit 1 has been erroneous.

Example 2

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>bit (2^2 + 2^1 + 2^0) = 110</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

So bit 6 is erroneous.

Hamming codes can be generated and checked with simple electronic circuitry, i.e. with dual full adders (Fairchild 9304). We have seen that hamming distances of "two" allowed for single-bit error detection (parity bit). A distance of "three" makes single error correction possible, a distance of four make double-bit error detection and single-bit correction possible, etc.

With integrated circuit prices falling, more complicated error codes could be implemented in the hardware. Another frequently used technique is based on a division of a
digital data pattern by a fixed digital number. The remainder is to be transmitted after
the digital data pattern has been transmitted. A similar division should be performed at
the receiver side, and the remainder could be compared. Suppose that a data word 10101100
is to be divided by 1101. The division will use the modulo 2 addition (* subtraction)

\[ 0 + 0 = 0, \quad 0 + 1 = 1, \quad 1 + 1 = 0 \]

hence

\[ 1 = -1 \]

So the division of 10101100 by 1101 gives 0011 as a remainder. This division could be per-
formed with a flip-flop shift register. The most significant bit of a series of data bits
will enter the register first. In this example three flip-flops are used. Two feedbacks
are applied, one to flip-flop 1 and one to flip-flop 3. To define the total number of
flip-flops, a feedback to the output of flip-flop 3 could be envisaged.

With the shift register initially on zero, three zeros are returned to the flip-flops by
means of the feedback connections, before the first message bit ≠ 0 is returned. At this
moment, the contents of the shift register will be 101.

A new shift will enter a new message bit into the first stage of the shift register.
The contents of the third flip-flop, however, are fed back to the first and third flip-flops.
So the new contents of the shift register are made out of the modulo 2 addition (subtraction)
of the contribution of the message bits and the feedback of the contents of flip-flop 3.
Hence:

<table>
<thead>
<tr>
<th>Shifted data</th>
<th>Waiting message bits</th>
<th>Feedback</th>
<th>New contents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.....................</td>
<td>1 1 0 1</td>
<td>1 1 1</td>
</tr>
<tr>
<td></td>
<td>1 0 1 0</td>
<td>(1100)</td>
<td></td>
</tr>
</tbody>
</table>
Again a new shift will enter the first of the waiting message bits into the shift register.

Shifted data

Waiting message bits

\[ \begin{array}{cccc}
1 & 1 & 1 & 1 \\
\end{array} \]

Feedback

\[ \begin{array}{cccc}
1 & 1 & 0 & 1 \\
\end{array} \]

\[ + \quad (\cdot) \]

New contents

\[ \begin{array}{cccc}
0 & 0 & 1 & 0 \\
\end{array} \]

It will be clear that the shift register performs the digital division. When the division is performed, the shift register will contain the remainder of the division; this remainder is now to be used as an error check code. The original message could be divided in the same way at the receiver side. Comparison of this remainder with the received remainder will reveal possible transmission errors.

In the example, message + error check bits will be 10101100-011. In order to see the possibilities one has with this division technique, a digital bit pattern will be written as a polynomial in \( x \). So the pattern 10101 = 1.2^4 + 0.2^3 + 1.2^2 + 0.2^1 + 1.2^0 is to be written in general terms of \( x \): 1.\( x^4 \) + 0.\( x^3 \) + 1.\( x^2 \) + 0.\( x^1 \) + 1.\( x^0 \) = \( x^4 + x^2 + 1 = P(x) \).

The polynomial, representing the message + error check bits of the example, will be

\[ P_{10}(x) = x^4 + x^2 + x^3 + x + 1 \]

The original message 10101100 gives a polynomial

\[ M_7(x) = x^7 + x^5 + x^4 + x^2 \]

and the remainder as \( R_2(x) = x + 1 \).

The feedback pattern defines the number by which the message has been divided, and is written as a polynomial \( G_3(x) = x^3 + x + 1 \). The division is now to be written in polynomial form:

\[ P_{10}(x) = x^3 \cdot M_7(x) + R_2(x) \]

The answer of the division could be represented by \( A_7(x) \)

\[ \frac{x^3 \cdot M_7(x)}{G_3(x)} = A_7(x) + \frac{R_2(x)}{G_3(x)} \]

With "addition mod. 2", \( R_2(x) \) is to be added on both sides giving:

\[ x^3 M_7(x) + R_2(x) = A_7(x) \cdot G_3(x) = P_{10}(x) \equiv V(x) \quad (VI.2) \]

\( M_7(x) \) is an example of a message polynomial
\( G_3(x) \) is called a "generator" polynomial
\( R_2(x) \) is called the "redundancy" polynomial.

\( V(x) \) is divisible by \( G(x) \).

If, instead of \( V(x) \), another bit pattern is received, represented by a polynomial \( H(x) \), this is to be written as \( H(x) = V(x) + E(x) \), in which \( E(x) \) stands for the polynomial of the bits, different from \( V(x) \).

If \( H(x) \) should again happen to be divisible by \( G(x) \), the error cannot be traced back.
Example

Message 10011, $M_4(x) = x^4 + x + 1$. $G_1(x) = x + 1$

\[
\frac{x^4 \cdot M'(x)}{G_1(x)} = \frac{x^5 + x^2 + 1}{x + 1} = \frac{x^4(x + 1) + x^5 + x^2 + 1}{x + 1} = x^4 + x^3 + x^2 + \frac{1}{x + 1}.
\]

$R(x) = 1$

$A_4(x) = x^4 + x^3 + x^2$.

$V(x) = x^3 + x^2 + 1 + 1 = x^5 + x^2$, corresponding to a bit pattern 100100.

Suppose one bit was wrongly received, giving a received pattern 100101:

$H(x) = x^3 + x^2 + 1$.

Hence the error polynomial $E(x) = 1$.

The question whether $H(x)$ is divisible by $G(x)$ is strictly the question whether $E(x)$ is divisible by $G(x)$:

\[
\frac{E(x)}{G(x)} = \frac{1}{x + 1}.
\]

Obviously, here $E(x)$ is not divisible by $G(x)$, so the error is to be detected. If two-bit errors had been introduced, $E(x)$ would have been divisible by $G(x)$; but with three-bit errors, detection would have been possible. The example of $G(x) = x + 1$ is one of the many possible cyclic error codes. The power of a code does not strictly depend on the length alone, i.e. $G(x) = x + 1$ already detects all odd numbers of errors.

Prove $V(x) = A(x) \cdot G(x) = A(x) \cdot (1 + x)$

take $x = 1$

$V(1) = A(x) \cdot (1 + 1) = A(x) \cdot 0 = 0$.

So the polynomial of $V(x)$ consists of an even number of $x$ terms. If, due to an odd number of errors, an odd number of $x$ terms should be added to $V(x)$, the total number of $x$ terms should be odd. $H(1) \neq 0$. Generator polynomials of higher order, such as $x^n + 1$, etc., will be at least as powerful. If we return to the shift register, we find that it was the feedback that defined the generator polynomial $G(x)$; that is:

\[
\text{digital data} \quad \rightarrow \quad + \quad x^3 \quad + \quad x^2 \quad x \quad + \quad 1
\]

\[
gives \quad G(x) = x + x^3 + x^4.
\]

It is obvious that flip-flop 1 does not play any role and has no significance. Only if a feedback to flip-flop 1 is applied, will it acquire any meaning. Consequently, $G(x)$ always has a "1" term. Hence, $G(x)$ never divides $x$. 
It is interesting to see how powerful cyclic codes are in respect to so-called "burst errors". It is very often the case that instead of single-bit errors, whole series of bits at a stretch are wrongly received. A series of \( b \) erroneous bits, one after the other, is called an error burst of length "\( b \)".

If \( V_n(x) \) is a code polynomial, it symbolizes "\( n \)" message bits and "\( n-k \)" error check bits. So \( V_n(x) = 1 + x + x^2 + \ldots + x^{n-k-1} + x^{n-k} + \ldots + x^n \). The generator polynomial of this code has consequently a degree \( n-k \), so \( G_{n-k}(x) \). An error burst of length \( b \) is symbolized by the error polynomial \( E(x) \):

\[
E(x) = c_i x^1 + c_{i+1} x^{i+1} + \ldots + c_j x^j,
\]

\( c_i, \ldots, c_j = 1 \) or \( 0 \), \( j + 1 - i = b \).

If \( E(x) \) does not divide \( G_{n-k}(x) \), the burst is to be detected:

\[
\frac{E(x)}{G_{n-k}(x)} = \frac{x^i(c_i \cdot 1 + c_{i+1} \cdot x + \ldots + c_j x^{j-i})}{G_{n-k}(x)}.
\]

\( x^j \) is not divisible by \( G_{n-k}(x) \) because \( G_{n-k}(x) \) is not divisible by \( x \). The polynomial 

\[
x^{j-1} + \ldots + 1
\]

is not divisible by \( G(x) \) if \( j - i < n - k \).

With \( j - i = b - 1 < n - k \), we get \( b < n - k + 1 \) or \( b \leq n - k \). This is to be formulated as: "Cyclic codes generated by generator polynomials of degree \( n-k \) detect error bursts of length \( b \leq n-k \)." Error codes have been the subject of studies for about twenty years, and many new powerful codes are regularly discovered. Implementation of error codes is today mostly done by hardware. However, if high flexibility is required, and sufficient computer capacity is available, the generation and control of the code could be done by software.

It is obvious that no attempt has been made to give a detailed survey of all the theories and techniques in data transmission. With applications at CERN in mind, only wired connections were treated as data channels. For practical applications of the theories and techniques mentioned, the reader is referred to the papers on this subject, some of which are listed in the attached Bibliography.
BIBLIOGRAPHY