híradástechnika

VOLUME LIX.

2004/6

Selected Papers

Wave theory

Photonics and Technology

Telecommunications policy

Scientific Association for Infocommunications

Contents

Cover: Tower of Babel – Since this event people speaking different languages can't understand each other, except for the engineers.

Foreword

lajtha.gyorgy@ln.matav.hu

While reading through the Híradástechnika issues of the first half-year, I selected for English publication several articles which more or less are supporting the development of the telecommunications and communications sectors. Several interesting studies were written from the component and technology field. This half-year may well emphasise the fact that the recent achievements of electronics, nano technology and photonics have reached the level where they can play a determinant role in telecommunications.

Disregarding their sequence in the journal, first of all I want to speak of photonics. Three articles (those of Tamás Marozsák, Gábor Kovács, Attila Kovács – Ildikó Deme, respectively) report on photonic devices, new laser types, wavelength switches, semiconductor amplifiers and packet switched optical routers, which elements will certainly step up the propagation process of optical telecommunications. The Hungarian researchers are also doing their best to find out how to reduce the number of the electric/optical transformations along the connection. The number of the functions to be realised on the optical level is growing steadily, thus permitting the reduction of the broadband telecommunications costs. As newer and newer technologies emerging in the fields of photonics fulfil urgent needs, we have been endeavouring to present the achievements of the national research works in this issue, as well. It is worth contrasting these articles with our previous English issue, where the optical burst switching, the distributed parameter and Raman amplifiers have been brought into the light (December 2003).

Because the service itself and also the production require nano technologic solutions, we will first include in the following group a review of the National Nano Technology Conference. These achievements, however, are not connected with the modernisation of networks of the near future, being marked only with a tendency. This trend is underscored by the technologic researches that contributed to the creation of highly efficient solar cells, with the laser beam being a promising factor in the mounting miniaturisation and the precise completion of bores in the micrometer range (Edvárd Kuthy, Péter Gordon–Bálint Balogh).

Réka Limbek and Péter Sziklai handle one of the potential fields of high quality coding. This group may also outline the presentation of ENUM, the solution enabling the improvement of the service quality and linking the e-mail addresses to the phone number. Quality and reliability are important and play role also in the field of computer-technology. The investigation of fault-tolerant computers (Zoltán Katona) and of the relevant protocols is equally strengthening this tendency.

Thus we were studying to outline the development backgrounds of telecommunications, hoping that all these results will in the short term contribute to the enhancement both of the prices and to the quality of the different services. Mentioned last, but put first in the journal is the contribution of Orsolya Ferencz, with the presentation of an actually important solution of the Maxwell equations.

There is a chance that the technological and experimental background that we became familiar with could be used extensively in the near future. The acceptance of the countrywide implementation of the theoretical findings is a crucial factor. By means of these articles we intend to attract the attention to the results and suggest their use by the telecommunications industry. In one or two years from now we might be able to report in our English-language issue the realisation of the first end-to-end optical connection based on the photonic achievements. Along with this we would like also to have a report on the improvement of the reliability of the computer technology used for the control and the management of such a connection, i.e. on the experience that the failure rate is tending to zero.

Dr. Lajtha György

Short impulse propagation in waveguides

ORSOLYA E. FERENCZ

Space Research Group, Eötvös University, Geophys. Department
spacerg@sas.elte.hu

Keywords: Maxwell-equations, radio waves, ionosphere, reflections

One of the most important research topics is the investigation of (short) impulse propagation in waveguides. The known solutions are based upon the well-known monochromatic approaches, examining the different frequencies separately or building the model and the theory on a fundamentally monochromatic starting point (e.g. permittivity tensor, which is defined originally by assuming an type solution form). In this paper a completely new theoretical model and solving method will be presented for a rectangular waveguide filled by vacuum, excited by an arbitrarily formed electromagnetic signal (Dirac or real, even short impulse).

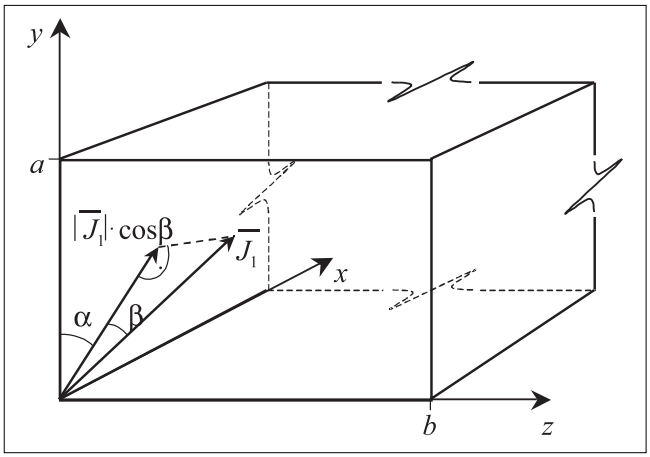
This method avoids the application of the former assumptions regarding the sinusoidal waveforms.

The obtained closed-formed solution leads back to the former ones known for monochromatic excitation, by using a sinusoidal excitation with a given frequency, but obviously the new formula results a general solution of the problem.

Introduction

The model, in which the new theoretical method will be presented, is a rectangular waveguide filled by vacuum and bordered by perfect conductor walls.

Fig.1. The structure of the model



The generally directed exciting current density is

$$\overline{J}_1 = J_{1x}\cdot\bar{i} + J_{1y}\cdot\bar{j} + J_{1z}\cdot\bar{k}$$
$$\left|\overline{J}_1\right| = \delta(t)\cdot\delta(x)\cdot B_1(y)\cdot B_2(z) \qquad (1)$$

where $B_1(x)$ and $B_2(z)$ are envelope functions containing the boundary conditions

$$B_1(0) = B_1(a) \equiv 0 \quad \text{and} \quad B_2(0) = B_2(b) \equiv 0 \qquad (2)$$

So the current density is (3)

$$\overline{J}_1 = \left|\overline{J}_1\right|\sin\beta\,\bar{i} + \left|\overline{J}_1\right|\cos\beta\cos\alpha\,\bar{j} + \left|\overline{J}_1\right|\cos\beta\sin\alpha\,\bar{k}$$

Further – without any theoretical restriction – let the following, sufficiently general form of the excitation be applied (4):

$$\overline{J}_{1x} \equiv 0$$
$$\overline{J}_1 = \delta(t)\delta(x)B_1(y)B_2(z)\cos\alpha\,\bar{j} + \delta(t)\delta(x)B_1(y)B_2(z)\sin\alpha\,\bar{k}$$
$$\left|\overline{J}_1\right| = J_1 = \delta(t)\delta(x)B_1(y)B_2(z)$$

The new solving method

The theoretical basis of the solving method can be found in [1, 2] for transient plane waves.

The equations to be solved are Maxwell' equations [3, 8]

$$I.\quad \overline{\nabla}\times\overline{H} = \overline{J}_1 + \varepsilon_0\frac{\partial\overline{E}}{\partial t} \qquad (5)$$
$$II.\quad \overline{\nabla}\times\overline{E} = -\mu_0\frac{\partial\overline{H}}{\partial t}$$
$$III.\ \overline{\nabla}\overline{H} = 0$$
$$IV.\ \overline{\nabla}\overline{E} = \frac{\rho}{\varepsilon_0}$$

Let the retarded potential be introduced on the well known way

$$\overline{\nabla}\times\overline{A} = \overline{H}$$
$$\overline{E} + \mu_0\frac{\partial\overline{A}}{\partial t} = -\overline{\nabla}\psi \qquad (6)$$

The Lorenz-condition is valid, as usual

$$\left(\overline{\nabla}\overline{A} + \varepsilon_0\frac{\partial\psi}{\partial t}\right) = 0 \qquad (7)$$

So, the equation to be solved is

$$\nabla^2 \overline{A} - \varepsilon_0 \mu_0 \frac{\partial^2 \overline{A}}{\partial t^2} = -\overline{J}_1 \tag{8}$$

As the excitation is a generally formed signal with exact starting point according to time and space, the Laplace-transformation can be applied, as

$$\begin{aligned} t &\xleftrightarrow{\mathrm{L}} s \\ x &\xleftrightarrow{\mathrm{L}} p \\ y &\xleftrightarrow{\mathrm{L}} u \\ z &\xleftrightarrow{\mathrm{L}} l \\ f(t,x,y,z) &\xleftrightarrow{\mathrm{L}} F(s,p,u,l) \end{aligned} \tag{9}$$

Because of the presence of derivative terms, initial values according to all coordinates will appear. Usually these initial values contain information regarding the energetic state of the medium. However, in this case the medium is considerable to be free of energy before the switching on of the excitation. Therefore in the further all the initial values have to be taken into account as 0.

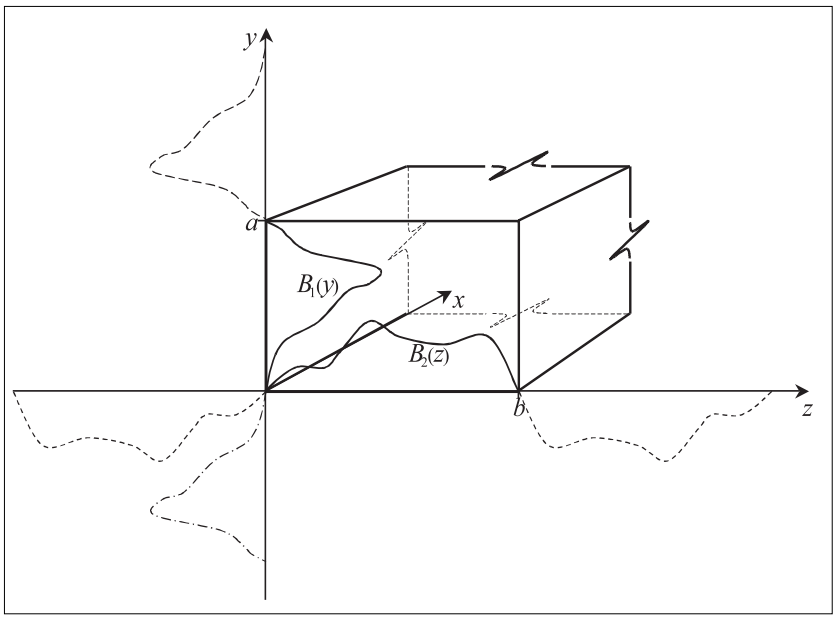
The transformed equations to be solved are:

$$\begin{aligned} H_x(s,p,u,l) &= u\, A_z(s,p,u,l) - l\, A_y(s,p,u,l) \\ H_y(s,p,u,l) &= -p\, A_z(s,p,u,l) \\ H_z(s,p,u,l) &= p\, A_y(s,p,u,l) \end{aligned} \tag{10}$$

further

$$\begin{aligned} E_x(s,p,u,l) &= \frac{1}{\varepsilon_0}\frac{1}{s}\left[p u\, A_y(s,p,u,l) + p l\, A_z(s,p,u,l)\right] \\ E_y(s,p,u,l) &= \frac{1}{\varepsilon_0}\frac{1}{s}\left[u^2 A_y(s,p,u,l) + u l\, A_z(s,p,u,l)\right] - \mu_0 s\, A_y(s,p,u,l) \\ E_z(s,p,u,l) &= \frac{1}{\varepsilon_0}\frac{1}{s}\left[u l\, A_y(s,p,u,l) + l^2 A_z(s,p,u,l)\right] - \mu_0 s\, A_z(s,p,u,l) \end{aligned} \tag{11}$$

Fig.2. The envelope functions



The Laplace-transformed form of the exciting current density is

$$\begin{aligned} J_1(s,p,u,l) &= \\ &= \iiint_0^\infty \delta(t)\delta(x) B_1(y) B_2(z) \cdot e^{-st} \cdot e^{-px} \cdot e^{-uy} \cdot e^{-lz}\, \mathrm{d}t\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \\ &= 1 \cdot 1 \cdot B_1(u) B_2(l) \end{aligned} \tag{12}$$

It is very important to choose B_1 and B_2 functions suitably. These envelope functions contain the boundary conditions resulted from the geometrical structure of the model. This can be seen in *Fig.2.*

The boundary conditions to be validated are

$$B_1(0) = B_1(a) = B_2(0) = B_2(b) \equiv 0 \tag{13}$$

As it is usual [3], the envelope functions B_1 and B_2 can be extended periodically, and it is possible to describe them by Fourier-series:

$$\begin{aligned} B_1(y) &= \sum_{m=0}^{\infty} C_m \cdot \mathrm{e}^{\mathrm{j}m\frac{\pi}{a}y} \\ C_m &= \frac{1}{2a}\int_{-a}^{a} B_1(y) \cdot \mathrm{e}^{-\mathrm{j}m\frac{\pi}{a}y}\, \mathrm{d}y \end{aligned} \tag{14}$$

and

$$\begin{aligned} B_2(z) &= \sum_{n=0}^{\infty} C_n \cdot \mathrm{e}^{\mathrm{j}n\frac{\pi}{b}z} \\ C_n &= \frac{1}{2b}\int_{-b}^{b} B_2(z) \cdot \mathrm{e}^{-\mathrm{j}n\frac{\pi}{b}z}\, \mathrm{d}z \end{aligned} \tag{15}$$

where C_m and C_n are Fourier-coefficients, a and b are geometrical parameters of the waveguide, m and n are integers

$$\begin{aligned} m &= 0, \pm 1, \pm 2, \ldots \\ n &= 0, \pm 1, \pm 2, \ldots \end{aligned} \tag{16}$$

The Laplace-transformed forms of (19) and (20) are

$$\begin{aligned} B_1(u) &= \sum_{m=-\infty}^{\infty} \frac{C_m}{u - \mathrm{j}m\frac{\pi}{a}} \\ B_2(l) &= \sum_{n=-\infty}^{\infty} \frac{C_n}{l - \mathrm{j}n\frac{\pi}{b}} \end{aligned} \tag{17}$$

The poles according to p can be determined from

$$p^2 + u^2 + l^2 - \varepsilon_0 \mu_0 s^2 = 0 \tag{18}$$

Investigating the poles according to u and l it becomes obvious, that only the poles originating from the excitation will yield terms different from zero in the amplitudes. Executing the well known steps of the inverse Laplace-transformation on the common way and using the

$$s = \mathrm{j}\omega \tag{19}$$

substitution, the spectral forms of the filed components depending on the spatial variables are:

$$H_x(\omega,x,y,z)=\sum_m\sum_n\frac{P_-C_mC_n}{2k_x(\omega)}\cdot e^{j[k_x(\omega)\cdot x\cdot M\cdot N]}-\sum_m\sum_n\frac{P_-C_mC_n}{2k_x(\omega)}\cdot e^{j[-k_x(\omega)\cdot x\cdot M\cdot N]}$$

$$H_y(\omega,x,y,z)=\sum_m\sum_n C_mC_n\sin\alpha\cdot e^{j[k_x(\omega)\cdot x\cdot M\cdot N]}+\sum_m\sum_n C_mC_n\sin\alpha\cdot e^{j[-k_x(\omega)\cdot x\cdot M\cdot N]} \quad (20)$$

$$H_z(\omega,x,y,z)=\sum_m\sum_n C_mC_n\cos\alpha\cdot e^{j[k_x(\omega)\cdot x\cdot M\cdot N]}+\sum_m\sum_n C_mC_n\cos\alpha\cdot e^{j[-k_x(\omega)\cdot x\cdot M\cdot N]}$$

and

$$E_x(\omega,x,y,z)=\sum_m\sum_n\frac{-P_+C_mC_n}{2\omega\varepsilon_0}\cdot e^{j[k_x(\omega)\cdot x\cdot M\cdot N]}+\sum_m\sum_n\frac{P_+C_mC_n}{(-2)\omega\varepsilon_0}\cdot e^{j[-k_x(\omega)\cdot x\cdot M\cdot N]}$$

$$E_y(\omega,x,y,z)=\sum_m\sum_n\frac{\left[(-jm\pi/a)(jP_+)-\varepsilon_0\mu_0\omega^2\cos\alpha\right]C_mC_n}{2j\varepsilon_0\omega\,jk_x(\omega)}\cdot e^{j[k_x(\omega)\cdot x\cdot M\cdot N]}+$$
$$+\sum_m\sum_n\frac{\left[(-jm\pi/a)(jP_+)-\varepsilon_0\mu_0\omega^2\cos\alpha\right]C_mC_n}{(-2)j\varepsilon_0\omega\,jk_x(\omega)}\cdot e^{j[-k_x(\omega)\cdot x\cdot M\cdot N]} \quad (21)$$

$$E_z(\omega,x,y,z)=\sum_m\sum_n\frac{\left[(-jn\pi/b)(jP_+)-\varepsilon_0\mu_0\omega^2\sin\alpha\right]C_mC_n}{2j\varepsilon_0\omega\,jk_x(\omega)}\cdot e^{j[k_x(\omega)\cdot x\cdot M\cdot N]}+$$
$$+\sum_m\sum_n\frac{\left[(-n\pi/b)P_+-\varepsilon_0\mu_0\omega^2\sin\alpha\right]C_mC_n}{(-2)j\varepsilon_0\omega\,jk_x(\omega)}\cdot e^{j[-k_x(\omega)\cdot x\cdot M\cdot N]}$$

where

$$P_+=m\frac{\pi}{a}\cos\alpha+n\frac{\pi}{b}\sin\alpha\qquad P_-=-m\frac{\pi}{a}\cos\alpha+n\frac{\pi}{b}\sin\alpha\qquad M=m\frac{\pi}{a}y\qquad N=n\frac{\pi}{b}z$$

$$k_x(\omega)=\sqrt{\varepsilon_0\mu_0\omega^2-\left(m\frac{\pi}{a}\right)^2-\left(n\frac{\pi}{b}\right)^2}.$$

It can be seen, that one term in the field components propagates forward, the other propagates backward, considering the location of the excitation as a starting point ($x = 0$) in the assumed infinitely long wave-guide.

The limiting wave-length (and the limiting frequency, respectively) can be obtained from (20) and (21) as

$$\varepsilon_0\mu_0\omega^2-\left(m\frac{\pi}{a}\right)^2-\left(n\frac{\pi}{b}\right)^2=0\qquad\lambda_{m,n}=\frac{2ab}{\sqrt{(mb)^2+(na)^2}} \quad (22)$$

By the application of the formal inverse Fourier-transformation, the complete time-space dependent, exact form of the propagating electric and magnetic field-components can be obtained as

$$H_x(t,x,y,z)=\frac{1}{4\pi}\int_{-\infty}^{\infty}\left\{\sum_m\sum_n\frac{P_-C_mC_n}{k_x(\omega)}\cdot e^{jT_+}\cdot e^{jM}\cdot e^{jN}-\sum_m\sum_n\frac{P_-C_mC_n}{k_x(\omega)}\cdot e^{jT_-}\cdot e^{jM}\cdot e^{jN}\right\}d\omega$$

$$H_y(t,x,y,z)=\frac{1}{4\pi}\int_{-\infty}^{\infty}\left\{\sum_m\sum_n C_mC_n\sin\alpha\cdot e^{jT_+}\cdot e^{jM}\cdot e^{jN}+\sum_m\sum_n C_mC_n\sin\alpha\cdot e^{jT_-}\cdot e^{jM}\cdot e^{jN}\right\}d\omega$$

$$H_z(t,x,y,z)=\frac{-1}{4\pi}\int_{-\infty}^{\infty}\left\{\sum_m\sum_n C_mC_n\cos\alpha\cdot e^{jT_+}\cdot e^{jM}\cdot e^{jN}+\sum_m\sum_n C_mC_n\cos\alpha\cdot e^{jT_-}\cdot e^{jM}\cdot e^{jN}\right\}d\omega$$

$$E_x(t,x,y,z)=\frac{-1}{4\pi}\int_{-\infty}^{\infty}\left\{\sum_m\sum_n\frac{P_+C_mC_n}{\omega\varepsilon_0}\cdot e^{jT_+}\cdot e^{jM}\cdot e^{jN}+\sum_m\sum_n\frac{P_+C_mC_n}{\omega\varepsilon_0}\cdot e^{jT_-}\cdot e^{jM}\cdot e^{jN}\right\}d\omega$$

$$E_y(t,x,y,z)=\frac{1}{4\pi}\int_{-\infty}^{\infty}\left\{\sum_m\sum_n\frac{\left[(-m\pi/a)P_+-\varepsilon_0\mu_0\omega^2\cos\alpha\right]C_mC_n}{\varepsilon_0\omega\,k_x(\omega)}\cdot e^{jT_+}\cdot e^{jM}\cdot e^{jN}-\right.$$
$$\left.-\sum_m\sum_n\frac{\left[(-m\pi/a)P_+-\varepsilon_0\mu_0\omega^2\cos\alpha\right]C_mC_n}{\varepsilon_0\omega\,k_x(\omega)}\cdot e^{jT_-}\cdot e^{jM}\cdot e^{jN}\right\}d\omega \quad (23)$$

$$E_z(t,x,y,z)=\frac{1}{4\pi}\int_{-\infty}^{\infty}\left\{\sum_m\sum_n\frac{\left[(-n\pi/b)P_+-\varepsilon_0\mu_0\omega^2\sin\alpha\right]C_mC_n}{\varepsilon_0\omega\,k_x(\omega)}\cdot e^{jT_+}\cdot e^{jM}\cdot e^{jN}-\right.$$
$$\left.-\sum_m\sum_n\frac{\left[(-n\pi/b)P_+-\varepsilon_0\mu_0\omega^2\sin\alpha\right]C_mC_n}{\varepsilon_0\omega\,k_x(\omega)}\cdot e^{jT_-}\cdot e^{jM}\cdot e^{jN}\right\}d\omega \qquad (23)$$

where

$$T_+=\omega t+k_x(\omega)\cdot x \qquad T_-=\omega t-k_x(\omega)\cdot x$$

Numerical results

As a first application of the obtained closed-formed solution, let an Earth-ionosphere wave-guide model be calculated.

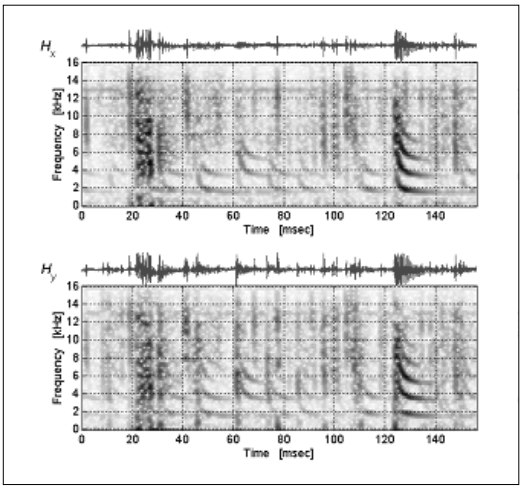
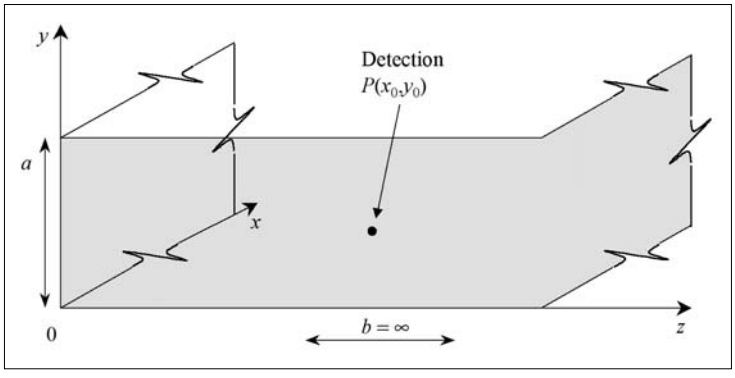


Fig.3. Observed dynamic spectra and time functions for two magnetic field-components (Marion Island, 2001.04.22. 02:30:07)

In the model, let $b = \infty$ be assumed (wave propagation between two infinite perfect conductor plain walls), the excitation is in the $x = 0$ plane. It is shown in *Fig.4.*

Fig.3. shows observed dynamic spectra detected on the terrestrial surface, propagated in the Earth-ionosphere layer. Many papers investigate the propagation of these signals, and it is known that the form of the dynamic spectrum is caused by the presence of the Earth-ionosphere waveguide [4, 5, 6, and 7].

Fig.4/a.



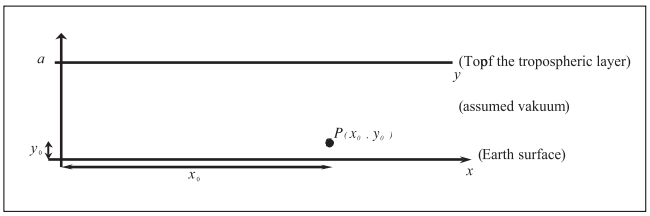


Fig.4/b. The model of the Earth-ionosphere wave-guide

But the theoretical description of the problem is based on the monochromatic wave-propagation of guided waves [8].

In the model-calculation the height of the bottom of the ionosphere is at 85 km, 2000 km is the traveled propagation path, 1000 m is the height of the antenna (location of detection). The excitation is a Dirac and all the Fourier coefficients are taken into consideration with 1 value, $\alpha = 45°$, and $m = 0,\ldots,5$ modes are taken into consideration (but all these parameters are flexible and modifiable).

The calculated time functions and dynamic spectra of H_x and H_y at the P point is shown in *Fig.5.*

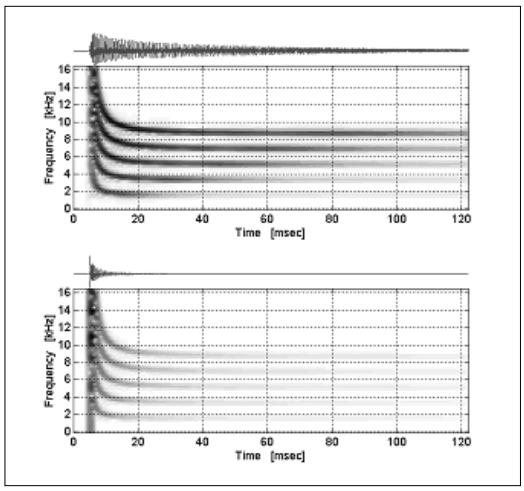


Fig.5. Calculated time functions and dynamic spectra of field components for Dirac excitation

In the comparison of Fig.3. and 5. some similarities are recognizable. The real impulse behavior can well explain the parallel traces in the spectra. The number of the branches depends on the number of Fourier-coefficients (modes) taken into consideration. Moreover, the distance among the branches depends on the height of the waveguide, or with other words, the thickness of the troposphere, the actual height of the bottom of the ionosphere (the height of the D layer).

Conclusions

In this paper a closed-formed solution is presented for guided waves excited by real transient signals.

The presented solving method is general and fully analytical. The excitation is a real arbitrarily shaped transient signal. Applying the presented solving method for Dirac excitation, the transfer function of the waveguide can be described. This result opens the way for investigation of real (short) impulse propagation in waveguides and the transient propagation phenomena. The geometrical structure can be developed further.

The closed-formed solution is well applicable for computers in order to calculate numerical results of the analytical solutions.

As a possible application of the new result in a special case, a geophysical example was presented. From the comparison of Fig.3. and Fig.5. it can be seen, that the model describes well some observed phenomena. By the application of this solution it became possible to monitor the bottom of the ionosphere continuously. It is not necessary to use the averaging of many observed spherics for the estimation of the height of the D layer, because the more exact new model makes it possible to use each individual observed spheric separately.

Moreover, the distance of the lightning source can be estimated from the dispersion fitting of the measured and calculated signals. The direction of incidence is well determinable from the ratio of H_x and H_y components.

Acknowledgement

This work was carried out by the aid of the contracts no. OTKA T034831, T037611 and F037603; the Bolyai János Research Grant of the Hungarian Scientific Academy; DAK-11/02 Hungarian-South African Intergovernmental Science and Technology Cooperation Program; and the Research & Development funds of the Hungarian Space Office.

References

[1] Ferencz O.E.:
Electromagnetic Wave Propagation in Different Terrestrial Atmospheric Models; Ph.D.Thesis, Budapest Univ. of Technology and Economics, 1999.

[2] Ferencz Cs., Ferencz O.E., Hamar D. and Lichtenberger J:
Whistler Phenomena, Short Impulse Propagation; Kluwer Academic Publishers, Astrophysics and Space Science Library, Dordrecht, 2001.

[3] Simonyi K.:
Foundation of Electrical Engineering, Pergamon Press, New York, 1963.

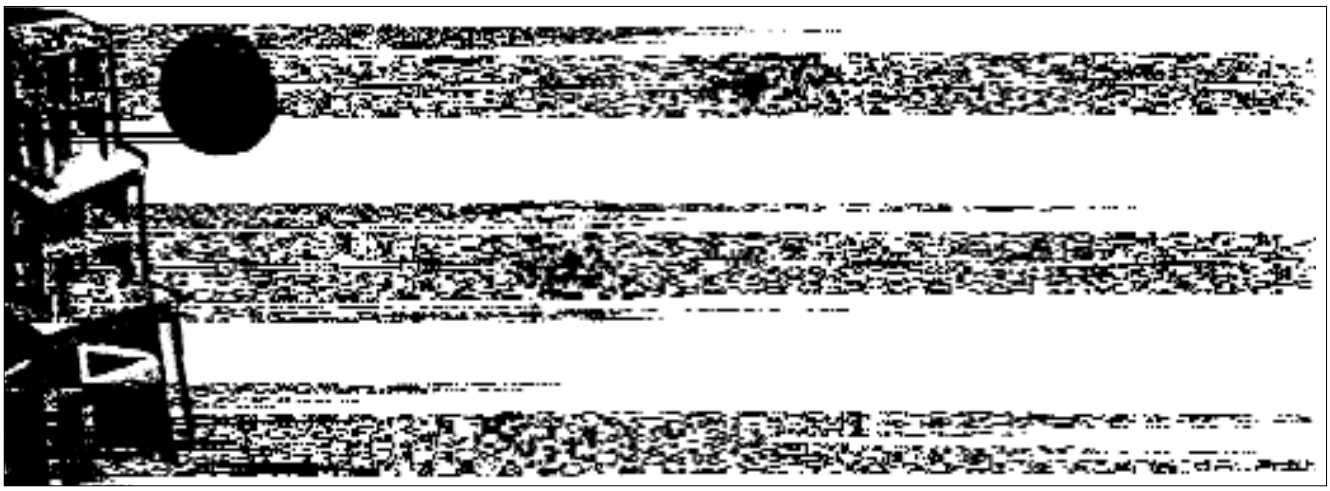
[4] Shvets A. V., M. Hayakawa:
Polarisation effects for tweek propagation, Journal of Atmospheric and Solar-terrestrial Physics, Vol. 60, No.4, pp.461–469, 1998.

[5] Shvets A. V., Lazebny B., V., Kukushkin A., S.:
Synchronous measurement of atmospherics on the sea surface and underwater, Journal of Atmospheric and Solar-terrestrial Physics, Vol. 16, No.3, pp.221–226, 1996.

[6] Hayakawa M., Ohta K., Baba K.:
Wave characteristics of tweek atmospherics deduced from the direction-finding measurement and theoretical interpretation, Journal of Geophysical Research, Vol. 99, No.D5, pp.10,733–10,743, 1994.

[7] Cummer S. A., Inan U. S., Bell T. F.:
Ionospheric D region remote sensing using VLF radio atmospherics, Radio Science, Vol. 33, No.6, pp.1781–1792, 1998.

[8] Budden K.G.:
Radio waves in the ionosphere; Cambridge University Press, London 1966.



Packet switched optical router

ATTILA KOVÁCS, ILDIKÓ DEME

University of Technology and Economics, Department of Broadband Infocommunication Systems
Laboratory of Optical and Microvawe Telecommunication

kovacs@mht.bme.hu, deme@mht.bme.hu

Reviewed

Keywords: WDM, optical packet switching, sub-carrier

Our laboratory is a participant of an international project where an optical core router development and research is made by university and industry members. The developed hardware based on wavelength division multiplex and subcarrier label technology. The concept, the stage of the development and the foregoing results will be demonstrated below very briefly.

Our laboratory is member of an European 3 years long project, called LABELS, where in optical backbone applicable core router will be developed. The participants are universities, manufactures and operators. Because of the research is not only theoretic but experimental also, the aim is to build and test a prototype. Our part is to realize the optical address processing unit.

The concept and the foregoing results will be detailed below.

Concept and specification

Relation to the design and specification of the device the next criterions were respected:

Because of the high speed WDM network application, the speed of the device was chosen to 10 Gbit/s per channel.

Other important respects were the testability and the cost effectively. Therefore minimum 8, maximum 16 channels routing are implemented.

Because of the IP network transparency and the available technology, O/E conversion and subcarrier label were used [2,3]. The various packet lengths take for the high speed 300 Mbit/s optical label. At this speed, in burst mode, the router device has to finish the label recovery and routing process along 1-2 µs. This time is given by the next calculation:

Because of the IP network transparency, we can calculate with 1500 byte long packets in average. Until the label processing the data packets are delayed by an optical delay line. This delay line is implemented by a few 100 m long optical fiber. The propagation speed in a fiber is n_1/c, where *c* is the velocity in vacuum ($3*10^8$ *m/s*), and n_1 is the refraction index of the fiber core. In the monomode fiber n_1=1,45. If the fiber length is L=400 m, the delay time is T=L n1/c=1.93 µs. The delay line length must be so calibrated, that the longest packet must be shorter in time than the delay line.

The label structure will be detailed in the next chapter. The specification has to involve the optical filters and the photodetector parameter, and the collision detection's method. The optical filter is a critical part of the system because of the required narrow band operation. The applied filter should be able to separate the subcarrier. While the distance in spectra between the baseband signal and the subcarrier is only 10 GHz, the optical filter must be very narrow band. The O/E conversion is made by an envelope detector. After the detection, the arisen baseband signal is processed by the label recovery circuit.

In case of collision the lower priority packet will be dropped by a collision detector. Based on the above detailed parameters the structure of the device and the label can define.

Optical address

In our case one channel is equal with one wavelength. The intensity modulated data speed is 10 Gbit/s. 20 GHz distance from the data in spectra can be found the NRZ (non return to zero) coded optical label information. The speed of the label is 300 Mbit/s. The optimal address length determination depends on 3 aspects:

• The address length should be enough to the synchronization because of the data recovery process. In case of open loop structure, based on the measurements 128 bits are needed for the safe synchronization. At the closed loop solution, this value is much more less. The preamble length can be reduced for 2-3 bytes. The drawback of the closed loop structure is the usage of the PLL. Because of this drawback, in our case the open loop recovery was used.

• The duration of the effective label information and the preamble together must be less than the 45% of the process time. If the duration of the 1500 byte length packet is 1,2 µs, and 100 ns time-protection is used for the elimination of the packet and the label overlapping, the available label recovery time is 450 ns. In this case 128 bits optical address can apply. For the routing process 600 ns additional time is available.

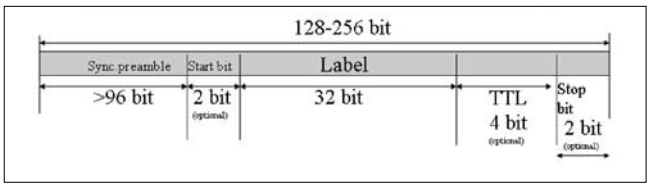


Fig 1. Optical address

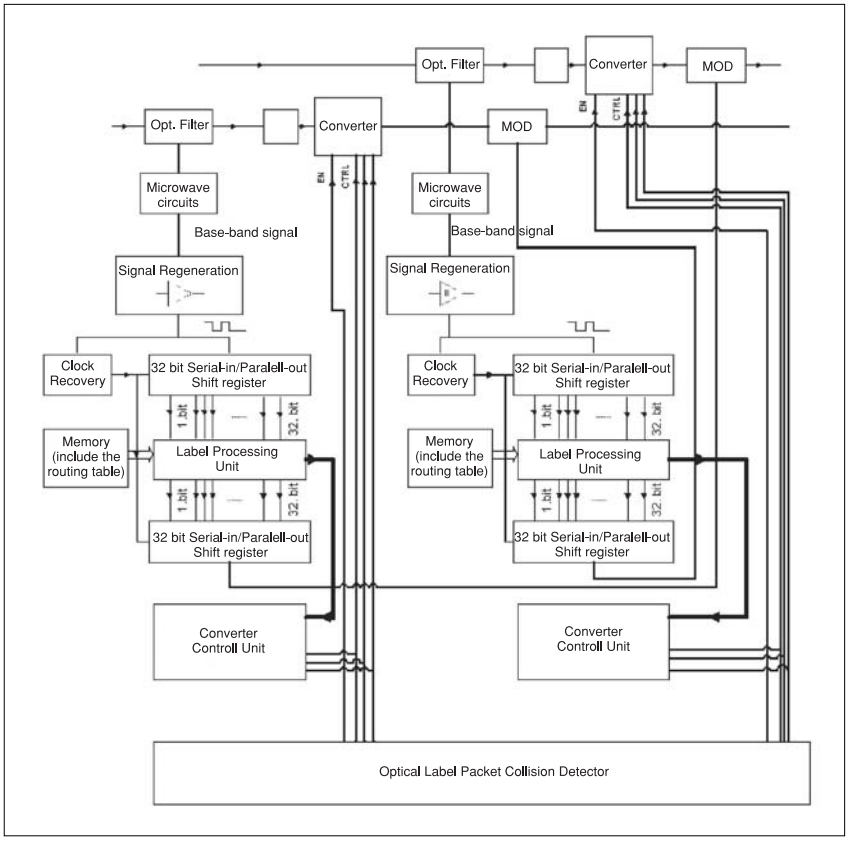
• At last the label contains start unique words to identify the beginning of the effective address information and TTL (Time to Live) bits to avoid the loop in the network. Because of the optical labels valid in the optical networks only, the 32 bits result adequate unique address *(Fig. 1)*.

Router structure

The developed router with two channels is illustrated by the *Fig 2*. At the input the incoming wavelengths are separated by a multiplexer to different channels. After the channel separation the optical label information is outfiltered. The data payload goes through an optical delay line in the same format.

During the delay time the optical label process must be done. The new optical label is generated, and the control logic sets up the wavelength converter. After the wavelength conversion (routing process) regenerated label information will be up-mixed and modulated on subcarrier of the packet carried on a new wavelength.

Fig 2. Two channel router block diagram



The chronologic order of the label process is the following:

The outfiltered optical label information is converted to a baseband electric signal by an envelop detector. The effective information will be recovered by using an open loop structure. Based on the effective label information and the static routing table the wavelength converter setup will be done. The optical label will be regenerated in the given format. The collision detection is based on packet priority.

The next chapter below details the open loop data recovery method.

Digital control logic

The open loop data recovery process made by an FPGA based (Field Programmable Gate Array) [1]. Because of the low cost, the high flexibility and the relatively high speed (in our case clkin = 420 MHz) the FPGA is able for network applications. The two main part of the control circuits are the user interface and the label processing unit.

The open loop data recovery based on an oversampling. The incoming bitframes are oversampled by a latch matrix. The elements of the D-latch matrix are divided by the clock and its phase shifted domains (90°, 180°, 270°). The oversampling structure is illustrated on the *Fig. 3*. The four phases results 4 time domain.

The right bit detection is made by a bit transition detector and time domain selection circuits. Because of the jitter the time domains are five latches long (Fig. 3). The bit transition detector selects the right time domain. The given data bits go through a serial/parallel transformation and after this are processed by a comparator. The beginning of the address information is signed by a special unique word. After the unique word matching the following 4 bytes are stored by a memory as the optical label information. The new wavelength is determined by the routing table. Finally the optical label is regenerated in the Fig. 1. which shows format, and added by an external microwave modulator to the subcarrier of the data packet.

Results and problems

At this moment the BER (Bit Error Ratio) measurement are on. To the BER measurement both transmitter and receiver circuits are needed. The optical address in predetermined format is generated by the transmitter circuit.

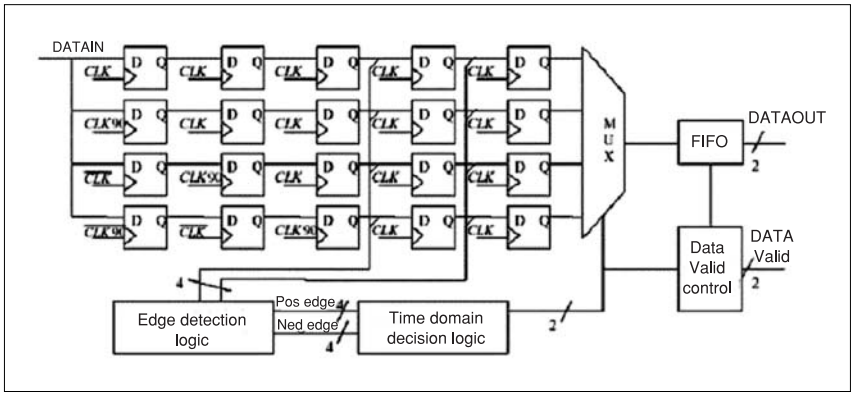


Fig 3. Data recovery

The transmitter includes random number generators (RNG). The random addresses and the random time between the packets are generated by the RNGs. The receiver recovers the burst mode label packets and stores the results. The BER is given by a comparator circuit. The asynchronous connection was simulated by using of different clock signal in both the transmitter and in the receiver. The clock frequency difference was a few MHz. The clock frequency of the transmitter was 200 MHz, and the receiver's clock frequency was tuned around 200 MHz ± 5%.

The synchronization time in function of the transmitter's clock ad the length of the preamble are illustrated on the *Fig 4.*

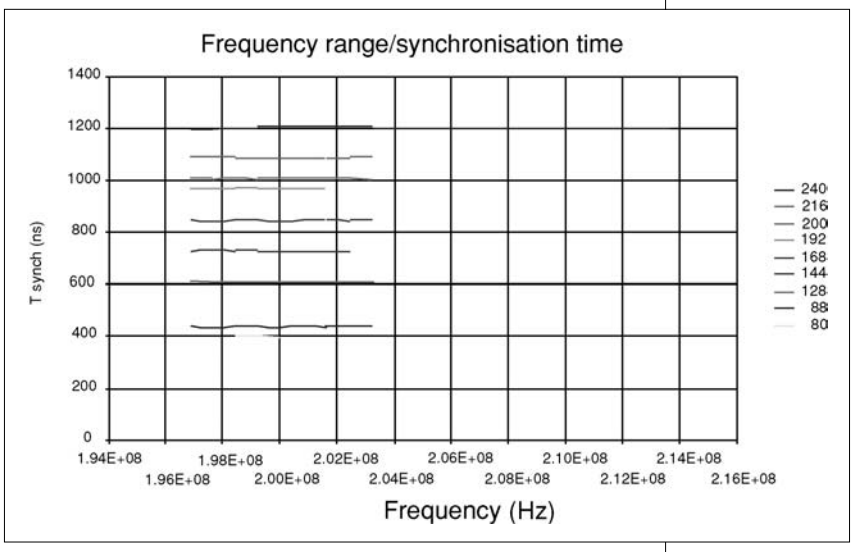


Fig 4. Measuring result

The minimal preamble length is 88 bit, and the minimal recovery time is 400 ns. The maximal possible frequency difference between the receiver and transmitter clock can be 2,2 MHz. It means 10% tolerance. This tolerance much more higher, than the requirements in the asynchronous networks.

At this moment our system fault are the maximum label recovery speed and the static routing.

Not faster than 200 MHz speed optical label can be recovered by the device.

In the future 300 MHz will have been achieved by the optimization of the digital circuits.

Only a static routing table can be used because of the available computing speed. The usage of dynamic routing table needs much more complex structure.

The developed device is capable only for wavelength routing. The physical routing is not supported. The physical routing can be achieved by using of 2 level structures, wherein the basic level is the wavelength router. In this case the wavelength routers are managed by an optical switch matrix.

Conclusion

We hope that we were able to give an adequate overview about our topical optical signal processing research.

We rely on that we find the solutions of our problems whereby other new questions and results could be finding up.

Believe the optical networks, it is the future!

References

[1] Attila Kovács, Ildikó Deme: "Clock and data recovery from high speed asynchronous NRZ coded data stream" Microcoll 2003, Budapest

[2] D. J. Blumenthal, A. Carena, L. Rau, V. Curri and S. Humphries: "All-Optical Label Swapping with Wavelength Conversion for WDM-IP Networks with Subcarrier Multiplexed Addressing," IEEE Photonics Technology Letters, Vol. 11, No.11, pp.1497–99, Nov. 1999.

[3] Tamás Marozsák, Attila Kovács, Ildikó Deme: "All-optical routing for packet switched networks" 3rd Hungarian WDM workshop, Budapest "Intelligence in Optical Networks", 8. April 2003.

Application and modeling of VCSELs in direct modulated optical links

TAMÁS MAROZSÁK

University of Technology and Economics, Department of Broadband Infocommunication Systems
tamas.marozsak@mht.bme.hu

Keywords: VCSEL, direct modulation, dynamic range, laser model

Vertical cavity surface emitting lasers, VCSELs, are very important light sources in optical communications. Their characteristics are very close to high performance edge emitting laser characteristics with low distortion, high modulation bandwidth and high dynamic range. At the same time their price can be an order less thanks to the new technology. This paper introduces the unique properties and application of these lasers in high speed direct modulated optical links. A novel circuit model is also shown, which is capable to simulate the spatial effects, like diffusion and spatial whole burning, in these lasers.

The semiconductor laser is one of the most important devices in today's communications systems. The optical carrier can transfer several GHz bandwidth digital signal or any kind of analog signals in this bandwidth. Even simultaneous transmission of baseband digital and radiofrequency analog signals can be applied [1]. The advantages of optical transmission, like small attenuation, high bandwidth and immunity to electromagnetic interference are evident, but the high price of optoelectronic devices still prevent this technology from being widely used. Fortunately, the ever increasing user demands and the fast development of new technologies transform the expensive technologies to everyday ones fast.

The VCSEL (Vertical Cavity Surface Emitting Laser) is an important milestone on this way, as it cuts the price of laserdiodes drastically and brings optical systems like Fiber To The Home (FTTH) into reality. On the next pages the main characteristics, some measurement results and modeling of these lasers are shown. The mathematical method used in modeling is general, and can be applied in circuit simulation of problems in other technical fields as well.

VCSEL structure

VCSELs carry the most important structural difference to conventional edge emitters in their names. The laser cavity is not a planar, but a vertical structure in the semiconductor wafer as shown in *Fig. 1.* As a consequence, this laser emits the light vertically to the surface. This is a very important feature in practice, because for testing the laser operation cutting of the semiconductor crystal is not needed as before. The individual lasers can be tested right after finishing the wafer because cutting of the semiconductor crystal is no longer needed for testing the laser. Cutting has no effect on the laser operation and therefore packaging or integration with other circuit elements is much easier.

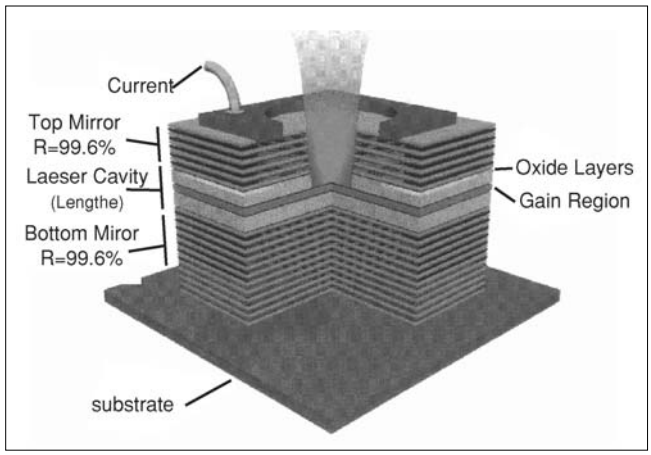


Fig. 1 Vertical structure of VCSELs

The structure of the vertical resonator is significantly different from the planar structure of edge emitters. The length of the resonator is defined by the thickness of some semiconductor layers grown on top of one other, and therefore it cannot be long: it falls into a few micron range. On such a small distance the round trip optical gain is small, therefore over 99% reflection is needed at the mirrors. Such a high value can only be made by distributed brag reflectors (DBR), which are formed by growing quarter wavelength thin layers having different refractive indeces on one other. To realize the required reflection, 20-30 layers are needed both at the bottom and the top. This layer structure is usually acchieved with molecular beam epitaxy (MBE). The active region between the mirrors usually consists of 1-2 quantum wells whose thickness is in the 10 nm range.

The number of layers in the mirrors is determined by the refractive index difference between the adjacent layers. The higher the difference, the smaller number of layers is needed. This is the reason why mostly 850-980 nm VCSELs are available. These lasers are made in the GaAlAs material system where sufficient refrac-

tive index difference can be achieved. This is not the case with InGaAsP, which is the material system for the 1300 és 1550 nm lasers.

There are some long wavelength VCSELs available, but this is still an intensive research area because the single mode fiber transmission systems require this wavelength. Therefore VCSELs are used mainly in multimode optical transmission systems where the dispersion limits the bandwidth seriously. The best wavelength optimized multimode fibers provide about 0.5 GHz·km, which is far less than the single mode fiber capacity. But it is inexpensive and many applications can work with this limited capacity. The optical bus can be mentioned as a good example, where several VCSELs are integrated into an array, therefore their light is easy to couple into a plastic fiber ribbon cable, and at the receiver side a Si photodiode array integrated with receiver circuits can detect the intensity at high speed.

Continuing the structure description, it is important to mention the transversal (parallel to the surface) properties. The cavity is a cylinder whose boundary is determined by the mirrors in the longitudinal direction but not in the transversal (radial) direction. In the beginning this boundary was created simply by disk shaped current injection as in the case of stripe lasers. The complex refractive index changes significantly where the current density is very high and this index difference from the surrounding material created the boundary and confined the optical field into the cavity. In this way only relatively large active diameter can be achieved, therefore an insulating oxide window is put into the structure, as shown also in Fig.1., to enhance the carrier confinement. In this layer a very small diameter hole can be realized, which means a very small resonator radius. The radial size of the cavity is important because it determines the number of possible transversal modes in the cavity. The intensity profile of the different transversal modes can be described by Bessel functions in a cylindrical coordinate system. The modes are denoted as $LP_{m,n}$, where the *m* and *n* indices refer to the number of periods in φ and *r* directions. *Fig. 2.* shows some of the lowest order modes in such a cavity.

The small radius and thickness of the active region causes very low threshold current compared to the edge emitting lasers. Typical values are less than 1mA and its variation is also significantly smaller. Another consequence of the small resonator length is that the longitudinal modes fall very far from each other in frequency and therefore only one of them can operate in the laser. Hence, the optical field has one longitudinal and more transversal modes in VCSELs.

This multiple transverse mode operation serves with some new interesting effects. The carrier density varies in space according to the optical field intensity. Where the intensity of a mode is high, the carrier density becomes locally low, which is called spatial hole burning (SHB). The dynamic evolution of these holes has fundamental effects on the VCSEL's static and dynamic characteristics. This will be explained better by simulation results later in the paper.

Single mode VCSEL can be fabricated by making a very small diameter oxide window. In this case the current density is very high in the cavity, which makes very high local heating. Because of this thermal problem, single mode VCSELs usually have small output power and therefore are not preferred. There are efforts to overcome this limitation, for example width fabricating mode selective mirror on the top of the laser. Recently 6mW output optical power was realized this way in a single mode device [2].

Modulation experiments

In short optical links the direct modulation of lasers gives the highest dynamic range [3]. In order to find this dynamic range, experimental investigations of laser diodes are needed. The bottom of the dynamic range is determined by noise and the upper limit is determined by nonlinear distortion.

The main source of noise in an optical link is the laser while the optical loss is not enough to attenuate it below the thermal noise in the receiver. This limit is generally 10dB optical loss in practice [3]. Therefore, in many cases the dynamic range is determined by the optical source itself.

The nonlinear distortion is not so important in digital transmission, but in analog transmission it has significant importance. The modulation depth cannot be increased over a limit in order to increase the signal to noise ratio because the second and third order intermodulation distortion (IMD) will degrade the performance. Cable television is a typical example where many channels are transmitted at the same time.

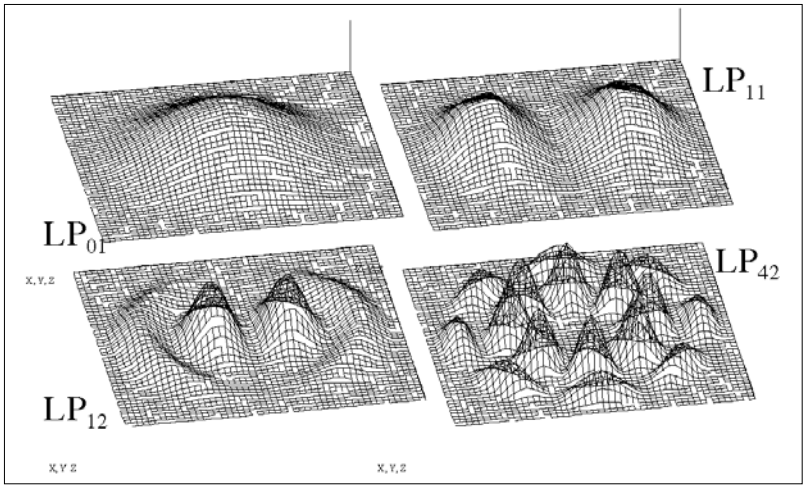


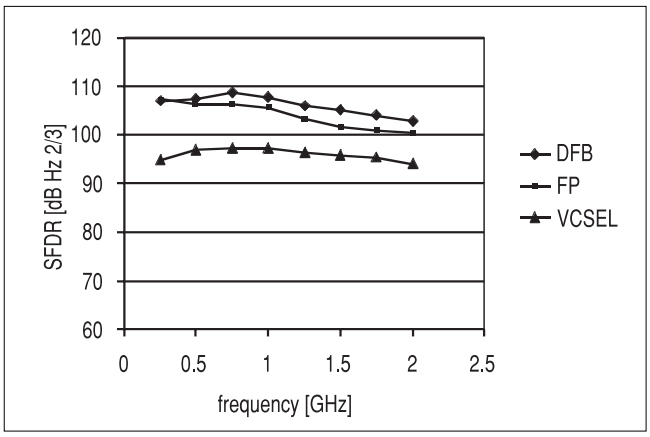
Fig. 2. Intensity mode profile of some fundamental modes in VCSELs

This is extremely sensitive for third order intermodulation because with increasing the power in the channels the third order IMD products increase according to the third power and rises over the noise level quickly. Therefore, a usual way to give the dynamic range of a device is to measure the difference between the output noise power and the signal power when the third order IMD product equals with the noise in a two-tone measurement. This parameter is called spurious free dynamic range, SFDR, and its unit is dB/Hz$^{3/2}$.

Such two-tone SFDR measurements were performed on conventional 1300 nm edge emitting lasers and on 850 nm VCSELs. The different operating wavelengths did not allow using the same photodiode. On 1300 nm single mode fiber, while on 850 nm multimode fiber was used. These differences were important because the same optical reflection level could not be ensured, which would be important to get consistent results, because the optical reflection affects the dynamic behavior of semiconductor lasers significantly. One of the edge emitting lasers was a Fabry-Perot, the other a DFB type, both fabricated on the same technology using multiple quantum wells.

The measurement results are shown in *Fig. 3.*, where the SFDR can be seen versus the modulating frequency. The main nonlinear effect in lasers is the relaxation oscillation, and approaching its characteristic frequency both the noise and the nonlinear distortion increases. Therefore we got decreasing dynamic range with increasing frequency. This indicates that high relaxation oscillation frequency can be important even in lower frequency applications.

Fig. 3. Measured dynamic range of different lasers



The optical reflection degrades the modulation performance. Even low level reflection causes increase and frequency peaking in noise and IMD. *Fig. 4.* shows measured results on 1300 nm Fabry-Perot lasers. The situation with VCSELs is similar but because the operating wavelength is 850 nm, the solution is more difficult because isolators for that wavelength are not available. The measurement showed that approximately 40 dB reflection loss is needed to make the laser operate near to its maximum dynamic range.

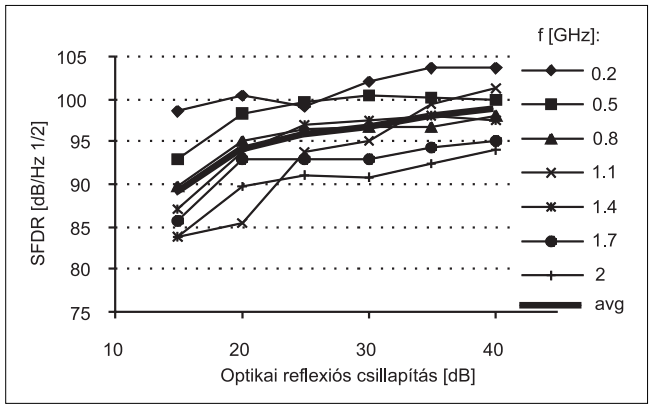


Fig. 4. Effect of optical reflection on the laser dynamic range

This could not be ensured in the VCSEL measurements in Fig. 3., therefore the real dynamic range should be higher than that shown in the diagram. Higher values are also reported in the literature [4]. In our case the dynamic range was smaller than the edge emitter's, but the measured 95 dB/Hz$^{3/2}$ value can be enough for most applications. For example the IEEE802.11b standard (wireless LAN) needs 94 dB at 2.5GHz [5], while a generic personal communication system, PCS, using $\pi/4$ DQPSK needs 72-83 dB/Hz$^{3/2}$ at 1900MHz [6].

Modeling of VCSELs

The accurate modeling of VCSELs based on the rate equations differs from the modeling of edge emitters. (In homogeneous medium the propagation of light can be derived by the Maxwell equations. In active medium containing free carriers the light effects are usually described by the rate equations. These can be derived from the Maxwell equations as well, but can be obtained by heuristics also by investigating the interactions between photons and free carriers. – *Editor's remark)*

Because these lasers operate in one longitudinal but more transversal modes, it is important to take the spatial processes into account. To simulate the spatial hole burning, it is essential to know the spatial distribution of the carrier density.

This means that in the rate equations the carrier density, in contrast with edge emitting lasers, can not be a scalar number but a function of space. The spatial distribution of the optical field is also different in the different modes as it was shown in Fig. 2. Therefore, the spatial dependence of the optical field also has to be taken into account. This can be done by writing separate rate equations for each modes using the modal gain theorem [7].

Finally, the following rate equations can be written:

$$\frac{dn_e(t,\underline{r})}{dt} = \frac{i(t,\underline{r})}{q} - \frac{n_e(t,\underline{r})}{\tau_e} - v_g \cdot g(t,\underline{r}) \cdot n_p(t,\underline{r}) \qquad (1)$$

$$\frac{dN_{p,i}}{dt} = v_g \cdot g_i \cdot N_{p,i} + \beta \cdot \frac{N_e}{\tau_e} - \frac{N_{p,i}}{\tau_p} \tag{2}$$

where n_e is the carrier density, t is time, $\underline{r}$ is the space vector, i is the pumping current, q is electron charge, τ_e is average electron lifetime, v_g is group velocity, g is optical gain, n_p is photon density, $N_{p,i}$ is the number of photons in the ith mode, g_i is the modal gain of the ith mode, β is the spontaneous emission factor, τ_p is average photon lifetime and N_e is the carrier number in the active region and can be expressed by n_e. $g(t,\underline{r})$ and g_i can be expressed by $n_e(t,\underline{r})$ and $n_p(t,\underline{r})$, and $n_p(t,\underline{r})$ can be obtained from the mode profiles and $N_{p,i}$. Index i means that we have as many photon rate equations as many optical modes are possible. This system of equations can be solved numerically after discretization of space. But it is computationally demanding, so difficult analyses like intermodulation takes very long time. Therefore I looked for a solution method, which does not has this drawback.

Circuit simulators solve systems of differential equations very efficiently. In addition it is very easy to choose between different types of analysis such as DC, AC transient or harmonic balance (for mixing and intermodulation) analysis. Therefore, I developed the circuit equivalent of these equations to simulate the main characteristics of VCSELs.

The main difficulty to overcome is that in circuit simulations only time or frequency can be the independent parameter, but in equations (1) and (2) both time and space are independent variables. The idea in eliminating this contradiction is that the optical intensity distribution of modes (the mode profiles) can be known a priori and they are not time dependent. The same can be done with the carrier density, it can be expressed as a linear combination of a limited series of orthogonal functions:

$$n_e(t,\underline{r}) = \sum_j n_{e,j} \cdot \Phi_j(\underline{r}) \tag{3}$$

where $\Phi_j(\underline{r})$ are given space functions and $n_{e,j}$ are their time dependent amplitudes. In this way the spatial dependence is represented by changing only the amplitudes of the different carrier density base functions. The space vector $\underline{r}$ can be changed to a simple radial coordinate, because the active region is very thin and therefore the carrier density does not change in vertical direction. The angular dependence can also be neglected for simplicity.

If the $\Phi_j(r)$ functions are orthogonal, then the electron equations (1) can be separated to j individual equations. For that (1) must be multiplied by each $\Phi_j(r)$ base function and integrated over the active volume. In the result only time varying amplitudes and space dependent integrals will remain. These integrals depend only on structural parameters, which are constants, and therefore these can be calculated in advance, and then any simulation can be carried out a number of times.

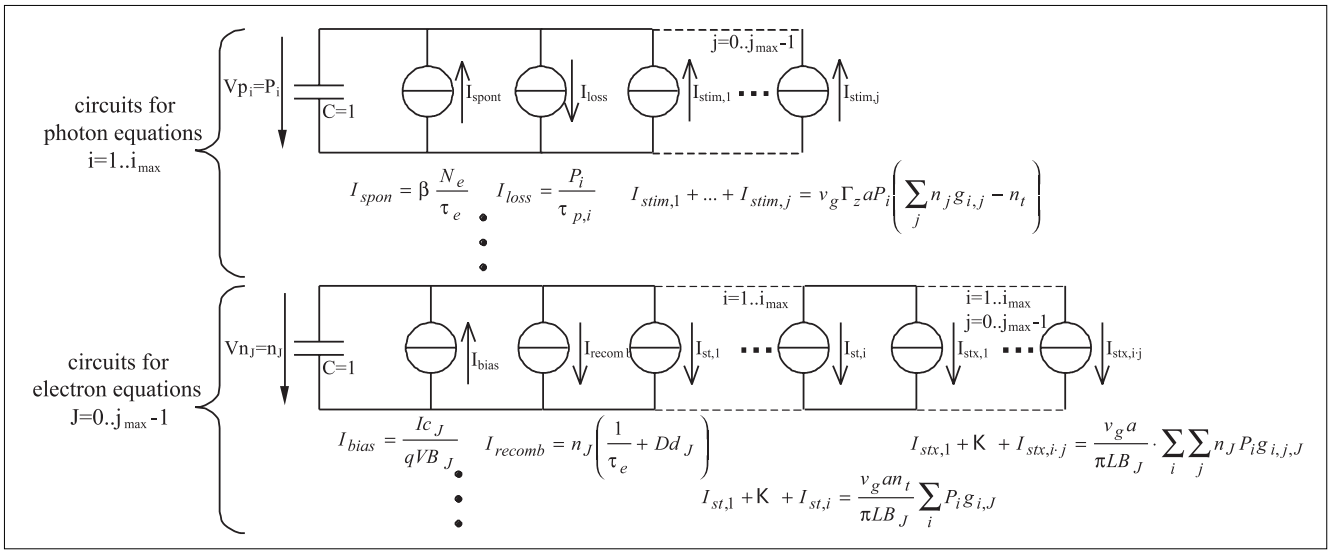
Doing all the derivations we obtain such a system of equations which consists of i photon and j electron equations, and mathematically independent of time:

$$\frac{dN_{p,i}}{dt} = v_g \Gamma_z a N_{p,i} \left(\sum_j n_{e,j} g_{ij} - n_t \right) + \beta \cdot \frac{N_e}{\tau_e} - \frac{N_{p,i}}{\tau_p} \tag{4}$$

$$\frac{dn_{e,J}}{dt} = \frac{I \cdot c_J}{q \cdot B_J} - \frac{n_{e,J}}{\tau_e} - \frac{v_g a}{\pi L B_J} \cdot \sum_i \sum_j n_{e,j} \cdot N_{p,i} \cdot g_{i,j,J} + \frac{v_g a \cdot n_t}{\pi L B_J} \sum_i N_{p,i} g_{i,J} \tag{5}$$

Here $g_{i,j}$ and $g_{i,j,J}$ constants are the results of spatial integrals which are called overlap integrals as they refer to the spatial overlap between the ith optical mode and jth electron distribution function. c_J are similar constants and B_J are normalizing constants. L is the length of the resonator, Γ_z is the longitudinal confinement factor, a is the gain constant (differential gain). The equations must be solved for $N_{p,i}$ and $n_{e,J}$, which are the time dependent amplitudes of the predefined distribution functions.

Fig. 5. Equivalent circuit for solving the multimode VCSEL rate equations



Carrier diffusion is also very important in VCSELs, therefore it has to be taken into account in the model. The electron rate equation, Eq. (1), must be extended by the term $+D\nabla^2 n(t,r,\varphi)$, where D is the diffusion constant, ∇ is the Laplace operator. If the $\Phi_j(r)$ carrier distribution functions are defined suitably, only the carrier rate equation will change by adding a $-n_{e,J}/(\gamma_J/R)^2$ term. γ_J and R are constants defined by the structure and material of the laser.

This idea to make the rate equations spatially independent can be found first in [8]. Later publications used this idea as well, but only for more simple cases and in many cases with not accurate mathematics.

The electrical equivalent of (4) and (5) can be derived as follows. The equations have the form similar to

$$C\frac{dV}{dt} = I_1 + I_2 + \ldots \text{ (assuming C=1)},$$

which is the Kirchoff equation for a circuit consisting of one capacitor and several current sources connected parallel. The photon number and the carrier density ($N_{p,i}$ and nJ in the equations) can be represented by the voltage on the capacitor, and the rates can be represented by the currents.

Using this idea, the equations can be modeled by the equivalent circuit network shown in *Fig. 5.* This electrical network consists of several subcircuits described above. There are i_{max} subcircuits for representing the i_{max} number of photon equations and j_{max} subcircuits representing the j_{max} number of electron equations. Each circuit contains one capacitor and a number of voltage controlled current sources equal to the number of terms on the right side of the equations. These terms are indicated in Fig. 5 below the generators. In these definitions the names of currents refer to the physical process that they represent. I_{stim}, I_{st} and I_{stx} mean different stimulated emission rate terms, the other names are obvious. In each subcircuit these currents are different as they are defined by constants having different i or J indices. In the expressions P_i and n_J become control voltages measured on the capacitors of the appropriate subcircuits (Vp_i and Vn_J). The input signal of the electrical network is given by I, which is common for the current sources defined by Ic_j/qdB_j in each electron subcircuit. In the simulation the network is solved for voltages Vp_i and Vn_J, which represent the number of photons, P_i, in the given mode and the electron density function amplitudes, n_J, respectively.

The method shown above for eliminating spatial dependence and creating the circuit equivalent was applied for VCSEL rate equations. But the principle is general and can be used to solve other problems as well.

The validity of the model and the circuit equivalent were checked by comparison to other models. The COST 268 action of the European Union had an open forum for comparing simulation results obtained from the same input parameters. The physical parameters of the laser for the following simulations were chosen from the modeling exercise of the action [9].

The simulations were done in APLAC, where using FOR statement arbitrary size ladder network can be defined in a few rows. The inputs of the simulation are material constants, parameters of the laser structure j_{max}. From these the $g_{i,j}$ and $g_{i,j,J}$, c_J és B_J constants are calculated, and then any type and any number of simulations can be carried out. In the following simulations two optical modes were supposed and j_{max}=11 was chosen giving rather accurate results.

Static characteristics, DC analysis

Fig. 6. shows the result of DC simulations in several bias points sweeping from 0 to 400 µA. It can be seen that lasing starts at I_{th} = 93 µA with the LP_{01} mode. As the optical intensity increases in this mode, a hole is created in the carrier distribution at r=0, as shown in *Fig. 7.* This affects the start of LP_{11} mode at I = 256 µA. Fig. 7 also shows that diffusion causes high electron density outside the r_i=3µm current aperture, which is also important in starting the LP_{11} mode.

Fig. 6. Result of DC simulation, bias current vs. optical power

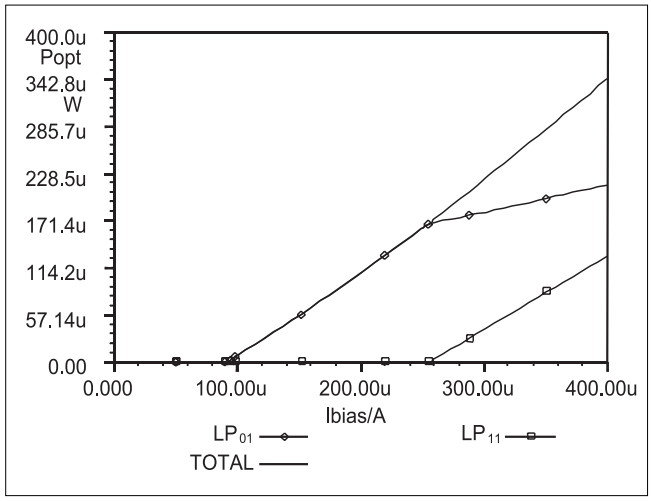
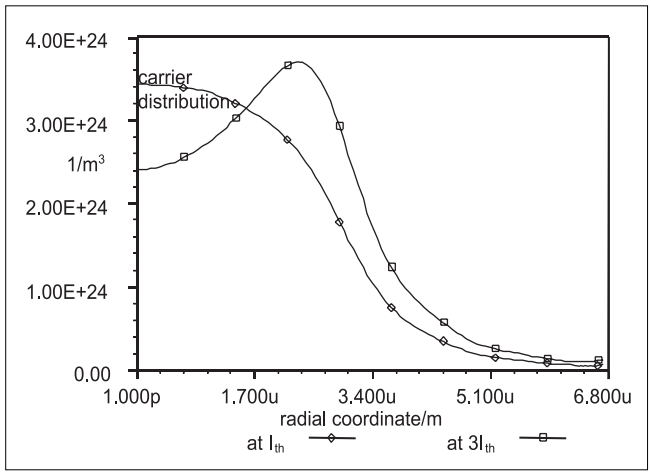


Fig. 7. Result of DC simulation, carrier distribution at I_{th} (LP_{01} starts) and $3I_{th}$ (LP_{11} starts)



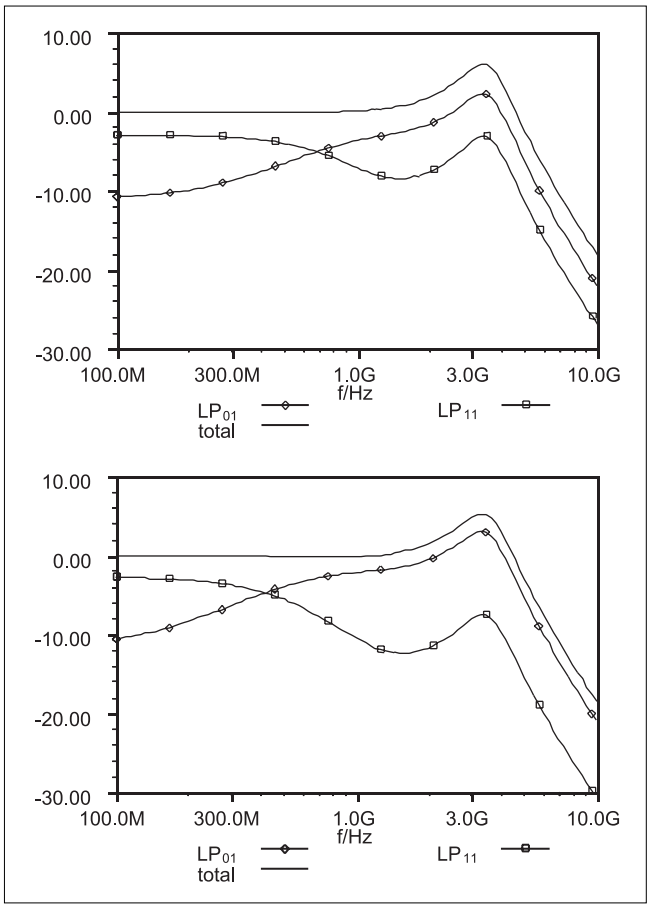


Fig.8. Results of AC simulation D=10·10^{-4} (up) and D=15·10^{-4} (bottom)

Small signal modulation characteristics, AC analysis

For the AC simulation such a bias point was chosen where both optical modes were lasing, I = 400 μA.

Fig. 8. shows the small signal modulation response for each mode and for the total intensity with two different diffusion constants. At low modulating frequency the LP_{11} mode has stronger modulation response than LP_{01}, which is a consequence of its higher slope efficiency shown by the DC simulation in Fig. 6. At high frequency the opposite becomes true, the LP_{01} mode has higher modulation response. The frequency of change depends on the diffusion constant, D, which determines how fast the diffusion can follow the modulation. In case of $D = 10 \cdot 10^{-4}$ m^2/s the characteristic frequency is around 700 MHz, while with $D = 15 \cdot 10^{-4}$ it is 450 MHz, which means that an increase in D by 3/2 times decreases the frequency to 2/3. What happens is that at higher modulation frequencies the LP_{11} mode can not gain from the diffusion process and its response to modulation decreases.

It means that the slope efficiency of the modes, which can be seen in the DC analysis, change with frequency. The total intensity does not show this change, because the carriers contribute to the stimulated emission anyway.

Distortion, harmonic balance analysis

The most interesting result of the paper is shown in *Fig. 9. and 10.*, which are the results of a single tone harmonic balance simulation. Up to the fifth harmonics of the modulating tone (at frequency f_m) was taken into account, but only the second and third harmonics are presented as most important in applications of direct modulated lasers. The amplitude of the modulating signal was $I_{bias}/10$, the bias point was $3I_{th}$. In directly modulated transmission systems the nonlinear behavior causes distortion and intermodulation. The generated second and third harmonic levels simulated here are in quantitative connection with the coefficients of the laser nonlinear characteristic and hence, have special importance.

The curves show that the laser nonlinearity increases toward the relaxation oscillation frequency in agreement with theory. It can also be seen that diffusion affects the higher order harmonics similar to the way it affects the fundamental one. The modulation response for f_m agrees with the result of the small signal (AC) analyses in Fig. 8.

Fig. 9. Harmonic balance simulation, harmonics level vs. modulating frequency

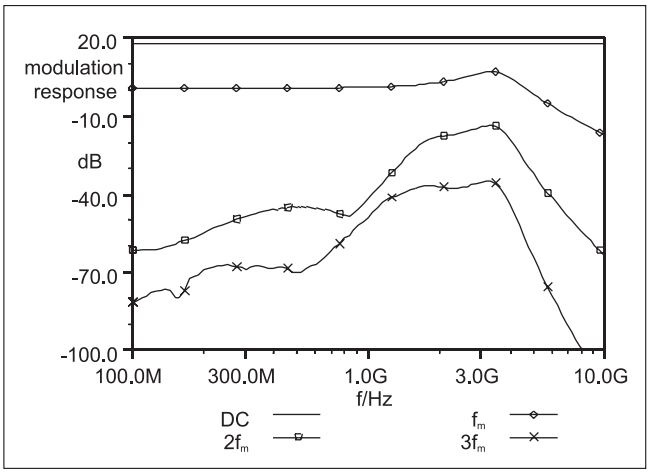
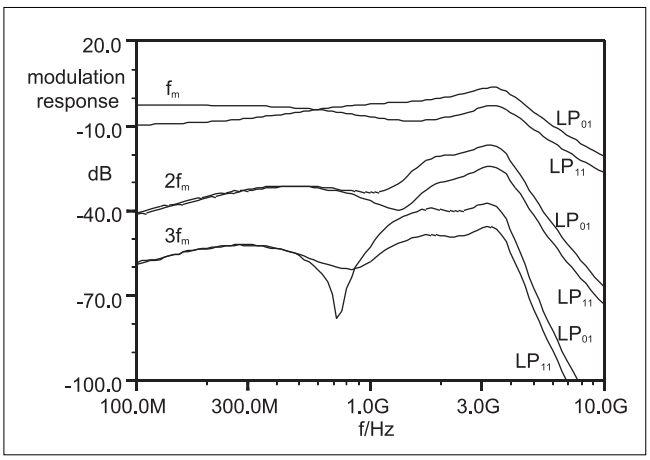


Fig. 10. Harmonic balance simulation, level of harmonics in different modes



As the modulating frequency becomes higher than approximately 700 Mhz, the LP_{11} response decreases in all higher order harmonics as well. This can be important in analog application of VCSELs because external reflectivity such as a coupling fiber or a photodetector surface can enhance one mode causing the modulation proprieties to change severely. These diffusion induced modes should be avoided in properly designed VCSELs. A low optical reflection environment is needed even when these modes exist with low optical power.

Interesting observation can be done by analyzing the numbers. The curves of the second harmonic ($2f_m$) in Fig.10. start from -40dB, while in Fig. 9, where the sum of the two modes can be seen, from -60dB. This means that the intensity of the two modes at the second harmonic are in antiphase.

Similar can be found at the third harmonic, and the opposite is true for the first harmonic (at the modulating frequency). This phenomena disappears above the characteristic frequency of diffusion, and the harmonic level, or distortion, increases. Between the two regions a deep notch can be seen in the third harmonic curve of LP_{01}, which is denoted to the antiphase behavior of the two main nonlinear processes, the SHB and the relaxation oscillation [10]. This effect is a bit different with the LP_{11} mode, and therefore it appears attenuated in the total optical intensity.

Conclusion

The basic properties and characteristics of VCSELs were introduced. The evaluation of measurement results showed that these lasers can have good performance even in demanding communication applications.

A novel equivalent circuit for modeling multiple transversal mode behavior in VCSELs was also presented. Using simulation results, the effect of spatial hole burning and carrier diffusion on the static and dynamic characteristics were shown. The different modes have different linear and nonlinear modulation properties which can cause problems with mode selective optical devices.

References

[1] Marozsak, T.; Udvary, E.; Berceli, T.:
A combined optical-wireless broadband Internet access: transmission challenges,
IEEE MTT-S International Microwave Symposium Digest,
pp.997–1000, May 2001

[2] Asa Haglund et al:
Single Fundamental Mode Output Power Exceeding 6mW in VCSELs with a Shallow Surface Relief,
will be published in
IEEE Photonics Technology Letters, Febr. 2004

[3] Tamás Marozsák, Attila Kovács, Eszter Udvary, Tibor Berceli:
Direct Modulated Lasers in Radio Over Fiber Applications,
MWP2002 International Topical Meeting on Microwave Photonics, Techn. Digest,
pp.129, Japan, 2002.

[4] Christina Carlsson et al:
Analog Modulation Properties of Oxide Confined VCSELs at Microwave Frequencies,
Journal of Lightwave Technology,
vol.20, Sept. 2002

[5] C. Faulkner:
RFIC Design Challenges for WLAN and 3G Systems,
Microwave Eng.,
pp.23–28, Jan/Feb. 2003

[6] J. C. Fan et. al.:
Dynamic Range Requirements for Microcellular Personal Communication Systems Using Analog Fiber-Optic Links,
IEEE Transactions on Microwave Theory and Techniques,
vol.45, pp.1390, 1997.

[7] L.A.Coldren, S.W.Corzine:
Diode Lasers and Photonic Integrated Circuits,
Wiley & Sons Publication, 1995.

[8] K. Moriki, H. Nakahara, T. Hattori, K. Iga :
Single transverese mode condition of surface-emitting injection lasers,
Electronic Communications Japan,
Part 2, vol.71, pp.81–90, 1988.

[9] http://www.ele.kth.se/COST268/WG1/DynVCSELTask/WGExcercise2.html

[10] J.S. Gustavsson et al:
Harmonic and Intermodulation Distrotion in Oxide-Confined Vertical Cavity Surface-Emitting Lasers,
IEEE Journal of Quantum El.,
vol.39, Aug. 2003

Wavelength converter solutions with semiconductor optical amplifiers

GÁBOR KOVÁCS

Budapest University of Technology and Economics,
Department of Broadband Infocommunications and Electromagnetic Theory,
Optical and Microwave Communications Laboratory,

gabor.kovacs@mht.bme.hu

Keywords: wavelength conversion, semiconductor optical amplifier (SOA), cross-gain modulation (XGM)

Semiconductor Optical Amplifiers (SOA) offer several possible solutions for wavelength conversion in the optical domain. In this paper different possible solutions are compared, pointing out the advantages and disadvantages of each. Then a simple and at the same time most promising method, the Cross-Gain Modulation (XGM) is investigated through performance measurement results.

Introduction

The main functions of optical telecommunication networks are the multiplexing and the routing of the information channels. The most widespread solution for optical networking is represented by the WDM (Wavelength Division Multiplexed) networks.

In these networks, channels can be separated and handled according to their wavelength, so that they can be transmitted through the same optical fiber. Connections can be simply point-to-point, but also complex network hierarchies can be built. In the second case it is essential, that for easy channel routing and to avoid wavelength collision at any point of the network, to make wavelength conversion possible at each node of the network, in order to transmit the same channel at different wavelengths on different links of the transmission route.

Nowadays the applied devices for wavelength conversion purposes are optoelectronic units. This means that the device converts the optical signal to the electrical domain – also regenerates it –, and retransmit it on an other wavelength. The advantages of this devices, that their technology is mature and they are available for trial use. But in optical networking aspect the long-term disadvantages of them is the opto-electronic conversion, which limits the possible operation speed (due to physical properties) and makes the device dependent on the signal coding scheme (not transparent for optical signals).

The main efforts in optical researches are therefore focused on *transparent* optical solutions. Though these solutions are mostly available only in laboratory environment, the possible performance of these devices according to the demonstrated results are really promising [1,2]. In the first part of this article we give an overview of the state-of-the-art of wavelength conversion using *semiconductor optical amplifiers* (SOA), than one of the possible methods is presented in details and investigated through measurement results.

Wavelength conversion methods

Wavelength converters should meet several requirements. They should operate at the 10 to 100 GHz range or even higher, and in order to build long-haul connections they should be cascadable. Further need is the low complexity of the device, high output power and fast tunability.

Optical gating methods

With this overall name we refer to those *all-optical* methods, where the new wavelength continuous wave (CW) signal is provided by an external source, and another device is applied to modulate the intensity of this signal according to the incoming optical information signal. The most typical solutions from this category are shown in the next subsection.

As it is widely known, in semiconductor optical amplifiers the key physical phenomenon is the population inversion, in other words the carrier density on the higher energy level in the active region. This carrier density is provided by external electrical pump. This population inversion gives the optical amplification in the amplifier due to stimulated emission. But through the manipulation of the carrier density by an optical signal (due to gain saturation process) it can also be applied for wavelength conversion.

The following methods use the device in this way.

Cross-gain modulation (XGM)

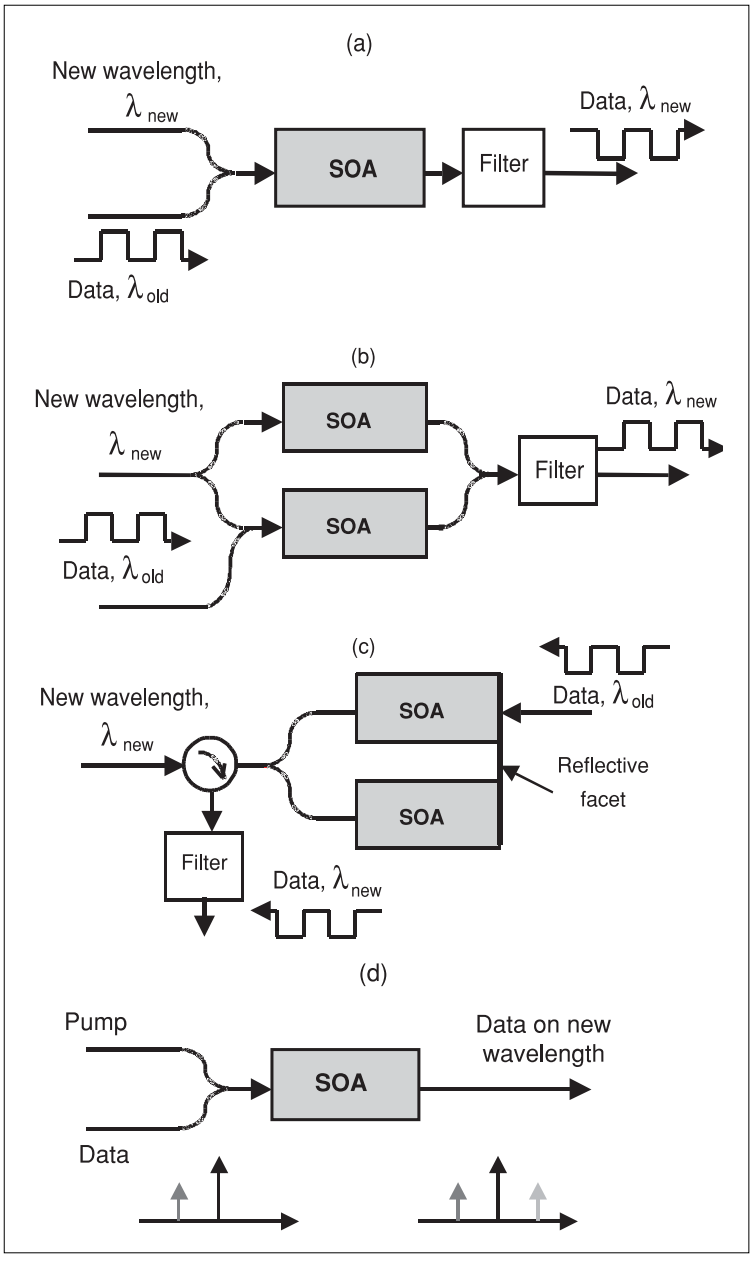
In this case the basis of the wavelength conversion is the gain saturation itself. The external pump source has a constant pump rate, and so the output power of the SOA is also limited. Therefore there is a limit in the incident optical power of the information signal, called the saturation power, until the SOA provides constant gain, above it the gain will significantly decrease. As the material is homogenous, the SOA will saturate at the whole operation wavelength spectrum, so that the

same gain variation will be for the CW signal at the new wavelength. In this way the CW signal will be modulated according to the input optical signal. The block scheme of the device is shown in *Fig. 1/a.*

The gain of the SOA is inversely proportional to the intensity of the incoming signal, thus it will work as an external modulator for the CW signal [3,4,5].

The advantages of this method are the simplicity of the device and the high transfer speed (approximately 100 GHz, limited by the carrier recovery time). But there are also some disadvantages. The input optical signal power should be high enough to saturate the SOA, a tuneable optical filter is needed in order to suppress the old wavelength, and the device inverts the signal levels (however the last two features can be avoided by special arrangements). And last but not least, this method is only suitable for conversion of intensity modulated signals, as the phase information is lost during the process. We investigate this method in details later in this article.

Fig. 1. Wavelenght converter schemes
a) XGM
b) XPM Mach-Zender interferometer
c) XPM Michelson interferometer
d) Wave mixing



Cross-Phase Modulation (XPM)

Placing two SOAs in the arms of an interferometer, it can also work as a wavelength converter. The CW signal at the new wavelength is transmitted into both arms, and the information signal to one of them. The intensity modulation of the information signal will lead to variation in the carrier density, and in the same time in the refractive index. The variation of the refractive index will result in variation in propagation time, which gives a phase difference between the two arms. This phase modulation can be converted to intensity modulation by an interferometer structure. Typically a Mach-Zender *(Fig. 1/b.)* or a Michaleson-interferometer *(Fig. 1/c.)* is used.

Compared to the XGM this interferometric scheme provides signal quality improvement. A stable operation can be achieved by integrating the interferometer to a single semiconductor chip, so that the two SOAs can operate at same conditions [6,7,8,9].

The possible operation speed is around 100 GHz, but 168 GHz has also been demonstrated.

Cross-polarisation Modulation (XPolM)

Wavelength conversion can be achieved through changing the polarisation of the CW signal at the new wavelength [10].

In this case we utilise a *polarisation beam splitter* (PBS), and its capability to filter optical signals according to their polarisation state. First the CW signal is transmitted through a SOA and adjusted to a PBS, so that the whole signal is transmitted through the system. Than the intensity modulated information wavelength is also transmitted into the SOA, and due to it birefringence the intensity variation will cause a polarisation rotation on the CW signal. This polarisation modulation will be converted into intensity modulation by the PBS.

The main advantage of this method is the high achievable extinction ratio, but the complexity of the device makes its application problematic.

Wave mixing

The earlier mentioned all-optical and transparent optical transmissions are not synonyms to each other. The gating methods are not transparent solution, as the only preserve the intensity information of the signal.

In this case we utilise the nonlinear behaviour of the SOAs, and the new wavelength signal is produced as the nonlinear combination of the input signals. Thus, if any of the input signals contains intensity, phase or polarisation modulation, it will be represented in the output signal. In this way, wave mixing is the only method with transparent optical properties, because all the optical properties of the information signal are unchanged during the conversion.

According to the number of the input signals there is *four-wave mixing* (FWM), *three-wave mixing* (TWM) and *difference frequency generation* (DFG), depending which order of nonlinearity is utilised.

In addition to transparency wave mixing is the only method which allows parallel conversion of more than one channel and capable for operating over 100 GHz. But on the other hand the optical nonlinear efficiency is very low, thus the implementation of this converters is complicated.

Investigation of XGM

After the general overview we investigate the most promising method of wavelength conversion, the cross-gain modulation. We studied XGM through simulation and measurements, and present the most important results in this article.

Saturation process

The most important process in semiconductor optical amplifiers in the cross-gain modulation aspect is the saturation.

As we mentioned earlier, population inversion is provided by an external electrical supply, which pumps carriers to the active region at constant rate. This carrier density is consumed by optical amplification through stimulated emission. When the input optical power is high than, to provide the optical gain, high output power is needed. Above a certain limit, the desired output power can not be provided by the SOA (because of the limited pump power), so the population inversion in the device and thus the gain will decrease.

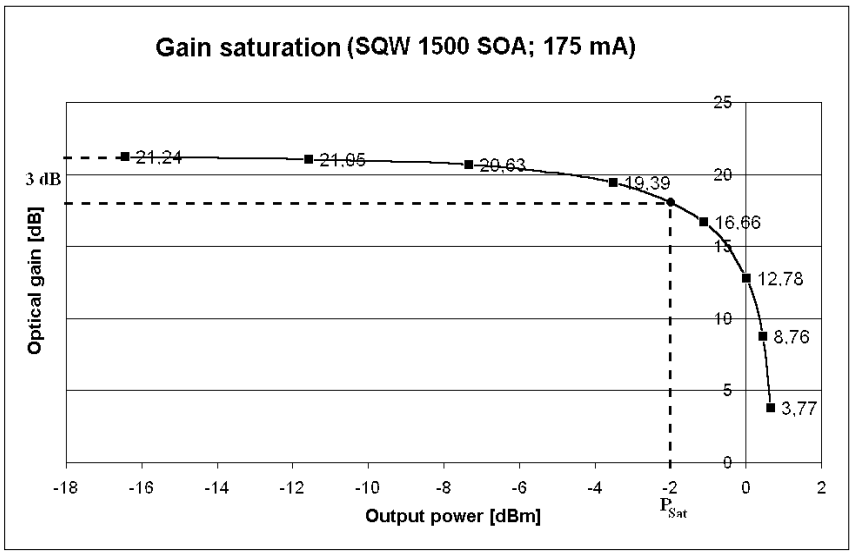
Plotting the optical gain versus the output optical power we get the saturation characteristic of the SOA, which is shown for the measured device in *Fig. 2.* It can be seen in the figure, that for a wide range of the optical power the SOA provides almost constant gain. The saturation power is defined by the 3 dB threshold, which is also shown in the figure. The maximum output power can also be estimated according to the figure.

As a normal amplifier constant gain for all intensities is preferred, for linearity reasons. But in our case it is desired to have big variation in the gain value due to the variation of the input optical signal, in order to have high extinction ratio. The low value of the saturation power is also advantageous, so that signals with low intensity can also be converted. This can be influenced by the material and the layer construction of the amplifier.

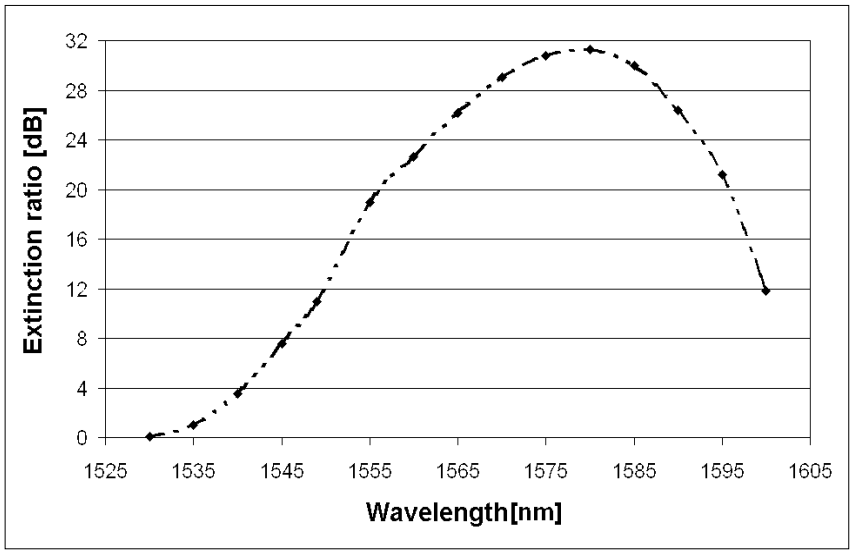
Extinction ratio measurements

Besides several interesting parameters *extinction ratio* is the most important one, as it highly determines the quality of the converted signal. Extinction ratio is defined by the quotient of the maximum and minimum

Fig. 2.
a) Gain saturation



b) Extinction ratio vs. wavelenght (SQW 1500 SOA, λ_{new}=1552 nm)



optical signal power, and it depends on the maximal achievable gain variation due to saturation. Measured extinction ratio versus the wavelength

The wavelength dependence clearly comes from the band structure, and the broadening of the energy levels in the semiconductor crystal. The measured device was designed to operate over the WDM wavelength range, and provides sufficient extinction ratio over the whole range. The maximal value for extinction was measured on the wavelength of 1582 nm and it was over 30 dB.

The operation speed is depends on the carrier recovery lifetime in the active region of the SOA. The typical value for this is around 1 ps, which allows operation up to 100 GHz. This, compared to the electrical devices, which are estimated to be limited at 40 GHz due to physical reasons, means that cross-gain modulation can be used for broadband telecommunication applications in the future.

Conclusions

Wavelength conversion using semiconductor optical amplifiers is one of the most intensively developing topic of optical telecommunication research. As it was shown, several approaches exist, with advantages and disadvantages. Up to the present matured, commercially available solutions are represented only by optoelectronic devices, but in long terms the increasing bandwidth requirements can be fulfilled only using all-optical, or rather optically transparent devices. Which method will be the winner? Device fabrication prices and market demands will decide.

References

[1] T. Durhuus, B. Mikkelsen, C. Joergensen, S.L. Danielsen, K.E. Stubkjaer: "All-optical wavelength conversion by semiconductor optical amplifier", IEEE/OSA J. Lightwave Technology, vol.14, pp.942–952, 1996.

[2] S. J. B. Yoo: "Wavelength conversion technologies for WDM network applications", IEEE/OSA J. Lightwave Technology, vol.14, pp.955–966, 1996.

[3] Derek Nesset, Tony Kelly, Dominique Marcenac: "All-Optical Wavelength Conversion Using SOA Nonlinearities", IEEE Communications Magazine, December 1998, pp.56–61.,

[4] A. Carena, M. D. Vaughn, R. Gaudino, M. Shell, Daniel J. Blumenthal: "OPERA: An Optical Packet Experimental Routing Architecture with Label Swapping Capability", JLT, vol.16, no.12, pp.2135–45, December 1998

[5] S. J. B. Yoo, Hyuek Jae Lee, Zhong Pan, Jing Cao, Yanda Zhang, Katsunari Okamoto, Shin Kamei: "Rapidly Switching All-Optical Packet Routing System With Optical-Label Swapping Incorporating Tunable Wavelength Conversion and a Uniform-Loss Cyclic Frequency AWGR", PTL, vol.14, no.8, pp.1211–13, 2002.

[6] R. G. Broeke, J. J. M. Binsma, M. van Geemert, F. Heinrichsdorff, T. van Dongen, J.H.C. van Zantvoort, E. Tangdiongga, H. de Waardt, X. J. M. Leijtens, Y. S. Oei, M. K. Smit: "An All-Optical Wavelength Converter in a Layer-Stack Suitable for Compact Photonic Integration", Proceedings Symposium IEEE/LEOS Benelux Chapter, Amsterdam, pp.95–98, 2002.

[7] J. Leuthold, G. Raybon, Y. Su, R. Essiambre, S. Cabot, J. Jaques, M. Kauer: "40 Gbit/s transmission and cascaded all-optical wavelength conversion over 1,000.000 km", El. Lett., vol.38, no.16, pp.890–892, August 2002

[8] Min-Yong Jeon, Dong Sung Lim, Hak Kyu Lee, Joon Tae Ahn, Do Il Chang, Kyong Hon Kim, Seung Beom Kang: "All-Optical Wavelength Conversion for 20-Gb/s RZ Format Data", Photonics Technology Letter, vol.12, no.11, IPTLEL (ISSN 1041–1135) pp.1528–30, November 2000

[9] J. Leuthold, C. H. Joyner, B. Mikkelsen, G. Raybon, J. L. Pleumeekers, B. I. Miller, K. Dreyer, C. A. Burrus: "100Gbit/s all-optical wavelength conversion with integrated SOA delayed-interference configuration", Electronics Letters, vol.36, no.13, pp.1129–30, June 2000

[10] Y. Liu, M. T. Hill, E. Tangdiongga, H. de Waardt, N. Calabretta, G. D. Khoe, H. J. S. Dorren: "Wavelength Conversion Using Nonlinear Polarization Rotation in a Single Semiconductor Optical Amplifier", IEEE Photonics Technology Letter, vol.15, no.1, January 2003

[11] J. P. R. Lacey et al.: "Four-channel polarization insensitive optically transparent wavelength converter," IEEE Photon. Tech. Lett., vol.9, no.10, pp.1355–57, 1997.

[12] D. D. Marcenac et al.: "40Gb/s transmission over 103 km of NDSF using polarization independent mid-span spectral inversion by four wave mixing in a semiconductor amplifier," Elect. Lett., vol.34, no.1, pp.100–101, 1998.

[13] R. G. Batchko, R. Roussev, M. H. Sher, L. R. Marshall: "All-Optical MgO:$LiNbO_3$ Wavelength Converter for Telecommunications",

[14] M. C. Cardakli, A. B. Sahin, O. H. Adamczyk, A. E. Willner, K. R. Parameswaran, M. M. Fejer: "Wavelength Conversion of Subcarrier Channels Using Difference Frequency Generation in a PPLN Waveguide", IEEE Photon. Tech. Lett., vol.14, no.9, pp.1327–29, September 2002

Crystalline silicon solar cells with selective emitter and the self-doping contact

EDVÁRD BÁLINT KUTHI

Hungarian Academy of Sciences, Research Institute for Technical Physics and Materials Sciences, Budapest University of Technology and Economics, Department of Electron Devices

kuthi@mfa.kfki.hu

Reviewed

Keywords: light-electricity conversion, solar cells, outlets, efficiency

One of the key questions of our world is the energy-supply of the people while protecting the environment suitable for life. One solution for this problem is the use of renewable energy sources. In Hungary the share of this energy amounts to only a few percents of the total energy-supply. While the using of the biomass and the geothermal springs are only near-term solutions, for long-term the utilizing of solar and wind energy has to be considered.

1. Introduction

The advantages of the solar energy are:

- In practice it is available without limits;
- Using this energy we can save fossile energy-sources for other needful applications;
- Environment-friendly, it does not contribute to global warming.

Photovoltaic systems convert directly the radiance of the Sun to electric power by solar cells, therefore they are very attractive energy sources, because they do not generate any pollution during operation. Their lifetime can be as long as 20-30 years and they need very little maintenance that consists of almost only keeping their surface clean.

For crystalline and multicrystalline silicon solar cells one research trend of the current is to prepare the front side contacts such way that optimizes the solar cell parameters. The main problem is the following. For the best solar cell parameters a low doped and shallow (so with high square resistivity) n-type silicon layer is necessary on the top of a p-type layer, altough on such layer we cannot prepare a good ohmic metal contact. For a good ohmic contact a highly doped and deeper (so with low sqare resistivity) n-type layer is required.

One solution for this problem is the use of selective emitter instead of homogeneous emitter. This metallization technology creates a highly doped layer under the contact, and a low-doped layer between the contacts, so both requirements are satisfied, and the efficiency of the solar cells increases as compared to the ones with homogeneous emitter.

Today for commercial crystalline and multicrystalline solar cells the manufacturers are using the screen printing technology for metallization, because it is simple and cheap. An important research trend is the improving and optimizing of this or other inexpensive technologies to make the production of solar cells more and more cheap. Hence it is worthwile to use this technology also for solar cell prototypes for research aimed as increasing the efficiency so that both research trends could achieve the aims after a successful cell improvement.

In this article after a short explaining of the basic terms in this section, we look on the theory of selective emitter and its realized technologies in the 2nd section, and in the 3rd section we examine the theory and realization of the self-doping metallization technology.

Basics [2,3]

Before the detailed exposition of the technologies of selective emitter we review the basic terms used in connection with solar cells (parameters, coatings, crystalline structure).

Parameters:

AM0, AM1 and AM1.5: the intensity of solar radiation in free space (Air Mass 0, AM0) is 1353 W/m^2 on the average with a special irradiance spectrum. Assuming one air mass at perpendicular incidence (AM1) a weaker irradiance power density reaches the surface of the Earth (925 W/m^2), because the molecules of the air filter the spectrum of the solar radiation. In case of a not perpendicular incidence, the air mass is bigger for the radiation, so the power density is less, e. g. at 45 degree incidence (AM1.5) it is about 844 W/m^2. However there is also an indirect (diffuse) radiation on the Earth because of the photons scattered by the molecules of the air. This increases the AM1.5 intensity to about 970 W/m^2, which is a good approximation for the usual terrestrial applications. For characterizations of solar cells the AM1.5 irradiance spectrum normalized to 1 kW/m^2 is used.

Short-circuit current density (J_{sc}): the short-circuit current of the illuminated solar cell (it belongs to V=0 V voltage) gives the maximum current in the device, because in practice it is equal to the current generated by the light (I_{sc}= -I_L, where I_L is the light generated current). Because the current depends on the surface

area, the current density is used instead of it ($J_{sc}=I_{sc}/A$, where A is the surface area). In addition it depends on the solar spectrum, the intensity of the light (the number of the incident photons), the efficiency of the current collection, and the optical parameters. In silicon solar cells the maximum available short circuit current density at AM1.5 spectrum is 46 mA/cm².

Open circuit voltage (V_{oc}): it is the maximum available voltage in the solar cells (it belongs to I=0 A current). Its value is:

$$V_{oc} = V_T \ln(1+\frac{I_L}{I_0}) \cong V_T \ln\frac{I_L}{I_0}$$

where V_T is the thermal voltage that is constant for a given temperature, I_0 is the saturation current. The open circuit voltage depends mainly on the saturation current of the device, because it can vary within several orders, and to a smaller degree it depends also on the light generated current (I_L). Hence I_0 depends on the recombination mechanism, we can say V_{oc} characterizes the recombination. The highest value that has ever been measured on silicon solar cells is 720 mV at one sun intensity, AM1.5 spectrum and room temperature.

Fill factor (FF): this is the ratio of the maximum delivered power ($P_m=V_mI_m$) and the product of the open circuit voltage and the short circuit current (the maximum available values) at a given illumination, so it characterizes the maximum power we can extract from a solar cell:

$$FF = \frac{V_m I_m}{V_{oc} I_{sc}}$$

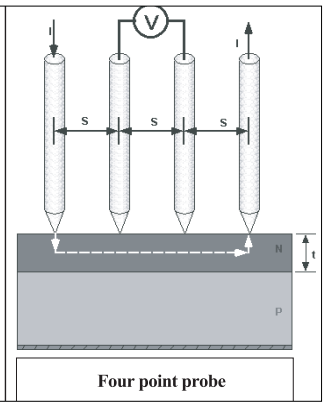
It can be calculated theoretically, but in fact it used to be a smaller number because of the parasitic resistances, so usually it is measured for a real solar cell. Its value depends mainly on the operating point, the open circuit voltage, and the recombination rate of the cell, so it depends on the the structural quality of the device. The greater the recombination rate, the less the FF and the V_{oc}. Because the V_{oc} does not change much in a material system it affects the FF to a small degree. The theoretical value of FF that belongs to the highest V_{oc} ever reached in a Si solar cell is about 0.85.

Efficiency: the ratio of the maximum delivered power and the power carried by the incoming photons: $\eta = P_m/P_{be}$. For a given material system, its value depends on the temperature of the solar cell, the solar spectrum and intensity of the incident light, and the technological and material properties of the solar cell structure. Hence for the measurement of the efficiency a standardized light source is used (with one sun intensity and normalized AM1.5 spectrum), and a constant room temperature (25°C) is ensured.

Sheet resistivity: for the top emitter layer of the solar cells, the resistivity as well as the thickness of the layer is often unknown or hard to be measured. However, a value known as the „sheet resistivity", which depends on both the resistivity and the thickness, can be easily measured using a „four point probe" for the top surface n-type layer. For a uniformly doped layer, the sheet resistivity is defined as: $\rho_{\square} = \rho/t$, where r is the resistivity of the layer, and t is the thickness of the layer. The sheet resistivity is increasing by the increasing of the resistance of the layer (so by the decreasing of the doping), and by the decreasing of the thickness of the layer.

The sheet resistivity of the top emitter layer is very easy to measure using a " four pont probe". A current is passed through the outer probes and induces a voltage in the inner probes. The junction between the n and p -type materials behaves as an insulating layer. The cell must be kept in the dark.
Using the voltage and current readings from the probe:

$$\rho_{\square}\left(\frac{\Omega}{\square}\right) = \frac{\pi}{\ln 2}\frac{V}{I}$$



Four point probe

Internal quantum efficiency (IQE): The quantum efficiency (Q.E.) is the ratio of the number of carriers generated by the solar cell to the number of photons of a given energy incident on the solar cell, as a function of wavelength or as energy. Internal quantum efficiency refers to the efficiency with which photons that are not reflected or transmitted out of the cell can generate collectable carriers. The QE curve caracterizes well the material parameters of the solar cells (surface recombination, bulk diffusion length, etc.)

Coatings:

Surface passivating layer: It is a transparent dielectric layer grown on the top surface of silicon, which has an appropriate material stucture for reducing the number of dangling silicon bonds and defects at the top surface, so the surface recombination velocity of the carriers decreases. The passivation effect can be increased further by bulk passivation using hydrogenation in the Si bulk to reduce bulk defects by hydrogen.

Anti-reflection coating (ARC): It is a transparent insulator layer with an appropriate thickness that increases the efficiency of the solar cell by reducing the photon reflection of the surface, optimized for a wavelength of 0.6 μm since it is close to the peak power density of the solar spectrum. Using a special material for it (e.g. SiO_2, SiN_x) this layer can also function as a passivating layer. The efficiency can be increased further by surface texturing that causes a mat surface. The increased area and the structured surface results in increased photon absorption and decreased reflection.

Crystalline structure:

Single crystalline Si (Cz-Si, FZ-Si): Solar cells need good quality and cheap bulk material with few crystalline defects so the minority carrier lifetime (and the diffusion length) can be fairly long to have a high probability for the generated carriers to reach the contacts.

Such quality single crystalline solar cell with silicon bulk material can reach an efficiency greater than 20%. In this case the whole Si wafer is a single crystal with the same crystalline orientation and overall structure. The manufacture of the silicon wafer needs high temperature, long time and good care, so its price is accordingly high. Cz-Si stands for Czochralski manufacturing process, and FZ for floating zone process. Cz-Si bulk material contains relatively high oxygen and carbon contamination levels that can decrease the bulk minority carrier lifetime, moreover the oxygen activates at high temperatures so this wafer is more sensitive to high temperature processing. The FZ-Si contains much less oxygen and carbon contamination levels so the minority carrier lifetime is of the order of millisecs. However this better quality bulk material is more expensive than Cz-Si.

Multicrystalline Si (mc-Si): The high cost of good quality single crystalline silicon materials were strong motivation to the utilizing of multicrystalline semiconductor bulk materials. In multicrystalline bulk material each crystalline grain has a crystalline structure, but each of their orientation is different. Other handicap of the multicrystalline material is the high defect density at the grain boundaries that decreases the minority carrier lifetime by increasing the local recombination rate, leading to a lower efficiency. On the other hand it is easier and cheaper to manufacture it. The handicaps decrease if the grain sizes are bigger than the diffusion length, so minimum a few millimeter. It is important that no grain boundary is allowable in parallel with the surface so the current does not have to flow through it. We can satisfy this requirement if the usual grain size is bigger than the thickness of the wafer.

2. Technology of the selective emitter

Theory [1,4,5,6,7]

The optimal emitter of the silicon solar cells with *homogeneous emitter* is a result of a compromise of the requirements of low dark current and low contact resistance. For passivated surface emitter the surface dopant concentration must be below 10^{20} cm^{-3} and the emitter must be shallow to keep the dark current (emitter saturation current density, J_0) low. However for a sufficient contact (where the surface thus has no passivation), especially for a screen-printed contact the surface dopant concentration must be higher than 10^{19}-10^{20} cm^{-3} for n-type silicon and 10^{17} cm^{-3} for p-type silicon, and the emitter must be relatively thick (>0.3 µm). Therefore the homogeneous emitter needs such high suface dopant concentration and emitter thickness (so 0.3-0.4 µm), but this makes the surface passivation almost impossible.

We must also note that for increasing dopant concentration the surface state (interface trap) density, N_{it} increases, so the recombination increases, too.

The overall series resistivity of the cell, including the effects of the interfaces and the electrode materials, has to be lower than 1 Ωcm^2, and the maximum of the contact resistance is 10mΩcm^2 in it by practice. On the basis of modelling and experiments for a non-screen-printed contact the optimal homogeneous emitter has a sheet resistivity of 60 Ω/□, a doping of $2x10^{19}$ cm^{-3}, and a depth of 1,4 mm. For a screen-printed Ag contact the sheet resistivity of the emitter has to be lower than 45 Ω/□ for an acceptable contact resistivity and a FF of not less than 70%. However such a highly doped homogeneous emitter decreases the response to short wavelength photons because of the high doping effects and the high recombination velocity that cannot be decreased sufficiently even by surface passivation.

The *selective emitter*, which has a high doping just under the metallization, can afford sheet resistivities even 70-200 Ω/□ between the contacts. Therefore the high doping effect in the emitter decreases, and good ohmic contact and high FF can be realized with low dark current and good carrier collection yield. In this case the surface of the solar cell can be well passivated between the metal contacts, and the emitter under the metallization is less critical for firing through by the contact. Note that for the same emitter sheet resistivity the deeper and less doped emitter causes less loss in the photogenerated current density than the shallower and higher doped emitter.

On the basis of a theoretical model [6] the emitter surface doping concentration under the contacts (where the surface recombination was considered as infinite) is optimally 10^{20} cm^{-3}, and its depth is 3-10 µm, and the doping profile has to be decreased persistently inside the substrate. However for the illuminated areas the surface doping concentration is ideally 10^{18}-10^{19} cm^{-3} constantly until a depth of 0,2-0,4 µm, where it has to be decreased abruptly.

The metallization of the silicon solar cells is a critical factor for the set of the overall electric and physical properties. The metallization can be a thin layer film (e.g. sputtering, vaporization) or thick layer (e.g. screen printing, stencil printing). The thick layer metallization technology is cheaper so it is often used for commercial and research manufacturing.

The thick layer metallization for the crystalline solar cells is used for realize the electric contacts and to assist the photocurrent collection. Advantages are the high photocurrent density, the low contact resistivity, relatively low line resistance, relatively good line resolution (50-100µm linewidth can be achieved), good solderability, adhesion and chemical strength, as well as the high productivity through the simple, quick and cheap technology. However the disadvantages compared to other techniques are the higher line resistance of the contacts caused by lower aspect ratio and higher resistivity of the paste material, and the higher contact resistance. The contact pattern is printed in one sweep of the squeegee, no vacuum or photolithography required for the metallization. The source of the problems is

the paste firing as its temperature is usualy in the range of 400-1000°C, and it is important that the contact metal must not penetrate the diffused layer too deeply. We can ensure this by firing the paste with the lowest possible temperature and duration. Hence nowadays the rapid thermal processing (RTP) technology is often used.

The *interdigitated back contact (IBC) cell structure* is simple and cheap, however it gives high effeciency.

Both of its positive and negative electrodes are on the back side of the cell, so on the illuminated side the contact metal does not overshadow the cell. On the front side only a very shallow and low doped homogeneous diffusion (for n-type substrate the optimal surface phosphorous doping is 10^{18} cm^{-3}) ensures the better surface properties. However on the back side there are also the n-type and p-type contacts, so in this case the local doping and hence the selective emitter is inevitable.

Realizations

Today in the industrial manufacturing of crystalline solar cells with *homogeneous emitter* conveyor belt systems are often used for the emitter diffusion (cleaning, screen printing, drying, annealing furnace etc.) [8]. In this process flow the diffusion source on the textured side of the p-type Si substrate is typically a sprayed or spinned on liquid containing phosphorous, or a screen-printed phosphorous paste. In the following drying step all solvents and organic components are evaporated or burned out of the doping source material. The wafers are further transported on the belt into a multizone furnace, where the appropriate emitter is made in temperatures typically above 900°C. Here the phosphorous diffuse into the substrate from the doping source. After that a passivating and antireflection layer is deposited on the surface followed by the screen printing of the metal contacts, and a subsequent drying and firing process (in 900°C). Lately instead of furnaces a conveyor belt system rapid thermal annealer is used, where the diffusion can be decreased to seconds.

The selective emitter structure was created in many ways. These technologies basically have two types: technologies that need scheme aligning, and self-aligning technologies. The scheme aligning process is more difficult, so it is more expensive. In case of aligning metallization to the underlying diffusion the misaligning is critical because if the metallization overlies from the highly doped area to the weakly doped area, the dark current of the device will increase and the contact resistance will be worse. It can also happen in this case that the metal get into until the p-n junction so it can short-circuit the whole cell. To avoid this the scheme of the highly doped diffusion has to be larger with the scale of misaligning error (for screen printing it is 1-2 µm, for photolitography it is much less). However in this case the dark current increases, and a dead layer formes where the highly doped area overlies the metallization. Furthermore in case of the laser overdoping technology (see IV.) the manufacturing cost is increased by the longer use of laser. In case of the etch-back technology (see II.) a dead layer also formes when the etch protective layer must be screen-printed aligned to the metallization. However in this item the most significant processes are grouped by the manufacturing techniques. *Table* shows us the published properties of the mentioned selective emitter solar cells.

I.) The double diffusion technique with *photolitography* is used for high efficiency cells, because you can reach the finest scheme resolutions with it, however it is the most expensive solution. The other techniqes (II.-V.) try to make this original technique cheaper, simpler and however able to produce still good quality selective emitter solar cells.

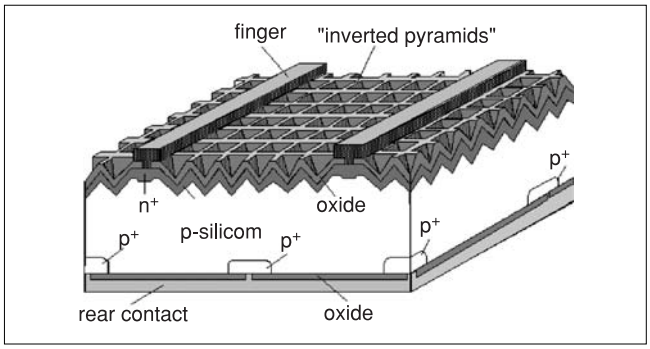
a) In the original technique an oxide mask layer is applied on the surface. This oxide is removed by photolitography where the contact areas are designed (applying the emulsion layer, drying, exposing through a photolitography mask, developing, firing, oxide etching on the opened areas, etching the emulsion from the remained oxide) followed by the forming of the highly

#	Cr.	Aera	Text.	H2	passiv.	ARC	Bulk res.	N. f. emit.	Contact	Efficiency
1	FZ	4 cm^2	yes	no	SiO_2	SiO_2	n/a	n/a	sput. or evap.	22,8 %
2	FZ	4 cm^2	yes	no	SiO_2	SiO_2	n/a	n/a	sput. or evap.	24,4 %
3	mc	1 cm^2	yes	n. a.	SiO_2	SiO_2	n/a	n/a	sput. or evap.	19,8 %
4	c	n. a.	no	no	SiN_2	SiN_x	n/a	n/a	screen-printed	13,2 %
5	Cz	100 cm^2	yes	no	yes	nincs	1 Wcm	70 Ω/□	sp. or e.buried	12,8 %
6	FZ	n. a.	no	no	SiO_2	SiN_x	1 Wcm	80 Ω/□	sp. or e.buried	14,26 %
7	mc	n. a.	n/a	n/a	yes	TiO_2	n/a	60 Ω/□	screen-printed	12,8 %
8	mc	25 cm^2	no	yes	no	porSi	n/a	100 Ω/□	sputtered	14,1 %
9	mc	130 cm^2	no	yes	yes	yes	n/a	100 Ω/□	screen-printed	13,1 %
10	Cz	100 cm^2	yes	no	SiN_x:H	SiN_x:H	n/a	n/a	screen-printed	17,3 %
11	mc	100 cm^2	n/a	no	SiN_x:H	SiN_x:H	n/a	n/a	screen-printed	15,3 %
12	Cz	25 cm^2	yes	no	no	no	n/a	130 Ω/□	screen-printed	13,05 %
13	Cz	90 cm^2	yes	yes	yes	yes	10 Wcm	n/a	screen-printed	18%
14	mc	100 cm^2	yes	yes	yes	kétrétegű	1 Wcm	n/a	screen-printed	16,9 %

Table

Reading of abbreviations:
Cr.: crystalline type of substrate
FZ – Floating Zone,
Cz – Czochralski,
c – single crystalline, unknown type,
mc – multicrystalline,
Text.: textured,
H2: hydrogenated,
passiv.: passivated,
ARC: antireflection coating,
Bulk res.: bulk resistivity,
N.c. emit.: sheet resistivity of not contacted emitter

doped areas by a conventional diffusion step. Then subsequent openings etched in the oxide by a photolitography step again, where the weakly doped areas are formed by diffusion, followed by etching the whole oxide layer. Then the passivation and antireflection oxide layer is formed, which is etched on the contact areas by a subsequent photolitography process. Here the metal contact layer is deposited followed by the forming of the contact scheme (perhaps by photolitography again). This technology was used for creating the selective emitter for the known highest efficiency crystalline Si solar cells (PERC – Table, line 1 [9], and PERL cells – lines 2-3 [10]). Note that each oxide deposition needs temperatures about 1000°C that increases the production cost. Another disadvantage is that two separate diffusion step needed that complicates the process. The long and high temperature annealing needed for highly doping can degrade the quality of the substrate. Further disadvantages come from multiple scheme aligning.



The structure of the PERL cell

b) PIII-diffusion with photolitography [11]. First the starting deep and highly doped homogeneous emitter is etched entirely in the uncontacted areas by photolitography process. Then a shallow emitter formed by PIII (plasma immersion ion implantation) diffusion on the entire front surface followed by the aligned screen printing and firing (950°C, 1 minute) of the contact paste, so it also activates the implantation (Table, line 4). This techniqe is simpler, needs only one photolitography mask to form the selective emitter, so it is quicker, while even fine scheme resolution can be achieved. Furthermore the metallization is located higher than the illuminated surface, thus in case of good aligning the overlapping of the metallization to the illuminated weakly doped emitter has less probability.However this technology is even relatively complicated, the metallisation is aligned, and there is a possibility that the emitter under the contacts does not join the illuminated emitter that can block current collecting.

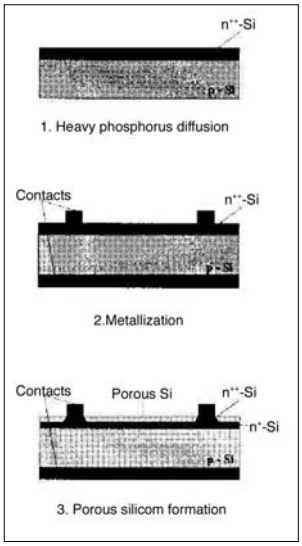
c) A simpler double diffusion technique: after the mask deposition the highly doped emitter formed in the area of the contacts, and after the mask etching a weak doping is made on the whole surface (Table, line 5 [12]). Its advantage compared to the previous process is the eliminating of the emitter etch step.

d) A more simpler technique is the following. The starting weakly doped homogeneous emitter is masked, then the highly doped emitter formed on the contact scheme area by this mask (Table, line 6 [13]). Its advantage compared to the previous technique is that the mask layer does not need to be etched if it has further functions (e.g. using SiN_x it can also serve as passivating layer and ARC). Fine resolution can be also achieved. Disadvantages are the need for metallization align, the relatively complicated technology (double diffusion), thus it is not suitable for mass production.

II.) *Etch-back techniques* are based on the selective etching of the homogeneous emitter using the metallization scheme to form a selective emitter. The top layer of the emitter with high surface doping density is etched between the contacts, so the emitter becomes thinner and the surface doping density becomes lower a little, so the sheet resistivity of the emitter becomes higher. Furthermore the metallization rises from the level of the illuminated surface so the overlap of the contacts to the etched emitter, and so the increasing of the contact resistivity and the dark current can be avoided easier. The etch-back process is usually time consuming and it is hard to control it, because the thickness of the highly doped layer is usually about 0,3 μm, and if the n+ layer remains due to the imperfect etching, then the power decreases.

a) A simple etch-back technology described as follows [14]. An n++-type front side homogeneous emitter (16-20 Ω/□) formed in a p-type Si from a screen-printed diffusion source followed by screen printing and firing (720°C) the back side metallization (Ag-Al) and screen printing and firing (645°C) the front side silver paste. After that a polymer protective layer screen-printed and dried (150°C) that protects the contacts against the influence of the etching liquid. Between the protected metal contacts the emitter etched-back by the HF/HNO_3 liquid until it reaches the appropriate sheet resistivity. With such a solar cell (Table, line 7) 0,5-1% absolute efficiency improvement can be showed compared to the conventional homogeneous emitter cell with 40 Ω/□ sheet resistivity. This technology is relatively simple, a little quicker than the previous ones and needs low temperatures. However the protective layer of the contacts has to be aligned to the metallization, but we can optimize the size of it if we know the width of the underetching. Another drawback is that the etching of the liquid is hard to set accurately so the reproducibility is lower.

b) The *porous Si* etch-back techniqe uses that during the porous Si formation the emitter is etched-back, and as the porous Si also acts as an ARC and passivating layer, no further layers needed. With appropriate etching liquid the texturing also creatable in the meanwhile (line 8 [15,16]). This is a very simple technique for making a solar cell, it is quick and cheap, and needs



Process of etchback selective emitter formed by porous Si

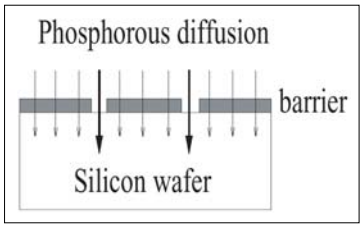
minimum of annealings and materials. It can be a disadvantage that the porous Si scatters the light inside the cell, so an amount of the photons gets in the highly doped emitter under the metallization, where the recombination probability is higher. In case of screen-printed contacts a further drawback is the weak etching of the metallization and the underlying Si surface due to the HF contained electrochemical etching, so the contacts become more fragile and their contact resistivity increases, thus the FF decreases. In case of sputtered metal contacts the etching does not degrade the quality of the contact and the FF, and so it becomes a self aligned process, however it is a more expensive and slower technique.

c) Another possibility is the *self-aligned plasma etchback* technique, where the screen-printed metallization is the mask layer. The highly doped homogeneous emitter between the contacts is thinned by plasma etching (RIE, using SF_6) [17]. A 0,35% avarage absolute efficiency improvement was reached with such a solar cell made by conveyor belt technique (Table, line 9) compared to cells with homogeneous emitter. The highest power can be achieved by etching about 90 nm Si layer (in case of p-type Cz-Si substrate) [18]. Its advance is the self-aligning, furthermore the series resistance and the bulk recombination decreases, thus the FF increases due to the plasma process.

III.) With the following methods we can *form the differently doped areas in one high temperature annealing step* thus the manufacturing time and cost decreases.

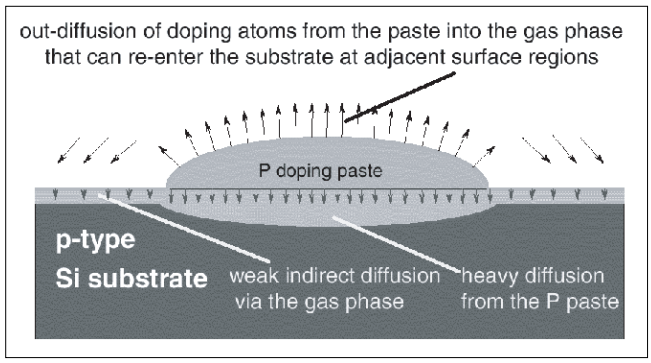
a) Using a *diffusion barrier* for the diffusion (100-1000 nm thick SiO_2) much phosphorous gets into the silicon through the opened windows of the oxide, and few phosphorous gets through the oxide barrier forming weakly doped and shallow emitter there [19]. Varying the properties of the diffusion barrier (thickness, permeability, doping) the doping rate of the underlying Si can be controlled, however it is relatively hard to set it accurately. Using SiO_2 paste for creating the barrier the concentration of the SiO_2 also influences the emitter doping.

Selective emitter created by phosphorous diffusion through diffusion barrier



b) By doping the oxide layer, the technique becomes the following: a highly phosphorous *doped SiO_2 (PSG)* deposited on the whole surface is etched between the contact areas with the help of a screen-printed mask layer followed by the etching of the mask layer and a subsequent deposition of a weakly phosphorous doped PSG on the whole surface. Then a rapid thermal heating (1000°C, 45 s) realizes the phosphorous diffusion. After that the metal paste screen-printed with aligning and fired (Table, lines 10-11 [20]). The properties of the emitters can be easily set by the amount of phosphorous in the PSG. These PSG layers are thin and transparent so it can used even to light-assisted diffusion. Disadvantages are the aligned metallization and the etching of the PSG that complicates the technology.

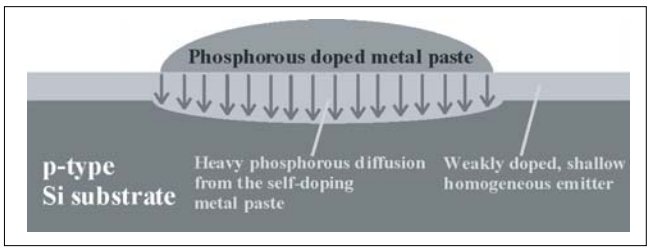
c) A simpler and cheaper similar technology is the following: a phosphorous *doping source paste* screen-printed on the areas of the metallization followed by the applying of a weakly doped spin-on material on the whole surface, and the diffusion annealing (950°C, 80 s). Then the formed glass etched and the metal contacted with aligning (Table, line 12 [21]). With such a cell 1% absolute efficiency improvement reached compared to the homogeneous emitter cells. Note that two screen printing step and aligning are needed.



Selective emitter created by auto-doping technique

d) The *auto-doping* technique: a phosphorous paste screen-printed in the metallization areas onto a p-type Si substrate followed by a diffusion step (960°C, 5 min), so a highly doped and deep n-layer formed under the paste, and a lower doped and shallower n-layer formed between the paste stripes in an indirect way (due to the evaporated P-gas). After the etching of the paste the Ag paste screen-printed with aligning (lines 13-14 [8, 22]). Note that less P source needed thus the cost is also lower. However aligning is needed, and the variance of the values increased due to the indirect diffusion.

e) With one screen printing step of *self-aligned doped (self-doping) metal pastes* onto a weakly doped homogeneous emitter (see next unit) the process becomes very simple and rapid. There is no need for aligning, the material input is economic and it needs minimal annealing. However the composition of the paste is hard to be set properly.



Selective emitter and metallization created by self-doping metal paste

IV.) Another method for making a selective emitter is the *laser overdoping* technique [23].

a) A spin-on diffusion source is used to form the homogeneous, weakly doped emitter followed by a local heating by a laser with appropriate energy in areas where the highly doping needed. It can reach linewidths in the range of 10-25 μm. However the metallization must be aligned, the process needs long time and it is complicated due to the using of laser.

b) However we can use also a *self-aligning contacting* that can be realized by electroless plating. This time the laser must burn through the whole thickness of the phosphorous doping source meanwhile the overdoping. Then the surface can be plated directly through the openings by an electroless plating solution. An appropriate doping source (e.g. doped SiO_2 or SiN_x) can be used as metal mask, passivating layer and ARC as well making the manufacturing simpler.

V.) A *buried contact* structure can be formed by any of the above technologies, thus the resistivity and the ratio of the metallized surface, so also the shadowing decreases.

This technique needs only a trench forming and a clean etching before the highly doping diffusion, then the contact metal can be filled in it. The mask of the trench forming can be equal to the mask of the undercontact diffusion and of the metallization, so it can integrated easily into the previous technologies. However it is more complicated, more expensive, and the metal/Si interface is increased, thus the V_{oc} decreases. On the basis of modelling the buried contact selective emitter cells has a 0,5-0,6% absolute efficiency enhancing compared to the non-buried screen-printed contact selective emitter cells [7]. Several technology integrated the buried contact forming, such as the double diffusion technique (mechanical grooving [24], laser drilling [25], chemical etching [12]) and auto-doping technique [12].

3. Doped metal pastes and the utilizing

Theory [4,26,27,28]

The metal paste

In case of screen printing or other thick layer applying technique the forming of a thick layer is made by special pastes. A paste is a dispersion of solids (conductor and inorganic phases) in an organic vehicle. A contact paste for a solar cell is hard to produce due to the many requirements:

1. Low contact resistivity to the silicon
2. Low line resistance (high electrical conductivity)
3. Negligible effect to the Si substrate (does not degrade the electrical quality of the silicon by creating recombination centers)
4. Little linewidth, good resolution
5. Good solderability
6. Good adhesion (makes a mechanically strong bond to the silicon)
7. Low cost
8. Applyable by an economical process (e.g. by screen printing)

The backside contact material is also expected to create a BSF (Back Surface Field). In the absence of other inorganic species the silver in paste often used for contacting the front side does not interact with the silicon or forms interfacial species until 830°C where a eutectic is formed [26]. The silver does not react with the anti-reflection coatings as well. Thus the inorganic components are essential because they make possible the physical contacting between the silver and the silicon due to the ARC cutting through. The so called glass frit is good for that, however it dissolves the silicon as well.

The self-doping metal paste

The screen printing technology creates a bad ohmic contact on the high sheet resistivity emitter, however the alloyed self-doping paste (SDP) can help that. For a self-doping contact material a plus requirement is the ability to allocate an amount of doping material into the underlying silicon. These requirements are sufficed by a known contact material, the aluminium in case of contacting a p-type silicon. The aluminium dopes the silicon by an alloying process that needs temperatures above the Al-Si eutectic point (577°C). The alloy contains 12,5% of Si and 87,5% of Al by weight. This eutectic electrode has a proper electric conductivity for the currents of the solar cell, and an excellent ohmic contact and interface to the silicon. In addition the aluminium is available in a form of screenprintable paste for a reachable price.

However there is no similar material for contacting the n-type silicon, but there is a possibility for making a selective emitter by using silver pastes alloyed with phosphorous or antimony doping material, which increases the doping under the contacts meanwhile the contacts formed. In this case there is no need for any more aligned screen printing, so the yield is higher and the cost is lower due to the simpler technology. This self-doping, self-aligned contacting technology can make contacts even onto emitters with sheet resistivity of 100-200 Ω/□.

Solving the short-circuit problem

The molten metal can consume some of the underlying silicon (about 0,3 μm) so the level of the regrown

Si can be lower than the level of the surrounding Si, thus the metal can easily short-circuit the p-n junction. Mixing Si in the self-doping metal alloy paste, the level of the metal-Si interface becomes higher than the level of the Si wafer. Using sputtered Ag/Si/B alloy a good quality p-n junction was succesfully created.

An alternative self-doping technology

A self-doping negative electrode can be made by starting with the applying of an unalloyed Ag layer onto the Si surface. Then the Ag and the substrate must be heated above the Ag-Si eutectic temperature (but below the melting point of the Si, e.g. 900-950°C for 2 minutes) in an ambient gas contained phosphorous. The Ag and Si begin to melt, meanwhile the P atoms of the ambient gas absorb into this Ag-Si liquid mixture in a larger amount than into a solid Si suface. Then with the decreasing temperature the melted Si reformes containing the P doping atoms in it, and below the Ag-Si eutectic temperature the Si formes a solid phase alloy with the Ag embedded in the substrate. This Ag-Si alloy is the final contact material. Due to the eutectic weight ratio much more amount of Ag will be in this material than Si, so good electric conductivity ensured. Note that the solid Si also receives the gas phase P atoms but in a much less amount than the melted metal.

The silver can be replaced by another metals (e.g. tin). For making a negative electrode the P can be replaced by other member of Group V of the Periodic Table, and for making a positive electrode it can be replaced by a member of Group III as a dopant vapor.

Unsolved problems

It is not clear yet that the growing conditions at the eutectic temperature how influence the quality of the grown Si, and how can it grow epitaxial as the part of the Si-metal alloy. It is possible that a spontaneous nucleation occures in the melted material. Therefore the cooling must be slowly at the eutectic temperature. The material quality of the Si growed by the self-doping metal system is an important subject of research.

Realizations

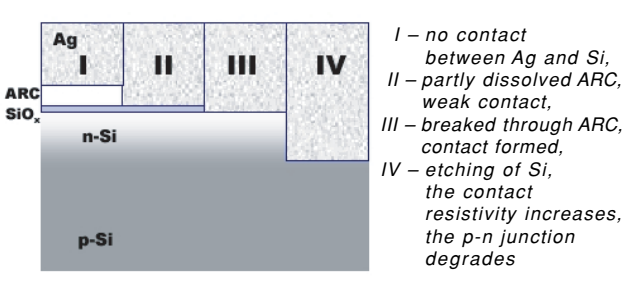
The firing of the metal paste and the contact resistivity

The screenprintable paste can be qualified by measuring the contact resistivity, because it governs the electric behaviour [26]. The contact resistivity depends on the sheet resistivity of the emitter that can be changed by the reaction of the metal and the silicon. Thus the interaction model of the emitter and the metal can be observed when the contact resistivity changes during the high temperature processes.

In case of n-type emitter the surface goes through four states during the firing of the metallization *(figure)*:

I. The antireflection coating (ARC) is unbreaked, high contact resistivity.

II. The ARC is breaked through in part (the TiO_x is dissolved but not the SiO_2 layer), high contact resistivity.



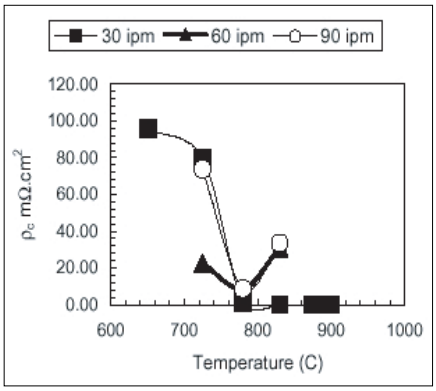
States during the forming of the Ag contact

III. The ARC is breaked through entirely, a good contact formed with the emitter, low contact resistivity.

IV. The emitter layer also etched, so the sheet resistivity of the emitter slowly increases, then the p-n junction damages due to the diffusion of the metal oxide.

To minimize the Si etching after the contacting a rapid thermal annealing must be used. Observing the effects of the annealing we can see that the contact resistivity decreases dramatically between 700 and 800°C, and the minimum can be reached at about 800°C. For higher belt speeds beyond 800°C the contact resistivity increases again, probably due to the interface processes of the quick metal firing.

So applying the Ag-Si eutectic point (830°C) or higher temperatures by rapid thermal annealing the formed contact is not the optimal, so several problem can occured (e.g. thermal expansion differences, Si etching).



Firing temperature and belt speed effection to the contact resistivity

Using the self-doping paste and the significance of the glass frit

The creating technology of a selective emitter solar cell by self-aligned self-doping P doped Ag paste was the following [29]. An n-type homogeneous diffusion layer with a sheet resistivity of 75 Ω/□ and a depth of about 0,25 μm created in a p-type wafer followed by the deposition of a SiN_x antireflection coating by PECVD. Then the back side Al layer screen-printed and annealed in a conveyor belt furnace (860°C, 2 min) followed by the screen printing of the self-doping Ag paste (DuPont PV168) on the top of the SiN_x layer, then it was dried and alloyed (900°C, 2 min).

Good contact resistivity reached (2 mΩcm^2) by fritted paste, however it increased quickly by the decreasing of the base doping (below 10^{18} cm^{-3}), and at 10^{15} cm^{-3} base diffusion it became too high due to the high sheet resistivity of the formed emitter (700 Ω/□).

If an n^- homogeneous emitter layer also existed then the contact resistivity became acceptable (1-12 $m\Omega cm^2$) until n^- layers with sheet resistivity of 100 Ω/□. The problem is that the frit of the paste etches a thin layer of the Si surface.

The selective emitter cells with 75 Ω/□ n^- layer had a little lower serial resistivity and higher FF compared to conventionally produced cells with 40-45 Ω/□ homogeneous emitter. The serial resistivity was about 0,75 Ωcm^2, the shunt resistivity was 2-25 $k\Omega cm^2$, the efficiency was increased by about 0,3%. These values can be made even better by better surface passivation. Based on modelling the front side recombination velocity must be below 10000 cm/s to make the selective emitter effective.

The properties of the surface passivated self-doping selective emitter cell

A similar solar cell with 100 Ω/□ n^- emitter made in another experiment, but under the SiN_x layer a SiO_2 passivating layer applied, and both side contacts were cofired [30]. The optimal heating process was a rapid thermal annealing at 900°C.

The dark current (recombination current) in the emitter of the 45 Ω/□ homogeneous emitter cells was J_{oe}= 337 fA/cm², while for the selective emitter cells it was 185 fA/cm².

The FF of the selective emitter cells (0,768) was slightly lower than for the conventional cells (0,785), however the fill factors of the cells with 100 Ω/□ emitter using conventional contact paste were much lower (0,479-0,709). The slight decreasing of the FF for the selective emitter cells was because the front side contact scheme was optimized to 45 Ω/□ emitters (gridline spacing of 2,2 mm), so it can be restored.

The short circuit current of the selective emitter cells was increased by 0,8-1 mA/cm² to 33,4-33,6 mA/cm² compared to conventional cells, and the open circuit voltage was increased by 9 mV to 635 mV, therefore the efficiency increased by 0,4% to 16,4%. The measurements showed that the selective emitter cells without an oxide passivation layer cannot enhance the properties considerable compared to the conventional cells.

Comparing the IQE curves we can see that the short wavelength response is much better for the selective emitter cells, if the surface passivation is good, so the J_{sc} is better as well. For further wavelengthes these cells are also better than the conventional cells, and the long wavelength response is also slightly better due to the high temperature annealing making the BSF deeper and better.

Self-doping contact on dendritic web Si substrate

The advantageous properties of the dendritic web Si substrate are that it is thin (100 μm), has a good electrical quality (the diffusion length of the minority carriers is multiple of the thickness), and it is cheap.

This substrate was used in an experiment with P contained fritless Ag paste (70 at% Ag, 0,07 at% P, this is the maximum solubility of the P in the Ag at room temperature) [31]. The optimal annealing was at 1000 °C for 10 minutes. The contact resistivity was 0,013 Ωcm^2, while for conventional Ag paste it was 1,9 Ωcm^2.

It was showed that the P atoms got into the Si several micrometers in depth. At the Si surface the doping concentration was higher, so it formed a thinner Schottky-barrier at the metal-semiconductor interface, thus the contact current contained mainly tunnel current through the interface. Therefore the contact resistance is exponentially proportional to the Schottky barrier height and to the inverse of the square root of the doping concentration. This paste could make a pn junction as well.

IBC cell with self-doping contacts

An IBC cell can be formed by self-doping negative and positive electrodes on a Si substrate as follows [4]. First an n^+ layer diffused in the front side of the substrate against the surface recombinations, followed by the dissolving of the so created diffuse glass and the screen printing and drying of a Ag-Ga paste onto the back side. Then a Ag-Sb paste screen-printed onto the back side in an interdigitated scheme followed by the burning out of the organic particles of both pastes (400 °C). This followed by the cofiring of the self-doping positive (Ag-Ga) and negative (Ag-Sb) electrodes, forming the p-n junction as well (with the Ga for n-base or the Sb for p-base), meanwhile a thermic oxide growed passivating the free Si surfaces on both sides of the wafer. This can be achieved by an RTP annealing (900°C, 2 min) in an oxygen ambient. Then an ARC deposited on the uncontacted front side of the cell, so the electrodes can be directly soldered for interconnecting the cells into modules. The positive electrode can be made by pure Al as well, however it has a lower conductivity, its surface is oxidized thus it is unsolderable. The above process is suitable for mass production as well.

An other version of the process applies liquid or solid phosphorous source onto the front side of the cell that forms the front side n^+ layer meanwhile the firing of the back side undoped negative Ag electrode, thus the evaporated P atoms get into the melted Ag-Si layer doping the underlying Si areas. Then the self-doping Al or Ag-Ga positive electrode pastes applied and fired.

A self-doping IBC cell can be made using dedrit web Si subtstrate as well [32]. In this case the front side n^+ doping can be formed during the crystal growing as well.

A commercial cell made using an n-type 20 Ω-cm dendritic web substrate, where the p-n junction formed by alloying a screen-printed Al, while a screen-printed self-doping Ag-P paste contacted the substrate. The Si_3N_4 ARC applied by PECVD. The 5 cm² area cells have J_{sc}=28 mA/cm², V_{oc}=0,55 V, FF=0,69 and the efficiency reaches 11%, and in the future it can reach even 15%.

4. Foresight

The aims of the developments are making cheaper solar cells and meanwhile keeping or improving the volume of the photoelectric efficiencies. This goes not only for chrystalline Si solar cells but for amorphous Si or for thin layer compound semiconductor solar cells (e.g. CIGS) as well.

Today it is not clear yet which way has the future. To decide it we must take into account the manufacturing cost, the usable photoelectric efficiency, the ability for mass production and the lifetime. For example now one of the cheapest solar cell type is the amorphous Si based structure, but its poor efficiency and its relatively short lifetime make it suitable using for only nearer terms. In special cases the area of the solar cell modules can also be an important requirement that can motivate the using of higher efficiency solar cells. Of course in parallel of the nanotechnology improvement the new type solar cell materials also appear (e.g. thin layer Si substrate, solar cell materials modified by quantum technology, photovoltaic carbon nanotube etc.), these will bring great enhancements in the world of photoelectric devices for the farer future.

However until that the actual technology must be improved, and one station for it is the more economic manufacturing of the selective emitter structure for the world's most applied crystalline Si solar cells, and therefore the using of self-doping screenprintable metal contact is an obvious solution.

Acknowledgements

The author would like to thank I. Pintér, Cs. Dücsõ, B. Põdör and other members of HAS RITPMS, and also I. Zólomy and other members of BUTE DED for many helpful discussions. This work was supported by OTKA via contract No. T-033101, and ADVOCATE Project FP 5 via contract No. ENK6-CT-2001-00562.

References

[1] Kuthi Edvárd Bálint:
„Foszfor diffúzióval kialakított sekély emitterekhez kontaktus készítése adalékolt Ag paszta segítségével és a p-n átmenetek vizsgálata", thesis,
Budapest University of Technology and Economics, Department of Electron Devices, 2002.

[2] Dr. Mizsei János:
„Napelemek", notes,
TUB DED, 1999., www.eet.bme.hu/publications/e_books/solar/napelem.zip

[3] Christiana Honsberg, Stuart Bowden:
„Photovoltaics PVCDROM Part 1, Photovoltaic Devices",
The University of Newsouthwales Photovoltaics Centre, 1999.

[4] Meier, Daniel L., Davis, Hubert P.:
„Method and apparatus for self-doping negative and positive electrodes for silicon solar cells and other devices",
USA patent 6180869, 30 January 2001.

[5] R. R. King, R. A. Sinton, R. M. Swanson:
„Studies of Diffused Phosphorus Emitters: Saturation Current, Surface Recombination Velocity, and Quantum Efficiency",
IEEE Transactions of Electron Devices,
Vol.37., No.2., February 1990., pp.365–371.

[6] K. Misiakos, F. A. Lindholm:
„Toward a systematic design theory for silicon solar cells using optimization techniques",
Solar Cells, No.17., 1986, pp.29–52.

[7] J. Nijs, E. Demesmaeker, J. Szlufcik, J. Poortmans, L. Frisson, K. De Clercq, M. Ghannam:
„Latest efficiency results with the screenprinting technology and comparison with the buried contact structure",
Proceedings 1st IEEE WCPEC,
1994., Vol.2., pp.1242–49.

[8] J. Horzel, J. Szlufcik, J. Nijs, R. Mertens:
„A Simple Processing Sequence for Selective Emitters", Proc. 26th IEEE Photovoltaic Specialist Conference, 1997, Anaheim, California, USA, pp.139.

[9] Andrew W. Blakers, Aihua Wang, Adele M. Milne, Jianhua Zhao, Martin A. Green:
„22,8% efficient silicon solar cell",
Applied Physics Letters,
Vol.55., No.13., 25 September 1989., pp.1363–65.

[10] Jianhua Zhao, Aihua Wang, Martin A. Green:
„19,8% efficient 'honeycomb' textured multicrystalline and 24,4% monocrystalline silicon solar cells",
Applied Physics Letters,
Vol.73., No.14., 5 October 1998., pp.1991–93.

[11] I. Pintér, A. H. Abdulhadi, Cs. Dücsõ, I. Bársony, J. Poortmans, S. Sivoththaman, H. F. W. Dekkers, G. J. Adriaenssens:
„Silicon solar cells prepared by PIII-RTP techniqe",
Proceedings of the 16th European Photovoltaic Solar Energy Conference,
1-5 May 2000, Glasgow, UK, pp.1743.

[12] Pirozzi, U. Besi-Vetrella, S. Loreti, P. Mangiapane:
„Screen printed contacts in buried silicon solar cells",
16th European Photovoltaic Solar Energy Conference, 1-5 May 2000, Glasgow, UK

[13] G. Arabito, F. Artuso, M. Belardinelli, V. Barbarossa, U. Besi Vetrella, L. Gentilin, M. L. Grilli, P. Mangiapane, L. Pirozzi:
„Electroless metallizations for contacts in buried structures",
2nd World Conference and Exhibition on Photovoltaic Solar Energy Conversion,
6-10 July 1998, Vienna, Austria,
pp.1558–61.

[14] J. Szlufcik, H. E. Elgamel, M. Ghannam, J. Nijs, R. Mertens: „Simple integral screenprinting process for selective emitter polycrystalline silicon solar cells”, Applied Physics Letter, Vol.59, No.13., 23 September 1991., pp.1583–84.
[15] R. R. Bilyalov, H. Lautenschlager, R. Schindler: „Multicrystalline silicon solar cells with porous silicon selective emitter”, 2nd World Conference and Exhibition on Photovoltaic Solar Energy Conversion, 6-10 July 1998, Vienna, Austria, pp.1642–45.
[16] M. Schnell, R. Lüdemann, S. Schaefer: „Stain etched porous silicon – a simple method for the simultaneous formation of selective emitter and ARC”, 16th European Photovoltaic Solar Energy Conference, 1-5 May 2000, Glasgow, UK
[17] S. Ruby, P. Yang, M. Roy, S. Narayanan: „Recent Progress on the Self Aligned, Selective Emitter Silicon Solar Cell”, Proceedings 26th IEEE Photovoltaic Specialist Conference, 1997, Anaheim, California, USA, pp.39.
[18] Nick Mardesich: „Solar cell efficiency enhancement by junction etching and conductive AR coating processes”, 15th IEEE Photovoltaic Specialists Conference, 12-15 May 1981, Kissimmee, USA
[19] J. H. Bultman, R. Kinderman, J. Hoornstra, M. Koppes (ECN Solar Energy): „Single step selective emitter using diffusion barriers”, 16th European Photovoltaic Solar Energy Conference, 1-5 May 2000, Glasgow, UK
[20] J. Horzel, S. Sivoththaman, J. Nijs: „Screen-printed rapid thermal processed (RTP) selective emitter solar cells using a single diffusion step”, 16th European Photovoltaic Solar Energy Conference, 1-5 May 2000, Glasgow, UK
[21] L. Debarge, J. C. Muller, B. Forget, D. Fournier, L. Frisson: „Screen-printed paste and spin-on source applied to selective emitter formation in a single rapid thermal diffusion step”, 16th European Photovoltaic Solar Energy Conference, 1-5 May 2000, Glasgow, UK
[22] J. Horzel, J. Szlufcik, J. Nijs: „High efficiency industrial screen printed selective emitter solar cells”, 16th European Photovoltaic Solar Energy Conference, 1-5 May 2000, Glasgow, UK
[23] Wenham, Stuart Ross, Green, Martin Andrew: „Self aligning method for forming a selective emitter and metallization in a solar cell”, USA patent 6429037, 6 August 2002.
[24] Jooss, M. Spiegel, P. Fath, E. Bucher, S. Roberts, T. M. Bruton: „Large area buried contact solar cells on multicrystalline silicon with mechanical surface texturization and bulk passivation”, 16th European Photovoltaic Solar Energy Conference, 1-5 May 2000, Glasgow, UK
[25] Shaoqi He, Yuting Wang, Xudong Li, Yuwen Zhao, Zhongming Li, Yuan Yu: „Laser grooved buried contact solar cell”, 2nd World Conference and Exhibition on Photovoltaic Solar Energy Conversion, 6-10 July 1998, Vienna, Austria, pp.1446–48.
[26] Richard J. S. Young, Alan F. Carroll (DuPont Microcircuit Materials): „Advances in front-side thick film metallisations for silicon solar cells”, 16th European Photovoltaic Solar Energy Conference, 1-5 May 2000, Glasgow, UK
[27] D. E. Riemer: „Evaluation of thick film materials for use as solar cell contacts”, Proceedings 13th IEEE Photovoltaic Specialist Conference (1978), pp.603.
[28] David D. Smith: „Review of Issues for Development of Self-Doping Metallizations”, www.sandia.gov/pv/smith.pdf Sandia National Laboratories, 2000. május
[29] A. Rohatgi, M. Hilali, D. L. Meier, A. Ebong, C. Honsberg, A. F. Carrol, P. Hacke: „Self-aligned self-doping selective emitter for screen-printed silicon solar cells”, 17th European Photovoltaic Solar Energy Conference and Exhibition, 22-26 October 2001, Munich, Germany
[30] M. Hilali, J.-W. Jeong, A. Rohatgi, D. L. Meier, A. F. Carroll: „Optimization of Self-Doping Ag Paste Firing to Achieve High Fill Factors on Screen-Printed Silicon Solar Cells with a 100 ohm/sq. Emitter”, 29th IEEE PVSC, New Orleans, Poster 1P2.17, May 2002.
[31] L. M. Porter, A. Teicher, D. L. Meier: „Phosphorus-doped, silver-based pastes for self-doping ohmic contacts for crystalline silicon solar cells”, Solar Energy Materials & Solar Cells, No.73. (2002), pp.209–219.
[32] Ebara Solar Inc.: „IBC cell process technology”, www.ebarasolar.thomasregister.com/olc/ebarasolar/cellibc.htm (2001-2002)

Parameter control of laser beams
in function of the pattern of multilayer structures

PÉTER GORDON, BÁLINT BALOGH

Budapest University of Technology and Economics, Department of Electronic Technology
gordon@ett.bme.hu

Reviewed

Keywords: lasertechnology, microelectronics, micro-via machining lasers, laser-material interaction

The development of electronic industry is strongly related to the permanently shrinking sizes, a challenge that in many instances cannot be coped with by means of the perfection of traditional technologies. The laser devices can be a useful means for their substitution, but as we do not know perfectly the processes that occur at the interaction of the laser with the material, we are unable to make the best of the laser technology. This contribution intends to highlight the features of this interaction, the impacts that the pattern applied on the substrate and the machining parameters have on the results of the machining process. We have set us the goal to exploit more efficiently the opportunities presented by the laser-based machining, by changing the laser beam parameters in function of pattern and material.

Introduction

The electronic industry is developing restlessly, though not necessarily at the space that had been predicted a few decades ago. Our need is constantly growing for smaller and smaller circuits and devices with even more reduced consumption, but with an ability to provide a greater range of services. As the sizes are getting smaller, the traditional manufacturing technologies are nearing by slow degrees the limits of their capabilities. Clearly, the manufacturers tend to reduce even further the strip width and to increase the number of the integrated functions. In this way, the IC sizes can just become smaller, entailing however the necessity of providing for them more connection points, i.e. bond-outs toward their environment. The high number of the leads, which can be as high as several hundreds, is impeding the easy handling of the integrated circuits by means of traditional packaging, because these would take up a much greater space on the surface being so small anyhow. Thus we are constrained to ignore the ease of handling, and to select the smallest possible packaging for the integrated circuits. In this way we are reaching the concept of the present CSP (Chip Scale Package): µBGA, flip-chip, TAB.

The rejection of „huge" packages can smooth things down for IC manufacturers, but the circuit substrate industry must face completely new challenges. The wires of the CSP IC are connected either to leads of line width down to 25 µm or to pads of 50 µm diameter. Therefore the substrate manufacturers are bound to refine their technologies to the order of magnitude of the hair diameter. But the traditional solutions used for this purpose are either unviable or not cost effective.

One of the alternatives for the solutions, which have been based for tens of years on the same principle and which have already achieved their limits of capacity, could be the laser technology. In the following we will outline the widespread use of the technology in our era, the advantages and perspectives it can provide and the compromises we are constrained to, while concentrating on one of the best-known material of the flexible circuit substrates, i.e. the polyimide.

The laser beam as tool

The individual features of the laser beam are well known: the beam is coherent, parallel and fundamentally monochromatic. The ideal laser beam can be focused extremely well, thus being able to produce energy densities of several MJ/mm^2 on the machined material. This feature permits to generate explosion-like processes on the surface, and makes it possible to remove material in small doses, in a tolerant way for the ambient environment. The focused laser beam as a machining tool (with diameter just down to 5 to 10 µm) can be moved along optically at a relatively perfect precision, without any defined feeding direction and, of course, any tear-down [2].

Drilling of through-holes or blind holes (vias) on circuit substrates by laser machining is incontestably more advantageous than mechanical drilling, irrespective of the larger diameter holes. In these applications, the history of laser machining began several decades before, so that micro-via machining lasers can be found in any substrate manufacturing line that fulfils the most recent needs. Due to the controllable, pulse operation mode of micro machining lasers, the laser beam can drill vias of preset depth in multi-layer structures as well, even in the case where the beam has to penetrate materials of clearly different physical features (copper, synthetic resin). This is achieved at a production capacity of several hundreds of vias per second [5].

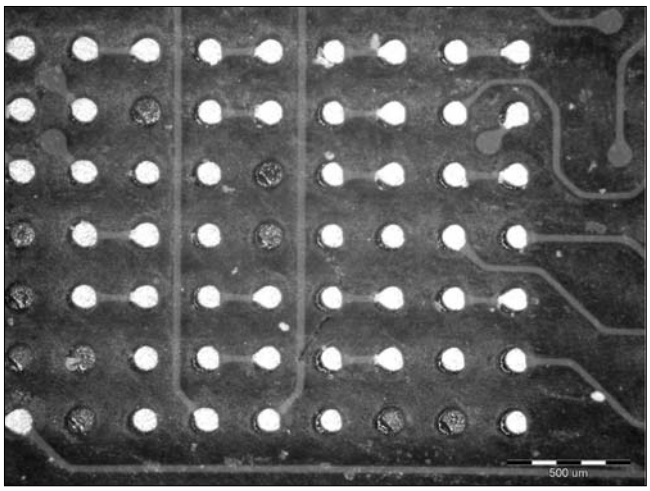
When creating conveniently fine patterns during the production of circuit substrates, the laser technology is

by far less commonly used than in the case of drilling. While drilling in a one-step technology could be replaced straight-out by laser machining, the multi-step pattern machining needs a definitive answer to the question whether and which steps could be qualified as viable by lasers. The arguments for the lasers are the extremely fine granularity to be achieved on the circuit substrates (line widths going incidentally down to the range under 1 mil), the substitution of up to 5 technological steps (photoresist application, exposition, development, etching, resist removal) by a single laser machining phase, the flexibility and the re-adaptation facility during the machining process [2]. The arguments against it are however the lower production capacity and the neighbourhood exposed to heat effects both in lateral direction and in the lower layers not affected directly. The patterned multi-layer structures have the ability to transform also thermally the subject of the micro machining into an inhomogeneous medium, so that the power input represented by the laser pulses – which can be fed also individually – can have from time to time different effects.

The interaction between the laser beam and the material is an exceptionally complex process. The users of the laser machining equipment, not obliged to have an insight into these processes, possess the skill of fitting the machine, pre-tuned roughly in a convenient manner, by means of a few settable parameters into the manufacturing line. The only question is, however, whether such appliances with a dedicated function are fully exploiting all the possibilities that the lasers could present. As a matter of fact, they do not change the machining parameters in function of the patterns of the inner layers, which thermally are just in connection. Production engineers, while setting present laser appliances by a "cut-and-try" method, are inclined to score this technique rapidly as unapt for certain complex jobs.

Fig. 1. shows an example of our investigations.

Fig. 1.
Laser-drilled 100 µm holes in the soldering-prevention layer – The non-leaded soldering pads were damaged in the presence of the same machining parameters



Here the effect of thermal conductance comes directly to light. A more sophisticated dilemma is when the patterned copper layer is not directly connected to the machined layer, but thermally it effectively influences the results.

If we knew both the interaction process between laser and the machined material and the influence which certain accompanying phenomena of the processes are exerting on this (thermal, optical, acoustic, etc.), we could have a model at hand, based on which we would be able to prepare the lasers used in the mass production for the solution of more specific and more complex problems, while keeping the flexibility. So the laser-based machining could be an alternative for the traditional technologies being at the limit of their performance.

Parameters of the laser-based machining

The abundance of the direct machining parameters is by itself representative of the complexity of the process. A possible grouping of the main parameters is shown in the following:

1. Parameters characteristic for a Q-switched laser source:
- wavelength,
- pulse width, time-dependent distribution of the energy,
- per-pulse energy, average power,
- pulse repetition frequency,
- beam quality (energy distribution over the beam cross-section),

2. Parameters describing the characteristics of the optical system:
- beam sweeping speed,
- diameter of the focal plot,
- focal range.

3. Parameters defining the machining geometry:
- pattern,
- grid spacing,
- number of the machining steps.

4. Environmental parameters:
- composed thermal conductance of the inhomogeneous, multi-layer structures,
- characteristics of the atmosphere (gas type, gas flowing speed) [3].

Due to the characteristics of the material to be machined and of the technical realities, the possible values of the majority of the above parameters will be constrained – fortunately – soon to a manageable interval. The laser type defines the wavelength. In the electronic technology, the 355 nm wavelength is prevalently used, which is produced of the Nd:YAG laser beam by frequency tripling. This prevalence is not by accident: the beam at the border of the visible and the UV range is absorbed well in almost every material used in the

rigid and/or flexible NYHL technology. (Therefore, we obviously cannot refer to the processes that automatically dominate the selective material removal, so it is necessary to maintain the controllability as much as possible during the machining process.) Similarly, both the width and the power to time function of the laser pulses emanate from the viability of the laser. As a well-proven fact, the machining "purity" is, in the sense of thermal load, strongly dependent on the duration of the interaction. The shorter the interval is in which we communicate energy to the surface, the more intensively we can cause the exposed part to ablate, as the heat has reduced time to propagate from the area directly affected [1].

We have greater chance to choose from all the multitude of the machining parameters those, which we are able to change at a rate that is comparable to the machining speed in function of the pattern. We cannot modify abruptly the energy and/or frequency behaviour of the laser beam without the change of the beam quality. (There are certain lasers that provide the possibility to compensate the thermal lens created in the laser crystal, but this may require several seconds.) The sweeping speed, the machining geometry, however, can be changed at any time. But before proceeding to examine the way in which it influences the machining results, we present the effects of thermal conductance in the case of copper-polyimide substrates.

The effects of thermal conductance

As referred to earlier, there is a need to fit the machining behaviour to the pattern in order to utilize completely the features of laser-based machining. The diversified pattern created on the substrate, in its several layers, may result in an extensive diversification of the thermal behaviour next to the surface. While the UV lasers have an obviously smaller thermal effect than their counterparts of greater wavelengths [4], this effect has all the same an important role when compared with the 355 nm wavelength of the frequency-tripled Nd:YAG laser. This statement is well underscored by the fact that the thermal conductance of a copper layer not directly affected influences the machining outcome in a detectable measure.

The patterns shown in *Fig. 2/a. and 2/b.,* at completely the same settings (frequency: 100 kHz, beam deflection speed: 300 mm/s, grid spacing: 10 µm, pulse energy: 3,7 µJ), have been produced at a 15x exposure. At places where no copper has been present on the substrate, almost all the polyimide volume could be removed. The copper foil, with a thermal conductance by order of magnitudes better than that of the polyimide substrate, has been capable of carrying away a part of the heat produced, thus a 10 µm thick polyimide layer could be formed on its surface.

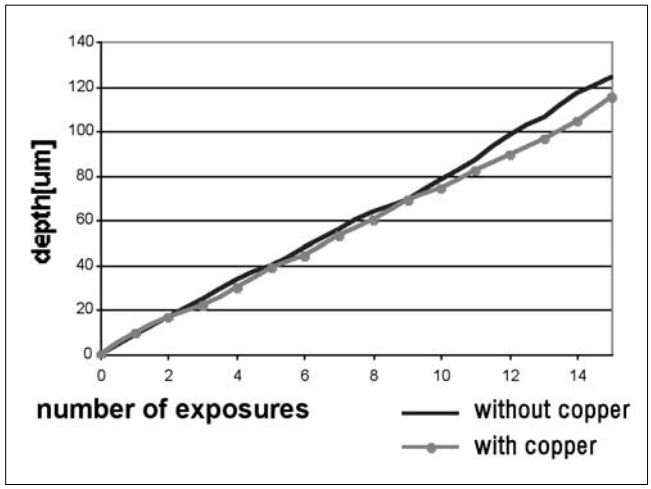
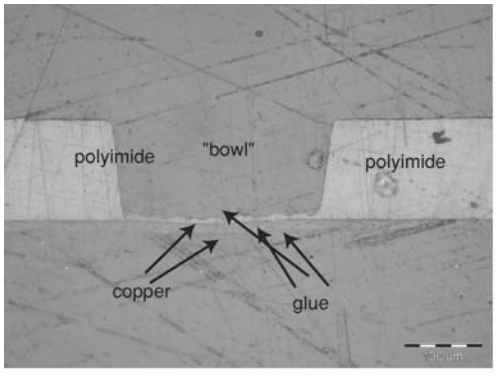


Diagram 1.
Depth as a function of the number of exposures

As it can be seen in *Diagram 1.,* each exposure step removes a ca. 8 µm thick polyimide layer. While maintaining the same settings, the copper foil bonded to the substrate did not produce any remarkable difference till the eleventh exposure. From this point on, the thickness, i.e. the heat insulation ability of the copper layer has also been reducing, so the copper could lead away certain part of the heat produced. Cleary, the ablation of the material of lower temperature requires higher energy, so that the same amount of energy permits to remove less material.

Thus it can be stated that the differences of the heat conductance caused by the patterns of the lower substrate layer have a noticeable effect on the quantities of the material removed. But modifying the laser parameters during the machining process can counter-



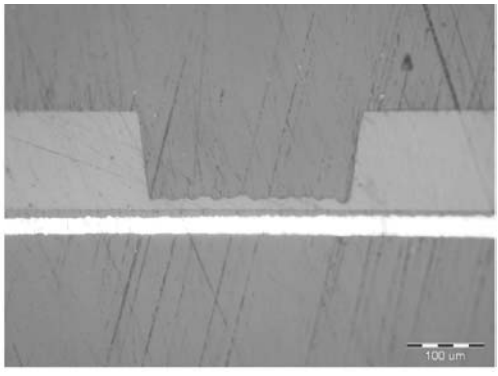


Fig. 2/a and 2/b.

Opening a window in the polyimide foil – the state after 15 exposures.

When the copper layer is present (Fig. 2/b), approximately 10 µm less material is removed.

balance this effect. In our experiments we investigated as well how the polyimide substrate responds to the changes of certain parameters. This will be shown in the following.

Effects of the pulse energy

It can be seen in *Diagram 2.* how the changes of the pulse energy, ceteris paribus, influences the depth of the "bowl" created. There is a linear correspondence between the amount of polyimide removed and the energy within a narrow range, where the depth of the grooves can be controlled easily by changing the energy. Above a specific level (25 µJ in this case), however, the increase of the pulse energy has little weight on the amount of material removed.

Because in the case of most appliances it can take several seconds to change the pulse energy, it is expedient to carry out this modification between the machining phases. We will compensate the pattern unhomogeneity by means of an other parameter, which can be changed more rapidly.

Diagram 2.
Depth as a function of the pulse energy

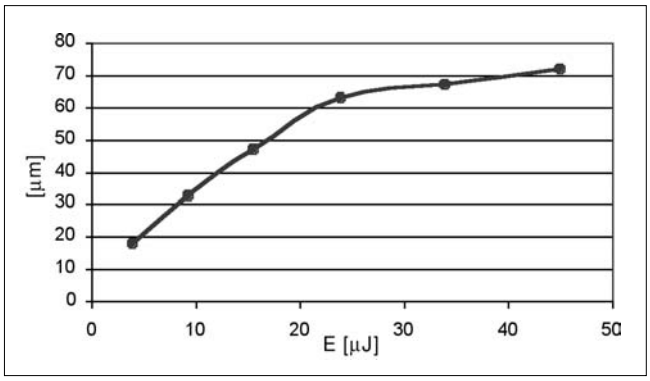
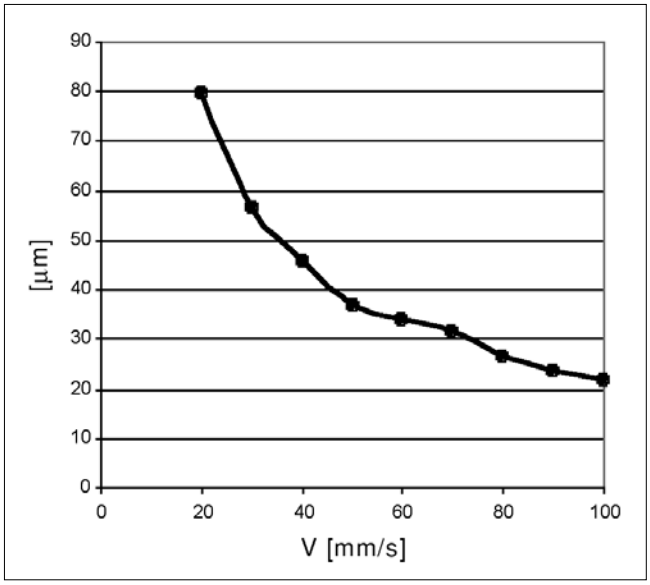


Diagram 3.
Depth as a function of the beam deflection speed



Sweeping speed

The widely used galvanometric beam deflection systems permit to produce jump changes in the sweeping speed. The speed increase produces a logarithmic reduction in the amount of the material removed, thus relatively small speed changes can compensate the high thermal conductance differences caused by the pattern. But at a given pulse repetition frequency, the beam deflection speed cannot be increased above all limits, because in this case the overlaps between the "prints" of the successive shots would diminish in such a measure that this could lead to ragging of the grooves produced.

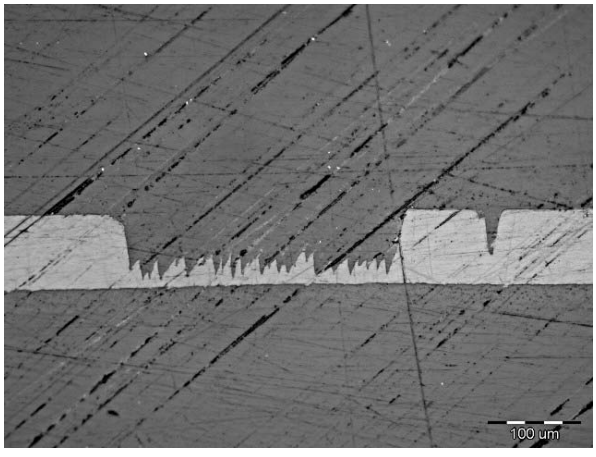
Grid spacing

When machining a given surface, the number of the adjacent grooves depends on the grid spacing. By reducing the grid spacing, the number of the lines and thus the machining time increases proportionally.

Along with the sweeping speed, the spacing of the grooves has proved to be the second definitive parameter, as it can be changed at any time, and it determines the energy fed into a unitary surface. On the other hand, the amounts of energy and its transmission interval have a direct effect on the quantity of the material ablated, i.e. on one of the most important factors of the phenomena we have investigated.

For mass production, it would be efficient to use large grids, which would permit to save time. We must, however, opt for sufficiently small groove spacing, because otherwise the surface of the material machined would be unreasonably uneven. *Fig. 3.* shows the cross section of a window manufactured at too large grid spacing. (The V shape shown on the right-hand side is the cross section of a laser-made groove.)

Fig. 3. Uneven surface due to the too large grid



Conclusion

The widespread use of CSP packages with a large number of bondouts calls for the introduction of laser-based technologies. In fact, these are not completely made suitable for the production of exact and reproducible results just down in the 10 micron range. We have seen that machining processes, as fine as this, results in the reaction of the ambience of the relevant volume on the material on the process itself.

Our experiments took place within the framework of two projects of the European Union, where we have been selecting and optimising the appropriate laser-based technologies for four sorts of rigid and flexible substrates, respectively.

Furthermore, our work aims at creating such a model, which could be a useful means for the material- and pattern-specific control of the laser beam parameters. For this purpose, we must define and simulate the dominant processes. Once this task performed, the model will be able to be built into the control program of an industrial laser-based machining station.

References

[1] Illyefalvi-Vitéz Zsolt:
Laser processing for
microelectronics packaging applications,
Microelectronics Reliablity 41 (2001), pp.563–570.
[2] Laser Machining Processes,
www.columbia.edu/cu/mechanical/mrl/ntm
(12.05.2003.)
[3] Gordon P., Berényi R., Balogh B.:
Controlled Laser Ablation of Polyimide Substrates,
36th International Symposium on Microelectronics,
IMAPS 2003, Boston, Massachusetts,
18-20. November, 2003, pp.725–730.
[4] Y.H. Chen, H.Y. Zheng, K.S. Wong, S.C. Tam:
Excimer laser drilling of polymers –
Microelectronics Packaging and Laser Processing,
SPIE, Singapore, 23-26. June, 1997, pp.202–210.
[5] William M. Steen:
Laser Material Processing,
Springer Verlag 1998.

News

Veraz Networks announced that MobilTel,
the largest GSM cellular operator in Bulgaria serving over two million customers, has deployed Veraz's high-compression I-Gate 4000 media gateways in its cellular network to carry domestic and international traffic. Initially, the media gateways will be used for a multi-point trunking application launched between three major cities in Bulgaria and an international point of presence (POP) in Frankfurt, with the opportunity to expand to more international POPs.

This is possible due to the advanced functionality of Veraz's media gateways to support voice, fax, modem and signaling transport while offering a unique migration option that protects existing investment and delivers a future proof, next-generation solution.

ITU has reached agreement on a new global standard
that specifies the application of the two main technologies used for encoding signals for DSL – Discrete MultiTone (DMT) technology and Quadrature Amplitude Modulation (QAM) – to VDSL (Very high-speed Digital Subscriber Line) technology. VDSL gives multi-megabit network access via ordinary telephone subscriber lines, allowing operators to offer a 'triple play' of services – multiple high-quality digital video streams, high-speed internet access and voice.

"Future evolution of the VDSL standard, promising even higher bit rates and longer distances, will be based on the DMT technology used for ADSL, thus establishing a single world-wide standard. This will allow the broadband telecom consumer to benefit from the economies of scale offered by global volumes as well as the technological innovation driven by competition." – said the Chairman of ITU-T Study Group 15.

VDSL can be deployed from central offices or from optical fibre-fed cabinets located near customer premises. Actual bit rates obtained will depend on the distance between the central office/cabinet and the customer premises and can be up to 50 Mbits downstream, but will typically be closer to 23 Mbits and 4 Mbits upstream.

Privacy homomorphisms

RÉKA LIMBEK, PÉTER SZIKLAI

Eötvös Loránd University, Faculty of Natural Sciences,
Department of Information Systems,Department of Computer Science
lreka@elte.hu, sziklai@cs.elte.hu

Reviewed

Keywords: data security, calulation- and data-delegation, coding

Cryptographic homomorphisms are used to charge a company with performing calculations where both input and output data are confident and therefore can be communicated in cryptographic ways only. This requires the use of special cryptography so that operations can be performed on coded data and decoding the results one gets the same result as if the calculation were made with uncoded data. This is the main point of the article which also introduces basic methods and possible attacks.

Introduction

Privacy homomorphisms were introduced by Rivest, Adleman and Dertouzos [6] in 1978 to solve the *computing delegation* problem in the first instance. This typically occurs when the data owner has only limited computing facilities, i.e. either the computation to be performed is too complex or an unmanageable bulk of data must be processed. In this case, the owner is constrained to deliver his data to a computer centre (data manager) that is able to perform the necessary computations. If the data are sensitive, i.e. of a confidential nature, there is the obvious problem that they could get into an environment which is not necessarily reliable.

This situation emerges in relation of a great number of Internet-based applications, typically when we are using a remote service. Even a modest software for currency conversion or for route planning can fall into this category, but the classical examples include programs for portfolio management and income tax calculation, running not on our computer but on a server operated by third parties. The problem of computing delegation may also occur in the field of academic research, where for example a medical research team uses a(n insecure) university mainframe for processing confidential healthcare records.

The question of *data delegation* represents a similar problem. The two kinds of delegation differ from each other in the fact that in the case of data delegation the results are relevant not for the data owner but for the data manager. As for the owner, the complete computing process is indifferent, he often does not possess the appropriate facilities for performing the computation, and "lends" his data only for the duration of the computing process. The data manager, for his part, wants to own not the raw data, but the processed data (for example the results of statistical analysis). As far as sensitive data is concerned, there is again the problem that this data will get out for the duration of the "lending" into a not necessarily reliable environment.

Data delegation typically occurs in the case of organisations having a federal-like structure. The state administration has generally a structure like this, but equally adequate examples are the European Union or the United States of America. Each member state has its own data (budget, registration of the population, etc.), which a central organisation would like to use for analysing purposes. In return for supplying the data the member states can require the opportunity to analyse the collected data. Because of data protection considerations, however, they have no possibility to store data of other member states, and have neither the computing capacity to perform the necessary operations. In turn, the central organisation, which is not facing any of these problems, has no excess capacity to perform for each member state the computing task they ordered. How and in which form could the collected data be transferred to the member states so that all these problems could be solved?

The secure solution of data delegation could also expand the application of smart cards. In this way, the need that a resource-demanding application must run on the data stored by the card would not present any problem. The card would simply export the data required for the computation to the unit running the application.

In both cases of delegation, there is a double challenge we must meet. On the one hand, we have to prevent the unauthorized access during the data transfer, and on the other, appropriate measures must be taken in order to impede that, in the case of computing delegation, the computing software, the computer and its personnel and, in the case of data delegation, the members can get (unauthorized) access to valuable information as regards the raw data, based on the data made available for them. The deep exploration of the former problem began as early as before the Internet had been going to spread out, especially due to the military and state security applications. Once the world wide net became the forum of more and more applications, the same question has been set in some other way, but the introduction and the convenient imple-

mentation of public key encryption have largely solved the data security problem.

The aim of this contribution is to investigate the second data security problem we mentioned above. Hardware and/or database management solutions are available in both cases, such as the use of physically protected processors or data fragments [8]. The cryptographic solution we are interested in, however, is the application of *privacy homomorphisms* that provide the possibility to transfer the data in an encrypted form to the unreliable level (computer centre, member states) and to perform there the relevant operations *without data decryption*, in their encrypted form. After the results have been sent back in an encrypted form to the reliable level (limited-capacity user, central organisation), we will get, after decryption, *the same result* as in the case where we would have performed the operations in the original form.

Higher security can be achieved by combining the different solutions, but the interpretation of these possibilities lies beyond the boundary of this contribution.

Terminology

Data appearing in a form that is interpretable for everyone is called *plaintext*, irrespective of the data type. The enciphering of the plaintext is called *coding* or *encryption*, the result of which is the *ciphertext* or *encrypted* text. If the legitimate user decrypts the ciphertext, we are speaking of *deciphering* of the encrypted text. The result of this is the original plaintext. If an attacker tries to decrypt the encrypted text, we have to do with *breaking* of the text. The decoding process is a sort of inversion of the coding process.

The encryption is performed by an encryption algorithm, which possesses a parameter, i.e. the *key*. Basically, the deciphering algorithm possesses this key parameter as well. In the case where both keys are identical, the process is called *symmetric* or *secret key encryption*. If, however, the two keys are different, it is the case of *asymmetric* or *public key encryption*, where one of the keys is generally public, i.e. accessible for everyone.

The attacker intending to break either the code or the encrypting algorithm seeks principally the way to break the encrypted text but occasionally also to define the secret key used for the coding. As the security of an encryption algorithm, i.e. in which measure it is resistant to the attacks, shall never depend on whether the algorithm itself is known for the public or not, we always suppose that the attacker knows the encryption method. For the attacker it is important to define the plaintext and the key used.

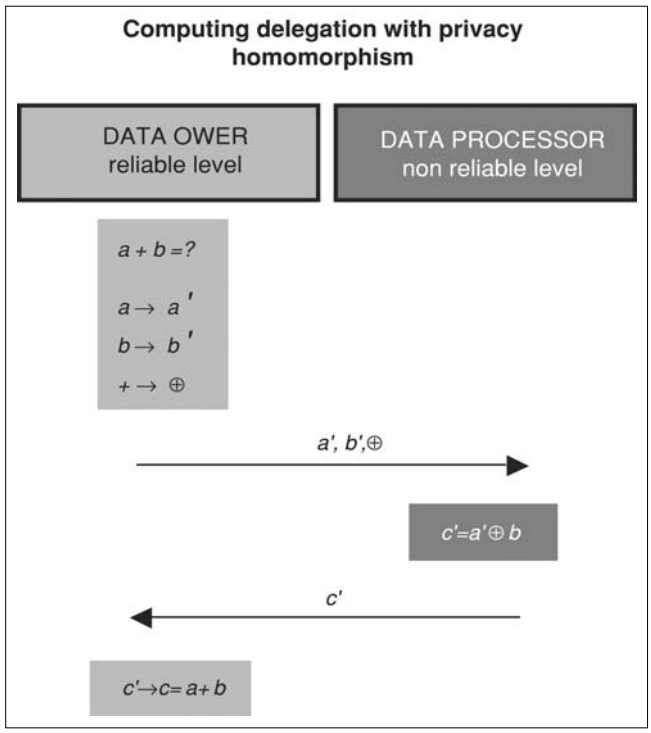
Privacy homomorphisms

Privacy homomorphisms can be used for processing encrypted data. These homomorphisms actually perform the encryption of the data so that the operations can be executed on the encrypted data in the same way as otherwise.

A privacy homomorphism represents an operation-preserving mapping from the set of plaintexts to the set of encrypted texts. Formally, if S is the set of plaintexts and S' is the set of encrypted texts, then an E_K: $S \rightarrow S'$ homomorphism can be defined where K is the key used as the parameter of the function.

Let the operations and predicates interpreted on S and S' be $f_1, f_2, \ldots, f_k$ and $p_1, p_2, \ldots, p_l$ and $f'_1, f'_2, \ldots, f'_k$ and $p'_1, p'_2, \ldots, p'_l$, respectively. Each operation or predicate within S corresponds with one interpreted on S', which generally differs from the original (as we select the operands from other sets), but it also can be very similar. The operation-preservation of E_K means formally that for

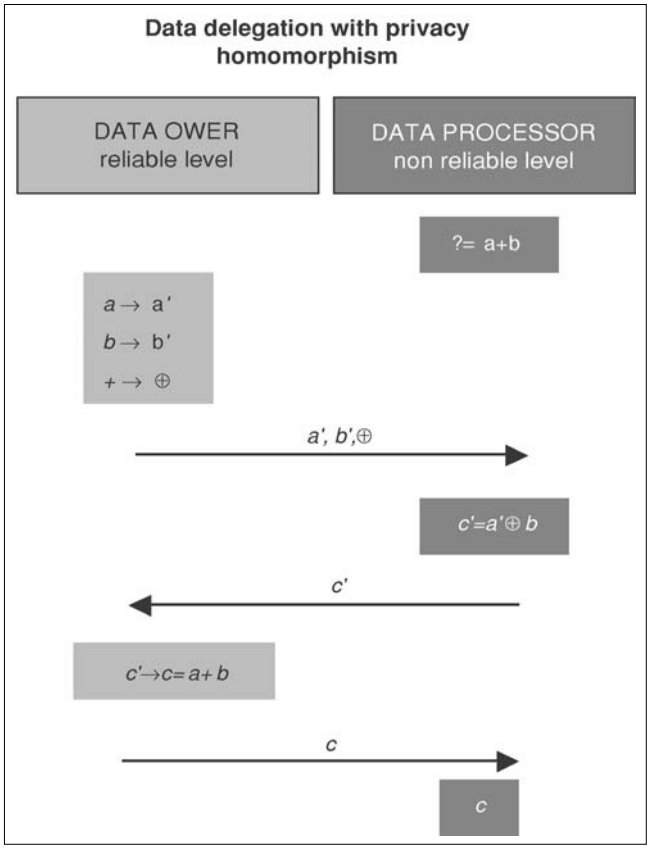
any $a,b,\ldots \in S$ and any $i = 1,\ldots,\mathbf{k}$
$E_K(f_i(a,b,\ldots)) = f'_i(E_K(a), E_K(b),\ldots)$,
as well as for any $j = 1,\ldots,l$
$p_j(a,b,\ldots) = p'_j(E_K(a), E_K(b),\ldots)$.



The deciphering (decryption) of the encrypted text is performed by a function D_K: $S' \rightarrow S$, with S' being the value set of the homomorphism E_K and K' the key. The function D corresponds to the inverse of E, so that it is also an operation-preserving mapping. This feature ensures that performing the operations on the encrypted text, then deciphering it yields the expected result, i.e. for any $a,b,\ldots \in S$ and i $f_i(a,b,\ldots) = D_K(f'_i(E_K(a), E_K(b),\ldots))$.

How does such a homomorphism work?

For illustration consider an RSA-related privacy homomorphism. As a key we select two large primes $K=(p, q)$, and denote the product of them by m. The set of plaintexts is $Z_m = \{0,1,2,\ldots, m\text{-}1\}$, the modulo m residual classes, with the conventional modulo m operators $+, -,$?. The encrypted correspondent of a plaintext $a \in Z_m$ is



formed by a number pair, the components of which are the remainders of the division of a with p and q, respectively. Formally it can be expressed by: $E_{(p,q)}(a) = (a \bmod p, a \bmod q)$. On encrypted texts as well, the modulo m +,-,? will be performed, in this case on every component. From a pair of numbers, (a mod p, a mod q), we obtain the original a by means of the Chinese remainder theorem, that is the decryption process is the application of the Chinese remainder theorem.

Chinese remainder theorem

If $m_1, m_2, \ldots, m_k$ are pairwise coprimes then the system of congruences

$$x \equiv a_1 \ (m_1)$$
$$x \equiv a_2 \ (m_2)$$
$$\ldots$$
$$x \equiv a_k \ (m_k)$$

has a unique solution mod $m_1 \_ m_2 \_ \ldots \_ m_k$.

Example:

$$x \equiv 1 \ (3)$$
$$x \equiv 3 \ (5)$$
$$x \equiv -1 \ (7)$$

We search a solution of form $x = 35y_1 + 21y_2 + 15y_3$. Then from the system of congruences

$$35y_1 \equiv 1 \ (3)$$
$$21y_2 \equiv 3 \ (5)$$
$$15y3 \equiv -1 \ (7)$$

A solution of it is:
$y_1 = -1$, $y_2 = 3$, $y_3 = -1$, from which $x = 13$ follows.

We will show the function of the homomorphism through a numeric example, and as we are interested in the demonstration, we had to decide on desperately small parameters. The secret key will be (3, 11), $m = 33$. Suppose that the formula to be computed is (12+4)×2 (with modulo 33 operations), thus the three plaintexts are 12, 4, 2, and the expected result is 32 mod 33. We send the encrypted plaintext in the following form: $E_{(3,11)}(12) = (0, 1)$, $E_{(3,11)}(4) = (1, 4)$, $E_{(3,11)}(2) = (2, 2)$. Of course, we also communicate to the entity performing the computation that we would like to multiply the sum of the first two members with the third factor. Accordingly, they compute the value of ((0, 1)+(1, 4))×(2, 2), where these operations are modulo 33 operations taken by every component. The computation goes like this: ((0, 1)+(1, 4))×(2, 2) = (0+1, 1+4)×(2, 2) = (1, 5)×(2, 2) =(2, 10). We obtain the result in the encrypted form: (2, 10), which we decrypt by means of the Chinese remainder theorem. We thus have to solve the congruence system $x \equiv 2 \pmod 3$, $x \equiv 10 \pmod{11}$. We are searching for the solution in the form $x = 3y + 11z$. As $3y$ is divisible by 3, the congruence $11z \equiv 2 \pmod 3$ must be fulfilled. For the same reason, $3y \equiv 10 \pmod{11}$. Hence $z = 1$ and $y = 7$, respectively, that is $x = 3 \times 7 + 11 = 21 + 11 = 32 \pmod{33}$.

Therefore, the privacy homomorphism becomes specifically useful for computing and data delegation problems because the “encrypted operations” corresponding to the “plain operations” can be performed on the encrypted text corresponding to the plaintext, and we obtain the appropriate “plain result” by deciphering the encrypted result thus achieved.

Why is a privacy homomorphism good?

First of all, a privacy homomorphism must be appropriate for the given application, that is it must preserve those operations, which are necessary from the point of view of the application. In addition, we can raise requirements basically on the *efficiency* and the *security* of the privacy homomorphism. These requirements include the ease of computation of the function itself, and the same must be valid also for the deciphering. A further requirement on the efficiency is that the computation of the f'_i-s and p'_i-s corresponding to the operations and predicates, respectively, of the plaintext must be performed quickly, and that the encumbrance of the encrypted text must not be much more than that of the plaintext.

From the cryptographic perspective, the question of security is more exciting. For this examination, we must overview the (passive) attacks that can be carried out against a cryptosystem.

Attacks

As far as privacy homomorphisms are concerned, we can mention essentially three kinds of attacks. All three are passive in the sense that the attacker tries not to

modify data or to make the computations physically impossible, but intends to get valuable information through the possession of certain data with respect to the plaintext, the encrypting algorithm or the key.

As we mentioned earlier, the security of the privacy homomorphism, but also that of any encryption system shall not depend on whether the attacker knows the encryption algorithm or not. The key used for the encryption is of course secret, but for the algorithm (or mapping) we always suppose that it is publicly accessible (though we do not explicitly intend to provide this access).

In the simplest case, the attacker has the least standing-ground to rely on, so he is facing the most complex task. Texts encrypted with the same key K are available for him; hence this attack is called *ciphertext only* attack. The attacker seeks then to abduct as much information as possible from this "knowledge base" with respect to the plaintext and the key used.

In the case of *known plaintext* attacks the attacker tries to get information related to the key K (and to the encrypting algorithm) based on plaintext-ciphertext pairs $(a, E_K(a))$.

The attacker possesses the most information in the *chosen plaintext* or *ciphertext* attacks. In this case, the attacker has the option to decide on the plaintext or ciphertext, and ask for the corresponding ciphertext or plaintext pair, and accordingly tries to find out the key and the encrypting algorithm.

The complexity of the attacks decreases in the above sequence, as the starting information available for the attacker is increasing. We can also say that the potency of the attacks is increasing in the sequence, because the attacker becomes stronger with a steadily growing arsenal. As regards, however, the preparation for the attacks, an inverse classification can be of use, as the attacker only has to tap the communication channel to execute the ciphertext only attack, while the chosen text attack needs obviously the addition of some other practices.

We define the safety of a privacy homomorphism in a "natural" manner: a homomorphism is safer if it resists stronger attacks. Generally speaking, the higher a privacy homomophisms's security level is, the more useful it is. But the necessary security level is greatly depending on the given application. The data delegation requires a higher security degree than the computing delegation. In the latter case, the data manager receives the encrypted text, performs the necessary computation, and sends back the encrypted result. The data owner carries out the coding and decoding of the data, and after decoding it, the communication with respect to the relevant data will not continue. Thus it is obvious that the data manager gets into contact only with encrypted texts, and therefore he can launch ciphertext only attacks, which is the weakest among all attacks. This attack can be made even weaker, if the data owner uses a different key for encoding for every single computation needed.

When data delegation is performed, returning the encrypted result does not terminate the communication between the data owner and the data manager, according to the fact that the data manager claims the decoded result. Accordingly, the data owner returns it to the data manager after decoding the encrypted result. In other words, the data manager will possess (plaintext, ciphertext) pairs even when the operation is normal, so that the premises are given for starting a known plaintext attack. Changing the key frequently is a useful tool to weaken the attacks in this case as well.

Attack against the RSA-related homomorphism

The RSA-related homomorphism can be broken, i.e. the values of p and q can be restored, by a known plaintext attack. The attack is performed as follows: The plaintexts $M_1, M_2, \ldots M_r$ as well as their encrypted correspondents $E_{(p, q)}(M_i)=(C_i, D_i)$ are available, with $M_i=C_i \bmod p$, $M_i=D_i \bmod q$ and $i=1, 2, \ldots r$.

According to the definition of the congruence, it is obvious that $p|C_i-M_i$, $i=1,\ldots,r$. Let us take the greatest common divisor of these differences, and let it be $p'=\gcd\{C_i-M_i : i=1,\ldots,r\}$. Similarly we define q'. As p and q, respectively, are divisors of the differences C_i-M_i and D_i-M_i, they are also the divisors of their greatest common divisor (p' and q', respectively), i.e. $p|p'$ and $q|q'$. Even for a small r there is a high probability of $p'=p$ and $q'=q$. If this is not the case, every new $(M, (C, D))$ pair lets the attacker come closer to the secret primes (i.e. the key) by introducing $p''=\gcd(C-M, p')$ and $q''=\gcd(C-M, q')$, respectively.

The generalization of the homomorphism presented in [4] permits to defend fortunately this kind of attacks. In addition, the RSA-related privacy homomorphism is so "successful", that this will be used as a starting point for the solution of a data delegation system. The next section shows more details on the results.

What is known about privacy homomorphisms?

The possibly most restricting feature in the use of privacy homomorphisms had been detected as early as at the beginning of the researches. If the person performing the computation tasks has an option to encrypt any constant, and is able to compare encrypted texts by means of the predicate $\leq$, then the privacy homomorphism is not secure, so that it can not resist even the weakest attack, i.e. the ciphertext only attack. In fact, the value of $a'=E_K(a)$ can be "caught" through a binary search, thus the value of a can be easily defined.

To start the search, the encrypted constants are needed, e.g. it may be known that $E_K(1)=1'$. Hence, due to the operation preservation, $2'=E_K(2)=E_K(1+1)$

$= E_K(1)‡E_K(1) = 1'‡1'$ can be calculated. + denotes the operation which the homomorphism will preserve, and ‡ is its equivalent in the image space. This process can be continued till the comparison lets us find a $2'^n$ where $2'^{n-1} \leq a' \leq 2'^n$. While continuing the search, we verify whether $a' \leq 2'^{n-1}‡2'^{n-2}$ is fulfilled. If so, we continue the search in this half of the interval by applying the bisectional method, i.e. by verifying the $a' \leq 2'^{n-1}‡2'^{n-3}$ condition; if not, we continue to bisect the other half of the interval and verify the fulfilment of the $a' \leq 2'^{n-1}‡2'^{n-2}‡2'^{n-3}$ condition.

At the end of the search, we will have the sum of the encrypted powers of two: $a' = 2'^i‡2'^j‡…‡2'^m$. That is $a' = E_K(2^i)‡E_K(2^j)‡…‡E_K(2^m)$, which means, due to the operation preservation, that $a' = E_K(2^i+2^j+…+2^m)$, i.e. $E_K(a)=E_K(2^i+2^j+…+2^m)$, or in other terms $a = 2^i+2^j+…+2^m$. The right-hand side of the equation can be calculated, so also the value of a can be defined.

It has been proven in [1] that the additive privacy homomorphisms cannot resist the chosen ciphertext attack. Note that for additive homomorphism, the set of ciphertexts can be regarded as a vector space above the body {0, 1}. Let us select a based in this vector space (for example the encrypted "powers of two"), and ask for the parent image of it, i.e. its plain version. Thus the following pairs are available: (1, 1'), (2, 2'), (2^2, $2'^2$),... Therefore, if a given $a' = E_K(a)$ encrypted text is supposed, based on which we want to define a, then logically a' can be written as a sum of the encrypted powers of two, i.e. in the selected base: $a' = 2'^i‡2'^j‡…‡2'^m$. From this point on, the attack can continue in the same way as in the case mentioned before.

The restriction of additivity outlined in [3] has the goal to defend this attack. The r-additive privacy homomorphisms permit the summation of only r members. So, if r is sufficiently small and the set of the ciphertexts is sufficiently large, the above-mentioned attack does not work, and the homomorphism is secure for ciphertext only attacks.

The [4] shows a generalization of the RSA-based homomorphism presented. Even in its original form, the homomorphism has preserved addition and multiplication, but as we have seen, it has not been secure against a known plaintext attack. The point of the generalization is that the former encrypted text (a mod p, a mod q) appears in a more complex form, so there is no possibility to set up the statements relating to the divisibility.

The result shown in [5] presents a breakthrough in the number of the operations preserved. Note that this is a privacy homomorphism, which preserves all four field-operations (+,-,?,/) and is at the same time resistant against the ciphertext only attacks. However, as the results highlighted in [1] say, the homomorphism of such type is able to achieve a security level where it can stand also against known plaintext attacks. Finally, [2] has included an advanced version of the additive and multiplicative homomorphism introduced in [4], which already facilitates the division as well and, at the same time, is provably a secure option in the case of the known plaintext attacks as well.

Based on this homomorphism, the prototype of a system delegating sensitive statistical data has been created and standardised by the authors as a "Method for secure delegation of statistical data" [7], so we have no further information on the features of the system.

The above-mentioned homomorphism is a useful tool for any computation, which needs field-operations (+,-,?,/) only. More complex financial or engineering computations may however require additional operations as well, for example logarithm and exponentiation.

An additional, interesting question is whether the personal income tax can be solved by means of the privacy homomorphism. Of course, the relevant tax calculation software tools can be downloaded from the Internet by means of which the most appropriate declaration version can be found out under the shelter of one's own computer, but an application might be more comfortable that could permit to specify in on-line mode the imposed sum of the tax, based on different input data. But because the user would certainly experiment with a series of values, he would not willingly send out such data in a form ready for processing.

Logarithmic calculation with privacy mapping

The logarithmic calculation can be solved by a privacy mapping operation, which, strictly speaking, cannot be regarded as a privacy homomorphism. Let R^+ be the set of plaintexts and ciphertexts, i.e. the set of positive real numbers. Let the key be an arbitrarily chosen positive integer r. The encrypted version of an $a \in R^+$ is $E_r(a) = a^r$. The decryption is not exactly the inverse of E, as we presume that we will have also a logarithmic calculation between the encryption and the decryption. The deciphering of a ciphertext a' is accordingly delivered by $D_r(a') = a'/r$.

It is easy to verify that if we decrypt the logarithm of the encrypted text, we obtain the logarithm of the plaintext, because $(\log(a^r))/r = (r \times \log(a))/r = \log(a)$. It can however be stated that this is not a classical privacy homomorphism, as in this case, the logarithmic operation interpreted on the set of plaintexts is the same as its counterpart interpreted on the set of the ciphertexts, thus the following equality will not be fulfilled: $E_r(\log(a)) = \log(E_r(a))$, as $(\log(a))^r \neq \log(a^r)$.

Impediments of the tax calculation

An on-line income tax calculation service is burdened with the problem of the computing delegation. In this section we outlined some concepts that could be of use for the exploration of an appropriate homomorphism.

This service could be operated for example by an accountant organisation addressed by the tax-paying entity ordering the calculation of the most advantageous taxation form. The taxpayer is clearly not in the possession of the extensive knowledge required for this work, and, on the other hand, he would not willingly expose his financial situation for the service provider.

Once the aspects able to influence the tax base and the tax extent taken into account, the calculations will land sooner or later at a point where we have to define from the tax schedule the sum of the tax relating to the income. Because the tax function is linear but changes interval by interval, ranging the income between the relevant limits must precede the definition of the function value. For this purpose, we need the encrypted version of the end points of the intervals and also an encrypted comparative predicate. The tax function can be supposed to be known for the service provider, and if so, the text pairs of the end points of the interval (plain, encrypted) are available for a possible attack.

Under such conditions the privacy homomorphism will not be secure, because a searching attack can be carried out by means of the comparison. Let us suppose that the encrypted image of the income representing the tax base is $x'= E_K(x)$. We intend to use the tax function for the calculation of the relevant tax value. First we shall define the tax range or interval in which x' lies. The breakpoints of the tax function are 0, a, b, c, and their encrypted correspondents are $0'= E_K(0)$, $a' = E_K(a)$, $b'= E_K(b)$, $c'= E_K(c)$. It can be supposed that $a'\leq x'\leq b'$. Then the value of x can be defined on the interval [a', b'] by means of the binary search described earlier.

But we can also choose the solution according to which we specify the tax schedule individually for each income sum comprised in every interval [1,N]. This means that we define the tax function for each value within a constrained definition domain, of course in an encrypted form. No doubt, the service provider may know the open tax schedule in its present form as well, that is he knows the pairs (x, $t(x)$), with t being the tax function. Although we also make available the pairs (x', $t(x)'$) for him, the data manager will not be able to compose from them the quarts without any additional information. Thus it seems that we did not have shown favour toward the data manager as a potential attacker by ensuring the conditions to perform a known plaintext attack. On the other hand, some apparent efficiency questions are also arising (encryption and transfer of a huge quantity of data) that could impair the application possibilities of the concept.

Summary

The privacy homomorphisms have been introduced in relation to data delegation and computing delegation, but because a great number of Internet-based services are marked by these problems, their application possibility has again come to the fore [8]. The privacy homomorphisms are confined into specific limits, but the performance of the statistically relevant calculations at an unreliable level is ensured by means of a mapping concept being provably secure, which preserves all the four field-operations (+,-,?,/).

At present there exists no convenient homomorphism which could be a viable solution for complex services, such as for example in the case of applications involving logarithm or limit value calculations. Finally we have outlined a few concepts, which could possibly be set as a guide for future researches.

References

[1] N. Ahituv, Y. Lapid, S. Neumann:
Processing Encrypted Data,
Communications of the ACM,
Vol.30, No.9, pp.777–780,
September 1987.

[2] J. Domingo-Ferrer:
A Provably Secure Additive and
Multiplicative Privacy Homomorphism,
Information Security 2002,
Lecture Notes in Computer Science,
Vol.2433, pp.471–483

[3] E. Brickell, Y. Yacobi:
On Privacy Homomorphisms, Advances in Cryptology,
EUROCRYPT '87,
Lecture Notes in Computer Science,
Vol.304, pp.117–125.

[4] J. Domingo-Ferrer:
A New Privacy Homomorphism and Applications,
Information Processing Letters,
Vol.60, No.5, pp.277–282.,
December 1996.

[5] J. Domingo-Ferrer, J. Herrera-Joancomartí:
A Privacy Homomorphism Allowing Field Operations
on Encrypted Data,
Jornades de Matemàtica Discreta i Algorísmica,
Barcelona, March 1998.

[6] R. L. Rivest, L. Adleman, M. L. Dertouzos:
On Data Banks and Privacy Homomorphisms,
Foundations of Secure Computation,
pp.169–179.,
New York, 1978.

[7] J. Domingo-Ferrer, Ricardo X. Sánchez del Castillo:
Method for secure delegation of statistical data,
P9800608 patent in Spain,
December 2000.

[8] C. Boyens, O. Günther:
Trust is not Enough: Privacy and Security in ASP
and Web Service Environments,
Advances in Databases and Information Systems,
6th East-European Conference, ADBIS 2002,
Lecture Notes in Computer Science,
Vol. 2435, pp.8–22.

Nano Technology Conference

DR. LAJTHA GYÖRGY

lajtha.gyorgy@ln.matav.hu

Keywords: microbiological devices, gene technology, industrial nanostructure materials

Information technology and telecommunications have ever been playing a leading role at the introduction of earlier new technologies, which allowed us to think that we must take again a major part in the proliferation of this new technology. With this concept in mind, the International Council for Communications and Information Technology (NHIT) and PKI Telecommunications Development Institute (PKI) have teamed up to deliver an overview of the development stage of this technology and the perspectives it promises for the different sciences.

It was just ten years ago that the first publications were being issued on the nano technology subject. The importance of this field was then far from obvious. But since the turn of millennium, the number of the application fields of this science is growing constantly. It is now clearly recognized that this cannot be regarded as an advanced branch of microelectronics.

In fact, in the micrometer order of magnitude, the fundamental material units remain unchanged and the different devices scaled in the micrometer range are based on the macro metric behaviour of the material. Although the last thirty years have seen the permanent decrease of the dimensions to achieve by now a few micrometers for the individual devices and a fraction of them for the wiring itself, this decrease has not entailed any fundamental changes. Manipulations ranging down to the order of magnitude of nanometers are already piercing into the interior of the cells and molecules, altering the crystalline structure and the internal links. This also spells the modification of the characteristics of the material.

Early 2003 we started to make arrangements for the establishment of an "educational" conference. It came shortly to light that we could not have a leading role in this process, as physicians, chemists and technologists had already unveiled actual results in this field.

We gradually became acquainted with the research nodes where significant experiences had been accumulated. Ranked first among them are the different institutes of the Hungarian Academy of Sciences, but we have been working intensively also at several universities to not only research this field but also to transfer learning about this subject. Thus we deemed necessary to get together a more general organizational team. The members of it are *József Gyulai* (HAS Research Institute for Technical Physics and Material Science), *Pál Bárczy* (Miskolc University, Institute for Material Science), *Zsuzsanna Mokry* (Ministry for Education), *Miklós Zrínyi* (Budapest University of Technology and Economics, Department of Chemical Engineering), *Sándor Imre* (BUTE, Department of Electrical Engineering), *Tamás Roska* (Péter Pázmány Catholic University and HAS, Computer and Automation Research Institute). We succeeded in getting the patronage of *Norbert Kroó* (Secretary of HAS), *Elek Straub* (Chairman-Chief Executive Officer of Matáv Rt.) and *László Pap* (Vice-Rector of BUTE).

The members of the organizational team undertook to set up a section for each field. These efforts have borne their fruit, as the initially half-day conference evolved gradually to a whole-day, then to a one and a half day, and finally to a two-day seminar with professional lectures read all over the time. Late 2003, being knowledgeable about the program, we concluded to organize the conference from 26 to 27 April, 2004. On the first day, the Hall of the Academy, on the second, the Tölösi Hall of Matáv Rt. Headquarters have been the locations where the audience of more than 200 could benefit from the lectures. The two co-chairmen of the conference have been *Ernő Simonyi* (NHIT) and *Vilmos Koralewsky* (PKI).

Now let us recite some eminent results of the conference. While doing this, we will commemorate some lectures in more details, and the others in brief. This proceeding may however not be deemed a value judgement, as all lectures presented a high professional degree and were delivering valuable results to contribute to the development of the nano science. The sequence of the themes is defined rather by our opinion which point we are stressing, which results may present themselves with innovative solutions to have a widespread use. The representation of all these achievements serves our goal to suggest that the nano technology is not an abstract science anymore, a scientist's toy, but a tool that can be used in numerous areas for the solution of emerging problems and the enhancement of classical methods.

The first results of the nano technology have been the carbonic nano tubes. Simultaneously, such measuring instruments and manipulators have been made that were able to perform the tasks in the nano range. An eminently attractive solution has been the application of the laser beam as a clip and as a motor driving tool. These "light-(laser)made and -operated devices" have been referred to by *Pál Ormos* (Szeged). The results have been quite surprising, and it has been shown which innovative tools completely different from their predecessors are able to solve specific tasks in the ranges invisible to the naked eye and hitherto unmanageable by any known devices. The HAS research institutes have also informed the audience on a great number of novelties. We also heard a report on the definition of high-resolution surfaces, on the creation of new material features and also on the formation of structures with required characteristics. The HAS Technical Development Funds have been represented by *József Gyulai*, academic, who held an introduction into the basics, and it has been clearly testified that in Hungary there exists an institute that can solve the emerging tasks by means of tools of the nano technology. HAS, itself, has been carrying out in-depth researches into the possibilities of GaAs/Au compounds so highly relevant in telecommunications *(Imre Mojzes)*.

Biological applications are also spreading out towards divers facilities. Let us just start with the statement, that we can control the aging processes on the basis of the approach from the nano technology, thus restricting the complaints and disabilities of the elderly to an acceptable level *(László Iván,* Semmelweis University). *László Lázár,* from the Hospital of Nyíregyháza, exposed, that backbone operations that had been regarded several years ago extremely hazardous and capital, could be radically changed. By exploring the problems by means of the tools of nano technology and by designing and manufacturing the necessary prosthesis for the healing in advance, they perform the operations using a minimum of cuts. As soon as 2 to 3 days after the op the patients are released from the hospital. In the biological section, researchers of the Biological Research Institute of Szeged published their experiences in gene technology and nano technology of bacteriology. Gene technology, through manipulating the DNS chips on animals, communicated quite surprising results. Although these have been advantageous in many instances both for the animal and for the progeny, we have concerns whether the human gene modification could be ethically acceptable.

In the section of nano-structured industrial materials, the practically most amazing lecture has been held on the utilization of nano-magnetism for measuring inequalities above rails. After this, the grinding can be performed at the minimum material loss. The measuring results can control the grinder at a precision of the nanometric range. The successive representations have focused on the details of influencing the characteristics of ceramics and metals by means of tools of nano technology. This facilitates both increasing their solidity and modifying the surface hardness, but polymeric composites can also be produced. The universities of Debrecen, Miskolc and Budapest and also the Zoltán Bay Institute have equally proven their skills in producing new materials of nano-structure, fulfilling specific needs.

The Section IV put prominently those fields into the foreground, where Hungarian researchers discovered solutions recognized as world-class innovations. They

spotlighted also a domain holding out much more promises for electronics than ever. By conceptually upgrading nano-structured silicon, these researchers succeeded in locating such applications that are not only smaller but also much more diversified than the previous silicon-based semiconductor solutions. In this section, prominent domestic achievements of manufacturing nano-structured alumino-silicates and metals have also been presented.

Nano technology has unveiled novel solutions for encryption and data protection in the fields of telecommunications and information technology. *Michele Mosca* of Canada recited the security to be achieved based on the quantum algorithm theory and its implementation. It was accompanied by an exposition on quantum cryptography that related the information elements to a photon. It has intended to realize security by the features of the photon, but its practical use is deemed to require further investigations.

The team of *Tamás Roska* (Péter Pázmány Catholic University) has summarized three terrains of nano technology (Nano Bio Info Cogno). By detecting the bioelectrical signals of cerebral processes, it is possible to control directly by thoughts diverse technical operations, machines or any other biological systems. This field is a promising area, with the results presented being near to be handled as a reality, all the same we are accustomed to take such ideas as science fiction. We were however astonished as we could learn from the studies of *György Karmos* and *István Ulbert* that this concept is already a reality. We had equally the sense of some futurology as we heard of the self-organizing nano systems, based on the function of one of the basic elements of the human organism, i.e. of the proteins.

The final section featured several lectures associated with this subject, that surpassed however the experiences of nano technology. *Aural A. Lazar* (Columbia University) presented a novel encrypting principle permitting the encryption and management of the data by means of information of the least amplitude and time intervals not requiring any synchronization. An exciting lecture could also be heard from Tamás Roska on similarly novel computers using the nano technology tools and concepts modelling and detecting the cerebral functions.

Angela Hullmann has been attending the Conference as a special guest, as a representative of the EU Research Development Board. She instructed the attendees of the Conference on the diverse programs, including a detailed presentation about the fields concerned by nano technology and the amount of the programs that are working all around this subject. Numerous have been the programs the Hungarian experts have been participating in, in close cooperation with the most successful researchers of Europe. We could get acquainted with the crucial elements of the next program; this event can be joined by anyone, but individuals are advised to keep in step with the joint development project during the period before 2007, to be familiar with the outputs and working methods of their researching associates, thus having the chance to enter the joint project.

Norbert Kroó (HAS) has been closing the Conference, outlining every future outbreak points of nano technology. He linked the results achieved with the possible effects of the present researches. This rendering helped also to rank properly the scope of telecommunications, which has been hitherto a neglected area. In addition, he has been mentioning strategic tools and diverse materials that are commonly used in our everyday life. Accordingly, the apparition of clothes is imminent in the near future that will be much more resistant in strength, but more casual in wearing. This overview brought altogether the final teaching that nano technology will certainly have a dominant impact on the different scientific areas and services, but its proliferation will take place at different rates. The grounds of this do not lie with the research achievements; the substitution of a lot of areas, as yet, is not justified by either the appearance of novel features or by the commercial constraints.

The foils of the Conference are available on CD (Katalin Forrás, 481-7456), and can be obtained by everyone concerned. Thus we are striving to support the basic goal of the Conference, i.e. to get Hungarian researchers and developers to be familiar with nano technology, and to inspire them to have a conception about it as an alternative possibility.

Let's migrate to ENUM

BALÁZS GÓDOR

godor.balazs@ln.matav.hu

Keywords: addressing, service co-operation, privacy

The development of information technology and telecommunications makes more and more electronic services available. These services (SMS, MMS, e-mail, instant messaging) facilitate an advanced communication between users. MMS, e-mail, fax, etc. define points of presence of a user. These points can be reached via different addressing schemes. If someone wants to present all his/her contact information on the same phisical medium, it is quite inconvenient when one of these information changes. In this case all pieces of the business cards (for example) must be reprinted. Using ENUM, this problem can be solved – one should print only one identifier on the businiess card, that wouldn't change over time. The question occurs, however: If ENUM is such useful, why is it not widely used? Present article tries to answer this question.

The name ENUM is an abbreviation, its meaning is tElephone NUmber Mapping. The first aim with this technology is to translate E.164 numbers into DNS domain names. These names could than be used as keys to search for address information of users in the DNS database. The address information can be an e-mail address, a GSM number, a SIP address, etc. These information are encoded in so-called Naming Authority Pointer (NAPTR) Resource Records in the form of URIs and service descriptions according to RFC 3402 [7].

The ENUM system is not a protocol, it is a convention for the use of a specific set of existing protocols, like:

- using E.164 numbers and the "e164.arpa" domain [4];
- the DNS protocol [5,6];
- using NAPTR Resource Records [7];
- and interpreting the URI results of the NAPTR lookups [8].

Using ENUM it is possible to provide a service where the user has to publish only one communications identifier, which serves as a pointer to all his addresses.

ENUM enables the convergence of conventional telephony and IP telephony and the usage of the well established E.164 numbers to access IP-based applications. This includes the access from terminals with numeric keypads, even from the TDM-based circuit-switched PSTN/ISDN network [3]. Platform independent addressing forecasts the vision of a unified telocommunications system [21].

The two problem statements of ENUM are the following [9]:

1) How do network elements (gateways, SIP servers etc.) find services on the Internet if you only have a telephone (E.164) number.
2) How can subscribers define their preferences for nominating particular services and servers to respond to incoming communication requests?

It should be noted that the introduction of ENUM itself does not require any change to the national numbering plans and will not imply any additional demand of E.164 number resources. However, new services and applications triggered by the availability of ENUM may generate demand for additional numbering resources. [1]

There are three different conceptional approaches to ENUM. These are "user ENUM", "operator ENUM" and "infrastructure ENUM". *User ENUM* refers to the use of the single public ENUM root domain (.e164.arpa). *Operator ENUM* refers to implementations of ENUM typically for large organisations using internal telephone and IP networks, and uses a private implementation of the ENUM principle rather than the public root domain. *Infrastructure ENUM* refers to implementations of ENUM within communications networks for network addressing and routing rather than end-user addressing purposes.

Present article concentrates only on user ENUM issues.

SIP or ENUM?

This question often arises during discussions in connection with VoIP in an explicit or an implicit way. However it is a badly formulated question. There is no answer because the two mentioned technologies are not alternatives of each other. While SIP is a protocol ENUM is rather a convention for the use of a specific set of existing protocols, as mentioned earlier. The reason for the confusion could be the fact that it is possible to initiate calls with E.164 addressing in a SIP server based VoIP system without using any additional system or technology. Only IP network, SIP server and some SIP user agents are nedded. So what is the added value with ENUM?

ENUM provides a global solution for unifying communication identifiers, while the VoIP solution with E.164 addressing and one (or more) call server (SoftSwitch, SIP proxy, etc.) provides an isolated system where one can use E.164 addressing as well. In this case it is the Call Server that makes the mapping (using a mapping table) between application level communication identifiers and E.164 numbers. The problem with this (and the reason why it is an isolated solution) is the propagation of mapping tables between Call Servers. There hasn't been any viable protocol that would solve this problem until recently. Users can reach only those subscribers via E.164 numbers who have registered themselves with the same Call Server. This approach is neither scalable nor universal.

In the ENUM system SIP is not the only protocol that can be used to reach services. There are some Internet Drafts [11] that define a basic set of enumservice descriptions that are intended for use in deployments of ENUM. These descriptions form a set of enumservice registration requests, as laid out in section 3 of [2].

The enumservice names are 'talk', 'voice', 'ivoice', 'video', 'msg', 'fax', 'sms', 'ems', 'mms', 'email', 'chat', 'tp', 'im', 'info', 'web', 'ft', 'srs', and 'all'.

DNS – The only database?

DNS [5,6] is a hierarchically arranged distributed database. It is used mostly on the Internet to do the mapping between IP addresses and domain names. The unit of data in the DNS is the Resource Record (RR). There are several sorts of RRs, but from the point of view of ENUM it is the NAPTR RR (Naming Authority Pointer Resource Record) [7] that is important. It specifies a regular expression based rewrite rule that, when applied to an existing string, will produce a new domain label or Uniform Resource Identifier (URI) [8]. Also the lists of ENUM services belonging to an E.164 numbers are stored here. One can acquire this information by using the method (algorithm) defined in the ENUM RFC [2]. According to this, E.164 numbers must be converted into domain names first in the following way.

Reverse the digits of the E.164 number, put dots between them and append the string '.e164.aprpa' to the end. For example the number +36-1-234-5678 would be transformed into

8.7.6.5.4.3.2.1.6.3.e164 .arpa.

This string can be used as a key to the DNS database to retreive information (records) about available services belonging to number +36-1-234-5678.

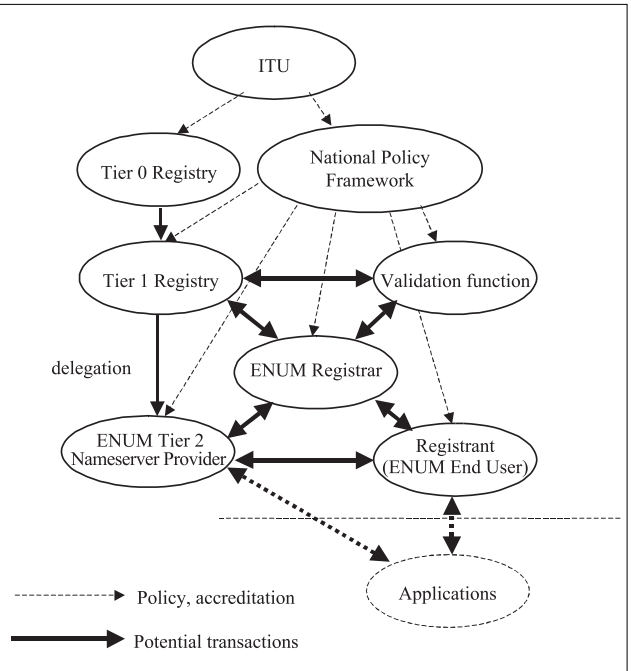
Using the DNS to store service information belonging to E.164 numbers is plausible, because DNS is a publicly available distributed database. The question arises however whether it is an optimal solution or not. There are some imperfections of the DNS that need to be examined thorougly [16]:

- DNS is insecure.
  TSIG [13], DNSSEC [14], PKI [15]
  could make it more secure.
  The question is, how mature are
  these technologies and how big is the gap between
  the theory and the implementation.
- DNS is variably timed.
- DNS is generally not well maintained.
- DNS is generally not well synchronized.
- There is no "DNS says 'no'",
  only an indistinct timeout.
- Putting regular expressions in the DNS is a fascinating complication.

Remark: Under ENUM this article means a system where the root of the hierarchical DNS database is 'e164.arpa'. ENUM systems with different root are regarded as ENUM-like systems.

Abbreviations

DNS	Domain Name System
DNSSEC	DNS Security Extensions
DoS	Denial of Service
E2U	ENUM to URI
ENUM	tElephone NUmber Mapping
ETSI	European Telecommunications Standards Institute
GSM	Global System for Mobile Communication
IN	Inelligent Network
IP	Internet Protocol
ITU	International Telecommunication Union
MMS	Multimedia Messaging Service
NAPTR	Naming Authority Pointer
PINT	PSTN/Internet Interworking Service
PKI	Public Key Infrastructure
PLMN	Public Land Mobile Network
PSTN	Public Switched Telephone Network
RR	Resource Record
SCN	Switched Circuit Network
SIGTRAN	Signaling Transport
SIP	Session Initiation Protocol
SMS	Short Message Service
SPAN	Services and Protocols for Advanced Networks
SPIRITS	Service in the PSTN/IN Requesting Internet Service
SS7	Signaling System 7
TIPHON	Telecommunications and Internet Protocol Harmonization over Networks
TISPAN	SPAN+TIPHON
TRIP	Telephony routing over IP
TSIG	Secret Key Transaction Authentication for DNS
URI	Universal Resource Identifier



ENUM functional model

ENUM Administration

The following *figure* shows the functional model of ENUM. [1]

To understand the tasks of the entities in the figure, some terms must be defined:

ENUM registrar:
entity that provides direct services to domain name registrants by processing name registrations

ENUM registrant:
entity initiating the ENUM registration process (end user or agent)

ENUM Tier 0:
level in the tiered architecture corresponding to the ENUM root, i.e. e164.arpa NOTE: Records at this level contain pointers to Tier 1 for an E.164 Country Code or portion thereof.

ENUM Tier 1:
level in the tiered architecture corresponding to the E.164 Country Code (CC), i.e. <CC>.e164.arpa NOTE: Records at this level contain pointers to Tier 2 for an E.164 number.

ENUM Tier 2:
level in the tiered architecture corresponding to the E.164 number, i.e., <N(S)N>.<CC>.e164.arpa NOTE: Records at this level contain NAPTR records for an E.164 number.

ENUM Tier 2 Nameserver Provider:
entity responsible for the servers within DNS that hold the NAPTR resource records NOTE: In some other documents this entity is also referred to as the ENUM Tier 2 Registry or the ENUM Tier 2 provider. [1]

The ENUM functional and administrative model is based on a ternary separation. The three distinct levels are: Tier 0, Tier 1 and Tier 2.

The main functions performed at Tier 0 level are the administration and technical management of ENUM domain. These functions are implemented by the Tier 0 registry that is a single international registry containing pointers to the Tier 1 registries.

The main functions performed at the Tier 1 level are management and operation of the ENUM domain corresponding to an E.164 country-code in the country or area identified by that given country code. These functions are implemented by the ENUM Tier 1 registry that is a national registry containing pointers to the ENUM Tier 2 Nameserver Providers.

The main functions performed at the Tier 2 level are the commercial provision of the ENUM functions. These functions are implemented by the ENUM Tier 2 Nameserver Provider and ENUM registrar which can be carried out by the same or separate entities. [1]

The 'Validation' entity is responsible for the authentication of users.

According to the 'opt-in principle': The assignee of a number must make an explicit request to participate in ENUM before the ENUM domain corresponding to that E.164 number can be registered and any NAPTR records for the number can be populated. [1]

ENUM risks and threats

Beyond the possibilities of ENUM it should be recognized that this system has some risks as well. These can be the following:

Unscrupolous use of information
ENUM system makes possible that an ENUM client attempting to initiate a call, based on an E.164 number, can retrieve all information about the communication identifiers of the called party. This makes possible to retrieve information on users by entering randomly an E.164 number.

Identity theft, Spamming
The previous point makes possible to build 'identity lists' that can be used for spamming.

Theft of user-provider relations
Provider specific information can be revealed in connection with a user, so one could offer alternative services based on this stolen information.

Denial of Service
'Flood attack' on the DNS NAPTR records can block the retrieval of any communication addresses.

Passing off
Passing off could occur when an entity provisions another end user's E.164 numbers in the DNS by having their own details inserted in the NAPTR records corresponding to another person's or company's number. This would undermine the trust in the ENUM system

Hijacking

"Hijacking" is where a provider of communications applications and services is inserted in a communications path without an end user's permission. In the context of ENUM, hijacking could occur when: a provider of communications applications or services arranges for end users' E.164 numbers to be provisioned in the DNS without their consent. [1]

Related issues (Problems, Protocols, Architectures)

Regarding the cooperation and convergence between IP and PSTN it is necessary to develop new and existing protocols. This section is about protocols and systems aiming to help the IP-PSTN cooperation.

The problem of locating the proper voice gateway is closely related to ENUM. VoIP applications are getting more and more popular that formulates the need for installation of more and more IP-PSTN voice gateways. For calls from the IP network to the PSTN, the caller must locate a gateway that is able to complete calls to the desired destination. There may be several available gateways, and selecting the most suitable one is a nontrivial process. Currently the gateway must be selected by the user or by the signaling servers. The selection and configuration of gateways to use involves manual work.

The list of available gateways must be configured into the signaling servers and updated when new gateways become available [10]. It would be nice to have a protocol that would help users to locate the most appropriate gateway dinamically. One solution can be the TRIP [17,12] protocol to this problem that is under development.

PINT [18] protocol specifies how to reach PSTN services from the IP. There is an other protocol often mentioned together with PINT, this is called SPIRITS [19]. It's RFC describes the architecture for supporting SPIRITS services, which are those originating in the PSTN and necessitating the interactions between the PSTN and the Internet (Internet Call Waiting, Internet Caller-ID Delivery, and Internet Call Forwarding are examples of SPIRIT services).

Specifically, it defines the components constituting the architecture and the interfaces between the components.

An other related framework is the SIGTRAN [20]. SIGTRAN defines an architecture framework and functional requirements for transport of signaling information (SS7, Q.931, etc.) over IP. The framework describes relationships between functional and physical entities exchanging signaling information, such as Signaling Gateways and Media Gateway Controllers. It identifies interfaces where signaling transport may be used and the functional and performance requirements that apply from existing Switched Circuit Network (SCN) signaling protocols.

Summary

As every tehcnology ENUM has some advantages and disadvantages as well. The aim is to make use of ENUM, provide services and convenience for users while minimizing the threats and risks that may arise. To reach this aim, the importance of standards can not be emphasised enough. It is ITU-T Study Group 2 and ETSI TISPAN working group 4 who do the majority of the standardization work.

Beyond standardization it is necessary to build experimental systems as well, to be able to reveal the problems of the system in real-life circumstances. Theory and implementation are often different matters.

The question, why ENUM is not widely used yet, can now be answered. There are some privacy and security issues where the answers are not well elaborated yet. Some of the problems can be solved by the development of the technology while others can be solved by policing. Both approaches are important and can not be neglected. It is better to introduce a good service a bit later than to introduce a bad one immediately.

References

[1] ENUM Admin. in Europe Technical Specification ETSI TS 102 051 V1.1.1 (2002-07)

[2] RFC 2916 E.164 number and DNS, P. Faltstrom. September 2000.

[3] Introduction to ENUM, Document version 0.1 Austrian ENUM trial platform

[4] RFC 3172 Management Guidelines & Operational Requirements for the Address and Routing Parameter Area Domain („arpa"), G. Huston, Ed. September 2001.

[5] RFC 1034 Domain names – concepts and facilities, P.V. Mockapetris. November 1987.

[6] RFC 1035 Domain names – implementation and specification, P.V. Mockapetris. November 1987.

[7] RFC 2915 The Naming Authority Pointer (NAPTR) DNS Resource Record, M. Mealling, R. Daniel. September 2000.

[8] RFC 2396 Uniform Resource Identifiers (URI): Generic Syntax, T. Berners-Lee, R. Fielding, L. Masinter. August 1998.

[9] Implications of ENUM, Geoff Huston September 2002. www.potaroo.net/papers/2002/enum.ppt

[10] TRIP, ENUM and Number Portability, Nicklas Beijar Networking Lab., Helsinki University of Technology http://keskus.hut.fi/opetus/s38130/k01/Papers/Beijar-TripEnumNp.pdf

[11] ENUM Services http://www.potaroo.net/ietf/ids/draft-brandner-enum-services-compendium-00.txt

[12] RFC 3403 Dynamic Delegation Discovery System (DDDS) Part Three: The Domain Name System (DNS) Database, M. Mealling. October 2002.
[13] RFC 2845 Secret Key Transaction Authentication for DNS (TSIG), P. Vixie, O. Gudmundsson, D. Eastlake 3rd, B. Wellington. May 2000.
[14] RFC 3008 Domain Name System Security (DNSSEC) Signing Authority, B. Wellington. November 2000.
[15] Public-Key Infrastructure (X.509) (pkix) Internet draft and RFC collection http://www.ietf.org/html.charters/pkix-charter.html
[16] An IETF view of ENUM, Geoff Huston, Executive Director, IAB http://enum.nic.at/documents/AETP/Presentations/Austria/0011-2003-03-Australia/huston.ppt
[17] RFC 2871 A Framework for Telephony Routing over IP, J. Rosenberg, H. Schulzrinne. June 2000.
[18] RFC 2848 The PINT Service Protocol: Extensions to SIP and SDP for IP Access to Telephone Call Services, S. Petrack, L. Conroy. June 2000.
[19] RFC 3136 The SPIRITS Architecture, L. Slutsman, Ed., I. Faynberg, H. Lu, M. Weissman. June 2001.
[20] RFC 2719 Framework Architecture for Signaling Transport. L. Ong, I. Rytina, M. Garcia, H. Schwarzbauer, L. Coene, H. Lin, I. Juhasz, M. Holdrege, C. Sharp. October 1999.
[21] Egységes távközlés a különböző infrastruktúrájú hálózatokon, Erdélyi Tibor, BME-AUT Híradástechnika, 2004/4.

News

Global Support for Information Society Targets

Targets set for improving access and connectivity to information and communication technologies (ICT) by 2015 at the first phase of the World Summit on the Information Society (WSIS) have received strong support in a global ITU survey. The Summit approved a Declaration of Principles and Plan of Action that set forth a roadmap to bring the benefits of ICT to underdeveloped economies. The Summit was organized by ITU under the patronage of UN Secretary-General Kofi Annan to ensure that social and economic development, which is increasingly driven by ICTs, will result in a more just, prosperous and equitable world. The survey shows overwhelming support for the belief that if the information society is to be one in which all citizens throughout the world can equally access and use information resources for sustainable economic and social development, that cyberspace should be declared a resource to be shared by all for the global public good.

ITU Standard gives operators brighter future

ITU has set a global standard for a new optical fibre that will make it easier for network operators to deploy bandwidth to maximize technology in their core networks.The development of standards in is important if network operators are to reduce costs and provide more innovative services to customers.

G.656 allows operators using CWDM (Coarse Wave Division Multiplexing) to deploy systems without the need to compensate for chromatic dispersion a phenomenon that at low levels counteracts distortion but at high-levels can make a signal unusable. Although it is not complicated, but do not the management of chromatic dispersion is crucial as the number of wavelengths used in WDM increase. G.656 also means that at least 40 more channels can be added to DWDM (Dense WDM) systems. In this case chromatic dispersion generate harmful interference over this – unprecedented – range of the optical spectrum.

ITU-T G.656 (Characteristics of a fibre and cable with Non-Zero Dispersion for Wideband Optical Transport) is the recent in the G-series which specifies the geometrical, physical, mechanical and transmission characteristics of optical fibres. Other Recommendations in this series include:

ITU-T G.652 – Characteristics of a single-mode optical fibre and cable
ITU-T G.653 – Characteristics of a dispersion-shifted single-mode optical fibre and cable
ITU-T G.654 – Characteristics of a cut-off shifted single-mode optical fibre cable
ITU-T G.655 – Characteristics of a non-zero dispersion-shifted single-mode optical fibre and cable

Does the Internet contribute to people's life?

ÁGOTA VISEGRÁDI

visagi@freemail.hu

Keywords: telehouses, distant learning, e-commerce, public web terminal

The information society will be fortified, developed by the generation of the future in the first place, the ones who were born into this world and grew up here. This will be natural for them and they will be the real users and usufructuaries. The spread of computers in the households and at homes will become as general as the television or the telephone. The ones who own a computer links up to the worldwide web as well. Teaching computer science has become accepted at schools by today.

If someone cannot have access to the worldwide web, it is a disadvantage in education even today. The questions of the examinations, the right solutions, the results of the examinees, the set books and the information sources, which are necessary to acquire the syllabus, can all be found on the Internet. When taking a language examination, for example, the examinees get a personal password with the help of which they can have a look at their own result on the homepage of the language school.

The students can get information in wider limits with the help of the Internet as specialized literature can be searched with the help of search engines. Going to libraries is not necessary searching for the required documents. This increases the effectiveness of learning a lot. The equipment with computers at schools is essential as the students do not have the chance to have access to a computer at home, it is the school that provides them with that. There is no doubt about the fact that the students get on with finishing their studies better with the use of the Internet.

The necessity of the Internet also appears in settlements and villages where the equipment with computers and the opportunity to access to the web is limited. The worldwide spread programme is to help that which urges to establish tele-houses, tele-centres and telehuts. Among others, the task of these information centres is to provide the inhabitants of villages with the access to the necessary information. Helping the equal opportunity and their success later on, the opportunity to have access to the worldwide web increases.

Life long learning in the economic competition is essential. A good proportion of the population studies at retraining and further training courses besides doing their daily jobs. The economic development of the society has produced the discontinuation of some recognized jobs so this problem can be solved by the retraining of the unemployed at the given settlement.

Studying in addition to your job reduces free time therefore the information background provided by the Internet is necessary.

I would mention an interesting possibility provided by the tele-education in South Africa where the young African students can listen to lectures of the partner universities through the medium of satellite connections. They can get a degree even at the Harvard University without leaving their own countries and also their cultural and economic surroundings and these degrees are also accepted by the partners cooperating in the network.

Services through the Internet

There are several search engines at the users' disposal to do their purchases on the Internet where the list of stores selling the certain product is put on the screen when the name of the searched product is written into the appropriate place. All the necessary information can be obtained on the homepages of these commercial firms. All the goods can be found in a list, with photos and also the features of the product. The consumer can decide on the way of purchasing the selected product. In the case of an online order the product can be ordered through the Internet and is delivered within one or two days to the given address or the consumer can collect it personally in the specified shop.

Study results certify that 98% of the population capable of earning a living are in the possession of an account. Banks and credit banks offer a certain part of their services online through the Internet. All the financial transactions can be concluded on their homepages.

Internet cafes, Tele-houses, libraries and more recently multimedia terminals sited on public areas (e. g. in shopping centres, in front of municipalities, at railway stations, etc.) can help the ones who do not have computers or Internet accession in their homes or at work. The new service offers an open accession to the worldwide web using a telephone card.

In Hungary the first public WebTerminal in the country was sited in Szeged by the Foundation C3 with the

support of the Matáv. These terminals are suitable for getting information quickly, having a look at any homepage, writing and reading letters.

In the following the appliance developed and installed by the Telecommunication Development Institute of the Post Office Research Institute (PKI) is shown.

The WebTerminal

The terminal is suitable for *browsing the Internet, electronic corresponding, making phone calls* and also *sending short messages* and using *game programmes*. It provides an *on-line advertising interface* which combines the fixed advertisements on the user interface, the advertisements of the Internet sites and the advertising pages running at standstill. The opportunity to advertise and also the amount taken down from the user's phone card mean a considerable source of income for the operator and according to received wisdom, it is necessary by all means to the profitable operation of such terminals.

The services on the terminal (e.g. telephone and browser) can be used at the same time, parallel, in this way the terminal provides the user with a feeling of real multimedia.

Although the terminal is suitable for free of charge distributing of information in the public interest, its most important unique feature is that it is capable of accepting *phone cards,* in this way the customer can pay for the different services with the value taken down from the phone card.

Charges

The tariffs of phone calls are exactly the same as the tariffs of a public phone box. The tariffs of browsing on the Internet work on the basis of time where the appliance charges the tariff by the minute. Sending and receiving electronic letters are charged bit by bit, irrespective of the size of the letter and duration of letter writing. Short messages and played games are also charged bit by bit.

The terminal displays the units taken down in the course of using the services so we can continuously be informed about the amount taken down from the phone card through the medium of a message window.

In the case of the Internet sites of special contents there is a possibility of charging based on the *content* where the amount shown in the head of the site is taken down from the card after the customer confirmed his intention. Databases whose access is restricted may be this type of content where the owner of the database can specify the tariff of the given site.

The way of a more complicated and sophisticated way of charges than now can also be solved. Such as a charge depending on the quantity of the data and the progressive and degressive charges depending on time.

User interface

Developers tended to establish maximal flexibility and convenience of use when creating the user interface of the terminal. The appliance can be used both with a keyboard and a touchscreen. The user can choose the most appropriate way of usage concerning the given activity. The intuitive touchscreen is more suitable for making calls and random Internet browsing and a physical keyboard is more likely to be used than a virtual keyboard when writing a letter or using the Internet. The terminal also contains a built-in webcamera and a microphone. Photos, moving picture and recording are also possible to attach to the written letters with the help of these.

Network connections

The terminal can have three types of network connections which can be combined in four different ways:

– analogue
– ISDN
– ADSL
– ISDN and ADSL combined

The analogue connection represents the past and as it does not provide this type of terminal with the service of a suitable level, it is likely to be dropped by the developers in the future.

The ISDN connection represents the present with the accession the electric circuit connected telephone and the mustered in Internet at the same time. National coverage can be carried out by this connection.

The ADSL connection represents the future. Its large bandwidth guarantees the Internet accession and IP telephone service of, satisfying all the requirements. If necessary the netterminal can access to the worldwide web through wide band ADSL connection. Besides the ADSL accession the ISDN connection is also necessary in future because of the electric circuit telephones and reliability.

Telephone function

The developers decided on the implementation of a clear software ISDN telephone based on a standard PC sound card and an ISDN card. These telephone programmes are given with the ISDN cards by the hardware producers and they can also be loaded from the Internet in the form of a shareware, however, none of them are suitable for using as public phone boxes. The reason for this is that these programmes do not instruct to write cards when calling, do not have telephone number analysis (which is to identify the free/paid/restricted numbers and call directions) and their user interface is also fixed and cannot be form to a similar one of the public appliances in accordance with the requirements of the MATÁV.

Therefore the software developers have made the telephone programme faithfully copying the public appliance with a Proxim card from the very beginning

on the surface of which the usual buttons and functions can be seen:

– volume control,
– selection languages,
– changing cards,
– short time breakdown,
– redialling,
– ABC quick call buttons,
– F1-F5 quick call buttons,
– display of the dialled number,
– display of advertising and information messages.

Apart from these the following functions also work to emphasize the multimedia feature and exploit the ISDN skills:

– loudspeaker,
– conference with three persons,
– display of the number of incoming calls.

The hungarian web terminal developed by László Bortel (Matáv-PKI FI)

The charts of figure analysis, tariffs and advertising messages are provided by the PMS 150 supervisory system.

Connecting of audio signs through the receiver, the internal multimedia speaker and the microphone is carried out with the help of a CNR card .

Browsing function

The browser – depending on setting – can provide a free or a restricted use of the Internet. The whole Internet address or certain groups of sites or the type of content given with the help of extension screening can be restricted. Browsing can also be free of charge or paid. On the surface of the browser the usual buttons can be found (Forth, Back, Refresh, Homepage, Delete, URL line) but there is not a list of menu, in this way activation of the menu buttons endangering safety can be eliminated (e. g. Save, Having a look at the source of a document, etc.)

Operating cards

The terminal accepts phone cards based on Eurochip II now and is prepared to accept cards from abroad that is to accept cards based on Eurochip II of other telecommunication companies – e. g. Deutsche Telekom, SwissCom and PTT Telecom. Above all – due to card operation by the standards PC/SC supported by the windows – other types of cards – e. g. electronic purse, chip bank card or loyalty card – are also easily accepted, it can be solved by the existing card reader installing another safety module into the terminal if necessary.

Special peripherals

– The radar recognizing approach built in the terminal is a special appliance that recognizes moving within the vicinity of a couple of metres of the terminal. The control programme reacts with interrupting the standstill advertising sequence and playing a picture/a sound/an animation, which arouses attention ('Touch the screen.').

– The IrDA infrared data-port makes the changing of data possible by mobile phones, laptops and pocket computers.

– The built in camera is to fix the photo attached to electronic letters, however, it may work as an appliance for video chats and a video telephones later on.

33 appliances of the above mentioned Web Terminals were sited in different parts of the country in December, 2002 and January, 2003.

The appliance has been developed by *László Bortel* (Matáv-PKI; bortel.laszlo@ln.matav.hu) which makes the significant extension of the domestic information services possible. It is likely to advance the use of the Internet on the economically depressed areas.

Equal opportunity, isolation and the Internet

An information centre can be established in community centres and small offices, which can be found at almost all settlements, using the existing telecommunication network and the Weberminal. The bases necessary to start can be established by a contract and with the help of the municipality.

On the basis of the market predictions such agricultural products can be produced that can expect solvent demand. It is also possible to know what tender their economy can be developed and modernized with.

They can join distant education programmes.

Distant jobs are also getting more and more important, which also help the population to remain in their settlements. One of the essential tasks of the telehouses is to create the conditions to do distant jobs for the inhabitants of the settlement (place, device, etc.)

The inhabitants can not only receive information through the medium of the Internet but can also send. Historical background, geographical features, economical, financial and demographical situation of their own settlements can be shown as well. Attention will be drawn and the interest of investors and tourism will also be aroused.

Isolation will come to an end by the Internet and it makes way for equal opportunity. This results in the qualitative change of the inhabitants' lives living there and contributes to the fact that young people will be likely to remain their native places, people will make plans again and have vision of the future. The Internet 'delivers' the opportunities they did not even dare to dream about to their homes.

The condition of spreading of the Internet

There is a certain part of the society that has never used a computer and they do not know how to work with it, they do not know the services and possibilities provided by the Internet. They think they could manage without it so far, there is no forcing impact that would change their standpoint. Children may remove their parents from this deadlock for whom the use of the Internet will be essential to finish their studies and find a job. The computer may not stand on the first place in the domestic budget of these families but sooner or later this investment will be done.

The families who are very hard up may like to have a computer and use the Internet, however, they cannot finance this investment.

The accession through public (WebTerminal) and community centres (tele-houses) shall be provided.

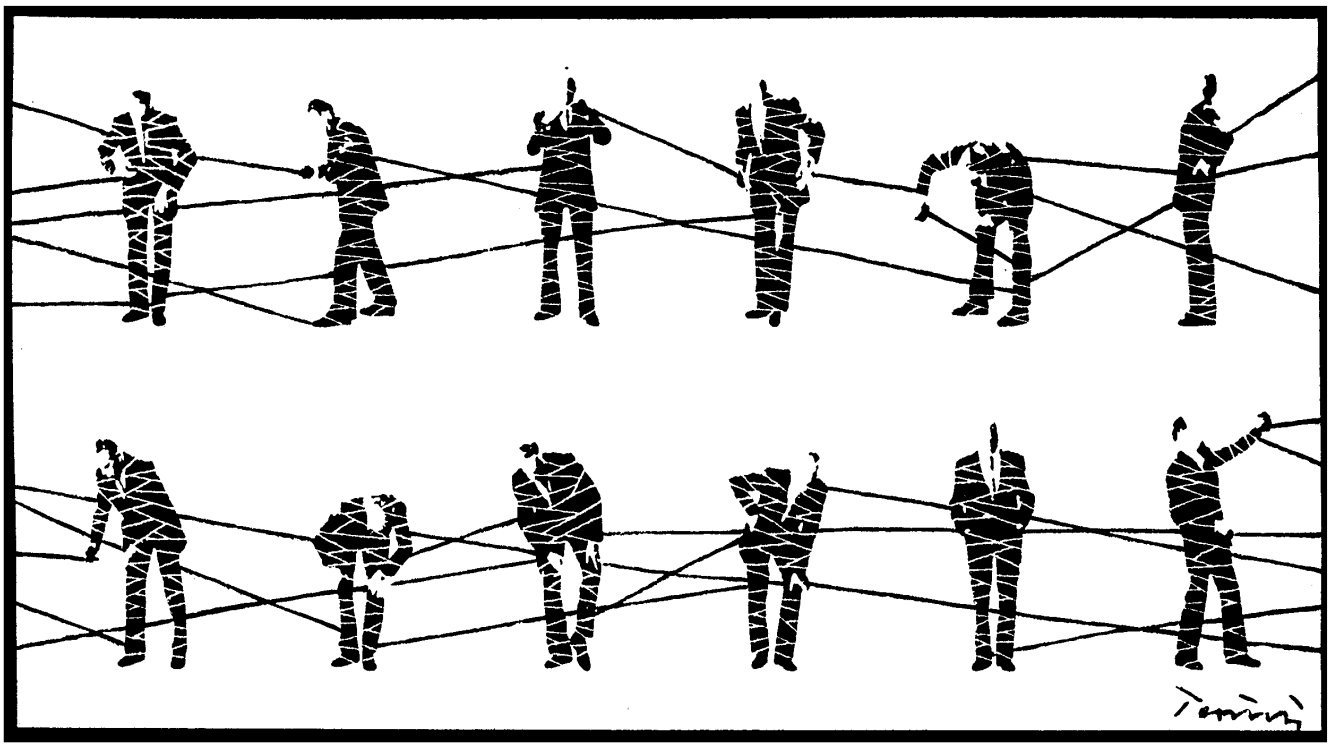
Teaching the use of the Internet free of charge may also contribute to the growth of the use of the Internet.

Analysis of expenses and benefits

As in the case of all investments, all the families and organizations make an analysis of expenses and benefits concerning the use of the Internet. Although this analysis is not always conscious, the advantages and the disadvantages are both examined.

On the side of the cost

a) The value of the investment in *money*: the price of the devices of information technology and the accession to the Internet, furthermore the costs of the

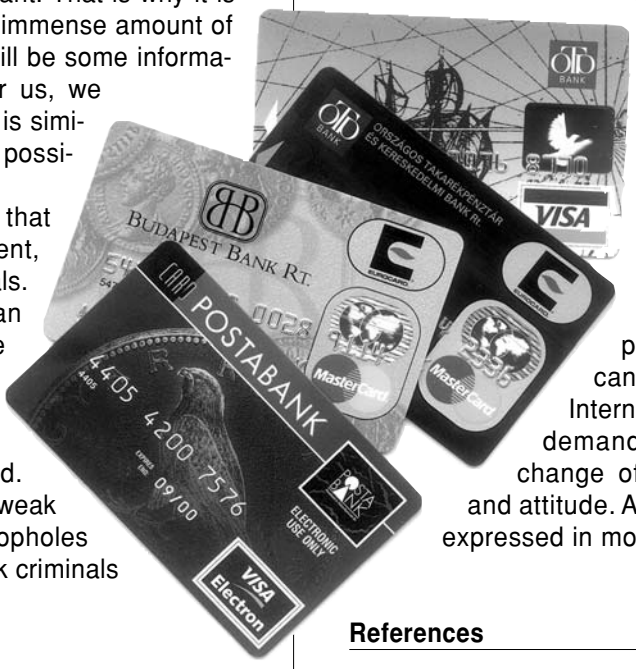


service and its effect on the domestic/organizational budget. What is to reduce or give up.

b) The feeling of *danger* of the user appears in different forms. Such as what happens if a wrong button is pressed? They do not know that a good system asks a question to be confirmed before activation. The systems have an internal safety system.

The language and the system of working up the documents of the search systems do not match. There is no logical connection between the viewpoints of searching and working up. It is proved by the fact that there are too many things found in the case of a search and several of them are irrelevant. That is why it is impossible to read through the immense amount of information, in this way there will be some information that may be important for us, we cannot get them. The situation is similar when using some tightening possibilities.

We often meet with the fact that rules do not protect the innocent, the victim but rather the criminals. Criminals always have human rights, however, these rights are often forgotten in the case of the victims. The appropriate sanctions shall be imposed so that the rules can be enforced. The missing Internet rules, the weak sanctions and the existing loopholes serve the interest of the network criminals only.



Benefit can only be seen

a) When the user is sure that the commercial firm he buys the product from works *honestly* when an online purchase is carried out. A perfect faultless product is delivered to the given address and they do not misuse with the fact that the customer did not see the ordered article.

b) The professional material that can be found on different Internet homepages contain real data, they are *authentic*. Misleading, deceiving the customer shall not happen. Untrustworthy, unreal news may shatter people's trust and then they will read the news on the worldwide web with suspicion.

c) In the case of different Internet bank and financial transactions the customers feel safe.

d) The users are sure that unauthorized persons do not break their secret code.

e) The worldwide web provides *easy access*. One of the main tasks of the tele-houses is to provide the inhabitants with the public and community access and teach them how to use it. The infocommunication background and the access to the worldwide web may be established by the tele-houses. The staff may provide the inquirers with training and help.

Extermination of unnecessary and ineffective information serves the expansion of the effectiveness of the Internet. A sophisticated search language might be a solution so that the unnecessary information does not appear. Sophisticated search tightening of the documents getting into the databases can be perfectly solved with different expressions (e. g. subordination, superiority, etc.), logical – and not accidental – compounds.

If the user orders a cinema, a theatre or a concert ticket, it would be convenient for him to collect the tickets 10 minutes before the performance. It would be necessary to solve the problem of electronic payment in order to ensure the entertaining or cultural institution that the tickets will be paid for sure and taken over. On the other hand, the customer shall be provided with a certain period of time up to the end of which the ticket order can be cancelled and resalable.

The listed difficulties, which put a check on the Internet, can be ceased. The use of the Internet will appear as a natural demand in the course of the change of people's way of thinking and attitude. After that the benefit may be expressed in money besides the expenses.

References

[1] Vilmos Bognár, Zsuzsa Fehér, Csaba Varga (editor): What is future? OMFB, ORTT, HÉA Strategy Research Institution, Budapest, 1998.
[2] Mátyás Gáspár, Andrea Wesselényi, Gyõzõ Kovács: Tele-houses and distant jobs in Hungary, Tuff Produkció Bt., Budapest, 1999.
[3] Dr. György Lajtha: Tele-houses (Teleházak) Magyar Távközlés, 1997/10, pp.1-2.
[4] László Bortel: Development of the public card Internet terminal PKI közlemények, Budapest 2002., 46., p.191.
[5] Zoltán Csörgõ, Csaba Varga: Intelligent regions in Hungary (A falu) 2001. Summer, pp.23-32.
[6] http://nws.iif.hu/NwScd/docs/eloadas/70/
[7] http://www.origo.hu/techbazis/internet/

SHORT IMPULSE PROPAGATION IN WAVEGUIDES

Keywords:
Maxwell-equations, radio waves, ionosphere, reflections

One of the most important research topics is the investigation of (short) impulse propagation in waveguides. The known solutions are based upon the well-known monochromatic approaches, examining the different frequencies separately or building the model and the theory on a fundamentally monochromatic starting point (e.g. permittivity tensor, which is defined originally by assuming an type solution form).

In this paper a completely new theoretical model and solving method will be presented for a rectangular waveguide filled by vacuum, excited by an arbitrarily formed electromagnetic signal (Dirac or real, even short impulse).

(from 2004/5)

PACKET SWITCHED OPTICAL ROUTER

Keywords:
WDM, optical packet switching, sub-carrier

Our laboratory is a participant of an international project where an optical core router development and research is made by university and industry members. The developed hardware based on wavelength division multiplex and subcarrier label technology. The concept, the stage of the development and the foregoing results will be demonstrated below very briefly.

(from 2004/2)

WAVELENGTH CONVERTER SOLUTIONS WITH SEMICONDUCTOR OPTICAL AMPLIFIERS

Keywords:
wavelength conversion, cross-gain modulation (XGM) semiconductor optical amplifier

Semiconductor Optical Amplifiers (SOA) offer several possible solutions for wavelength conversion in the optical domain. In this paper different possible solutions are compared, pointing out the advantages and disadvantages of each. Then a simple and at the same time most promising method, the Cross-Gain Modulation (XGM) is investigated through performance measurement results.

(from 2004/2)

APPLICATION AND MODELING OF VCSELS IN DIRECT MODULATED OPTICAL LINKS

Keywords:
VCSEL, direct modulation, dynamic range, laser model

Vertical cavity surface emitting lasers, VCSELs, are very important light sources in optical communications. Their characteristics are very close to high performance edge emitting laser characteristics with low distortion, high modulation bandwidth and high dynamic range. At the same time their price can be an order less thanks to the new technology.

This paper introduces the unique properties and application of these lasers in high speed direct modulated optical links. A novel circuit model is also shown, which is capable to simulate the spatial effects, like diffusion and spatial whole burning, in these lasers.

(from 2004/2)

CRYSTALLINE SILICON SOLAR CELLS WITH SELECTIVE EMITTER AND THE SELF-DOPING CONTACT

Keywords:
light-electricity conversion, solar cells, outlets, efficiency

One of the key questions of our world is the energy-supply of the people while protecting the environment suitable for life. One solution for this problem is the use of renewable energy sources. In Hungary the share of this energy amounts to only a few percents of the total energy-supply. While the using of the biomass and the geothermal springs are only near-term solutions, for long-term the utilizing of solar and wind energy has to be considered.

(from 2004/1)

PARAMETER CONTROL OF LASER BEAMS IN FUNCTION OF THE PATTERN OF MULTILAYER STRUCTURES

Keywords:
lasertechnology, micro-via machining lasers, laser-material interaction

The development of electronic industry is strongly related to the permanently shrinking sizes, a challenge that in many instances cannot be coped with by means of the perfection of traditional technologies.

The laser devices can be a useful means for their substitution, but as we do not know perfectly the processes that occur at the interaction of the laser with the material, we are unable to make the best of the laser technology. This contribution intends to highlight the features of this interaction, the impacts that the pattern applied on the substrate and the machining parameters have on the results of the machining process. We have set us the goal to exploit more efficiently the opportunities presented by the laser-based machining, by changing the laser beam parameters in function of pattern and material.

(from 2004/1)

Summaries • of the papers published in this issue