Acta
Technica
Jaurinensis

Vol. 7., No. 2., pp. 106-113, 2014
DOI: 10.14513/actatechjaur.v7.n2.242
Available online at acta.sze.hu

A Finite Difference Method of High Order Accuracy for the Solution of Two-Point Boundary Value Problems

P.K. Pandey

Department of Mathematics
Dyal Singh College (University of Delhi)
Lodhi Road, New Delhi – 110003, India
E-mail: pramod_10p@hotmail.com

Abstract: We present a new high order finite difference method for second order differential equation $y''(x) = f(x,y)$ subject to boundary conditions $y(a) = \alpha$ and $y(b) = \beta$. The method is based on rational function approximation and its development is based on power series expansions. Under appropriate conditions, local truncation error calculated and order of method estimated six. Our finite difference method leads to nonlinear system of equations. Numerical examples are given to illustrate the effectiveness, efficiency and high order accuracy of the method.

Keywords: differential equations, sixth order method, finite difference method, rational approximations, Troesch's test problem.

1. Introduction

In this paper, we consider numerical solution of the boundary value problem

$$y''(x) = f(x,y) \quad , \quad x \in [a,b] \subset \mathbb{R}, \quad f(x,y), y(x) \in \mathbb{R} \tag{1}$$

subject to boundary conditions $y(a) = \alpha$, and $y(b) = \beta$, where α, β are real finite constants.

The existence and uniqueness of the solution for the problem (1) is assumed. The specific assumption on $f(x,y)$, to ensure existence and uniqueness will not be considered [1, 2].

A second and fourth order three point discretization for problem (1) is well known [2, 3, 4]. A sixth order three point discretization for problem (1) is given by Chawla [5]. These methods based on the polynomial functions, which are normally smooth and sufficient continuous derivatives. These methods are used to solve specific differential equation.

In the present paper, we describe a sixth order finite difference method for differential equation (1). The method is based on rational function approximation and its development based on power series expansions (Taylor and binomial). Our difference method leads to a nonlinear system of equations. A significant feature of our method is that each discretization of the differential equation at an interior grid point is based on three evaluations of $f(x,y)$. In section 2, we discuss the derivation and development of the method. In section 3, under appropriate conditions obtain the local truncation error. In section 4, we consider numerical examples to illustrate the effectiveness and efficiency of the method. We also compare the performance of the present method with the existing sixth order method given by Chawla in [5].

2. Derivation and Development of the method

An idea of rational function approximation method for solution of two-point boundary value problems and development can be found in the literature [6, 7].

Let N be positive integer, $h = (b-a)/N$ and $x_i = a + i.h$, $i = 0\ (1)\ N$. The value of exact solution $y(x)$ at grid point x_i are denoted by y_i, similarly $y_i'' = f(x_i, y_i) = f_i$ etc.

Suppose we have to determine a number y_i, which is an approximation to the value of the theoretical solution $y(x)$ of the problem (1) at the nodal point $x_i, i = 0\ (1)\ N$. Assuming local assumption that no previous truncation errors have been made [8] i.e. $y_{i-1} = y(x_{i-1})$. For obtaining an approximate value y_i, of the theoretical solution $y(x_i)$ of the problem (1), that purpose, following the ideas in [9,10], we propose multistep difference method as,

$$y_{i+1} - 2y_i + y_{i-1} = \frac{h^2\left(a_3\left(y''_{i+1}+y''_{i-1}\right)+a_4 y''_i\right)}{G(x_i+h)+a_0 y_{i+1}+a_1 y_{i-1}+a_2 y_i} \tag{2}$$

where $G(x)$, is an unknown sufficiently differentiable function of step length h and coefficients $a_i, i = 0\ (1)\ 4$, are unknown constants to be determined under appropriate restrictions and conditions. Also we assumed that

$$G(x_i + h) + a_0 y_{i+1} + a_1 y_{i-1} + a_2 y_i \neq 0 \text{ for any } i = 1\ (1)\ N-1$$

Let define a function $F_i(a, h, y)$ and associate (2) with it as,

$$F_i(a,h,y) = \left(y(x_i+h) - 2\,y(x_i) + y(x_i-h)\right) \times \left(G(x_i+h) + a_0 y(x_i+h) + a_1 y(x_i-h) + a_2 y(x_i)\right) - h^2\left(a_3\left(y''(x_i+h) + y''(x_i-h)\right) + a_4\, y''(x_i)\right) = 0 \tag{3}$$

Let us expand $G(x_i + h)$, in Taylor series about point $x = x_i$,

$$G(h) = G(x_i) + hG'(x_i) + \frac{h^2}{2!}G''(x_i) + \frac{h^3}{3!}G^{(3)}(x_i) + \frac{h^4}{4!}G^{(4)}(x_i) + \frac{h^5}{5!}G^{(5)}(x_i) + O(h^6) \tag{4}$$

where G' denotes first derivative of $G(h)$ w.r.t. h, G'' ..... etc.

Expand $F_i(a, h, y)$ in Taylor series about point x_i, we have

$$\left(h^2 y_i'' + \frac{h^4}{12} y_i^{(4)} + \frac{h^6}{360} y_i^{(6)} + O(h^8)\right) \times \left(G(x_i + h) + (a_0 + a_1 + a_2) y_i + h(a_0 - a_1) y_i' + \frac{h^2}{2!}(a_0 + a_1) y_i'' + \frac{h^3}{3!}(a_0 - a_1) y_i^{(3)} + \frac{h^4}{4!}(a_0 + a_1) y_i^{(4)} + \frac{h^5}{5!}(a_0 - a_1) y_i^{(5)} + O(h^6)\right) - h^2 \left((2a_3 + a_4) + a_3 h^2 y_i^{(4)} + a_3 \frac{h^4}{12} y_i^{(6)}\right) = 0 \quad (5)$$

Using (4) in (5) and rewrite it as,

$$h^2(A_0 - 2a_3 - a_4) y_i'' + h^3 A_1 y_i'' + \frac{h^4}{12}\left((A_0 - 12a_3) y_i^{(4)} + 6A_2 y_i''\right) + \frac{h^5}{12}\left(A_1 y_i^{(4)} + 2A_3 y_i''\right) + \frac{h^6}{360}\left((A_0 - 30a_3) y_i^{(6)} + 15A_2 y_i^{(4)} + 15A_4 y_i''\right) + \frac{h^7}{360}\left(5A_3 y_i^{(4)} + 3A_5 y_i''\right) + O(h^8) = 0 \quad (6)$$

where

$$A_0 = G(x_i) + (a_0 + a_1 + a_2) y_i$$

$$A_k = G^{(k)}(x_i) + \left(a_0 + (-1)^k a_1\right) y_i^{(k)} \quad , \quad k = 1(1)5$$

If we equate the coefficient of $h^j, j = 2(1)7$ in (6) both side, we will obtained

$$(A_0 - 2a_3 - a_4) y_i'' = 0$$

$$A_1 y_i'' = 0$$

$$(A_0 - 12a_3) y_i^{(4)} + 6A_2 y_i'' = 0$$

$$A_1 y_i^{(4)} + 2A_3 y_i'' = 0$$

$$(A_0 - 30a_3) y_i^{(6)} + 15A_2 y_i^{(4)} + 15A_4 y_i'' = 0$$

$$5A_3 y_i^{(4)} + 3A_5 y_i'' = 0$$

If we assume that $y_i^{(p)} \neq 0$, $= 1(1)6$, and

$$A_0 - 30a_3 = 0 \quad (7)$$

Thus we obtained,

$$A_0 - 2a_3 - a_4 = 0$$

$$A_1 = A_3 = A_5 = 0$$

$$(A_0 - 12a_3) y_i^{(4)} + 6A_2 y_i'' = 0$$

$$5A_3 y_i^{(4)} + 3A_5 y_i'' = 0 \quad (8)$$

Solving system of equations (7-8), we obtained

$$a_4 = 28\, a_3$$

$$G(x_i) = 30a_3 - (2a_0 + a_2) y_i$$

$$G^{(2m)}(x_i) = -2a_0 y_i^{(2m)} + 3a_3\left(-\frac{y_i^{(4)}}{y_i''}\right)^m, m = 1,2$$

$$G'(x_i) = G^{(3)}(x_i) = G^{(5)}(x_i) = 0 \tag{9}$$

with the help of (4) and (9), from (2), after neglecting $O(h^6)$, we find that

$$y_{i+1} - 2y_i + y_{i-1} = \frac{8h^2(y_{i+1}''+28y_i''+y_{i-1}'')(y_i'')^2}{240(y_i'')^2 - 12h^2\, y_i''\, y_i^{(4)} + h^4\left(y_i^{(4)}\right)^2} \tag{10}$$

If we replace $y_i^{(4)}$ in (10) by second order difference formulae

$$\frac{(y_{i+1}'' - 2y_i'' + y_{i-1}'')}{h^2}$$

we have

$$y_{i+1} - 2y_i + y_{i-1} = \frac{8h^2(y_{i+1}''+28y_i''+y_{i-1}'')(y_i'')^2}{240(y_i'')^2 - 12\, y_i''(y_{i+1}''-2y_i''+y_{i-1}'') + (y_{i+1}''-2y_i''+y_{i-1}'')^2} \tag{11}$$

Using a notation as defined above $y_i'' = f(x_i, y_i) = f_i$, from (11) we get our rational approximation method for numerical solution of problem (1) as

$$y_{i+1} - 2y_i + y_{i-1} = \frac{8h^2(f_{i+1}+28f_i+f_{i-1})(f_i)^2}{240(f_i)^2 - 12\, f_i\,(f_{i+1}-2f_i+f_{i-1}) + (f_{i+1}-2f_i+f_{i-1})^2} \tag{12}$$

A non-linear multistep method is usually designed for dealing with unconventional problems such as initial-value problem, problem with singularities for which classical methods generally performs poorly. It was noticed that some of the non-linear multistep method which designed for singular problems do not perform on non- singular problems [11]. The proposed method (10), a nonlinear multistep method designed to solve unspecific problem (1). For each nodal point, we will obtain the nonlinear system of equation given by [12] or a linear system of equations if the source function is $f(x)$. Thus the obtained resulting system of nonlinear equations (12) is solved by Newton-Raphson method in considered model problems. In numerical section we will see that the proposed method is suitable for both linear and nonlinear problem. The computational results show that proposed method is effective and accurate.

3. Local Truncation Error

It follows from construction of the method (10) that the new method is at least of six orders. In fact if we consider the local truncation error as defined in [12,13], of the proposed difference method (10), may be written as

$$T_i = y_{i+1} - 2y_i + y_{i-1} - \frac{h^2}{30}(y_{i+1}'' + 28y_i'' + y_{i-1}'') \times \left(1 - \frac{h^2}{20}\left(\frac{y_i^{(4)}}{y_i''}\right) + \frac{h^4}{240}\left(\frac{y_i^{(4)}}{y_i''}\right)^2\right)^{-1} \tag{13}$$

Write Taylor series expansion for each term in (13) about point x_i , we have

$$\left(h^2 y_i'' + \frac{h^4}{12} y_i^{(4)} + \frac{h^6}{360} y_i^{(6)} + O(h^8)\right) - \frac{h^2}{30}\left(30 y_i'' + h^2 y_i^{(4)} + \frac{h^4}{12} y_i^{(6)}\right) \times \left(1 - \frac{h^2}{20}\left(\frac{y_i^{(4)}}{y_i''}\right) + \frac{h^4}{240}\left(\frac{y_i^{(4)}}{y_i''}\right)^2\right)^{-1} \quad (14)$$

Assume $h^4\left(y_i^{(4)}\right)^2 - 240(y_i'')^2 < h^2 y_i'' y_i^{(4)} < h^4\left(y_i^{(4)}\right)^2 + 240(y_i'')^2$, then by application of binomial theorem on right side of (14), we have

$$\left(h^2 y_i'' + \frac{h^4}{12} y_i^{(4)} + \frac{h^6}{360} y_i^{(6)} + O(h^8)\right) - \frac{h^2}{30}\left(30 y_i'' + h^2 y_i^{(4)} + \frac{h^4}{12} y_i^{(6)}\right) \times \left(1 + \frac{h^2}{20}\frac{y_i^{(4)}}{y_i''} - \frac{h^4}{600}\left(\frac{y_i^{(4)}}{y_i''}\right)^2 + O(h^6)\right) = \left(h^2 y_i'' + \frac{h^4}{12} y_i^{(4)} + \frac{h^6}{360} y_i^{(6)} + O(h^8)\right) - \frac{h^2}{30}\left(30 y_i'' + \frac{5h^2}{2} y_i^{(4)} + \frac{h^4}{12} y_i^{(6)} + O(h^6)\right) = O(h^8) \quad (15)$$

Thus from (15), we conclude that proposed method (10) has at least six order accuracy.

4. Numerical Experiments

In this section, the proposed method (12) is applied to solve four different model problems. The resulting system of nonlinear equations so obtained solved by Newton-Raphson method. All computations were performed in GNU FORTRAN language, using double precision. Let y_i , the numerical value calculated by formulae (12), an approximate value of the theoretical solution $y(x)$ at the grid point$x = x_i$. The maximum absolute error

$$\text{MAE}(y) = \max_{1 \le i \le N-1} |y(x_i) - y_i|$$

are shown in Tables 1-4, for different value of h ,the step length. Also we have computed y_i , using formula given by Chawla [5] and shown in the tables.

Example 1. Consider the following two point boundary value problem

$$y''(x) = \frac{2}{x^2} y(x) - \frac{1}{x} \quad , \quad 1 < x < 2$$

with the boundary conditions $y(1) = \frac{11}{19}$, $y(2) = 0$. In Table 1, the maximum absolute error presented in exact solution $(x) = (19x - 5x^2 - 36/x)/38$.

Example 2. Consider the following nonlinear two-point boundary value problem

$$y''(x) = \frac{3}{2} y^2(x) \quad , \quad 0 < x < 1$$

with the boundary conditions $y(0) = 4$, $y(1) = 1$ .In Table 2, the MAE presented in exact solution $y(x) = 4/(1 + x)^2$.

Example 3. Consider the following nonlinear two-point boundary value problem

$$y''(x) = \frac{1}{2}(1 + y + x)^3 \quad , \quad 0 < x < 1$$

with the boundary conditions $y(0) = 0 = y(1)$. In Table 3, the MAE presented in exact solution $y(x) = -x - 1 + 2/(2 - x)$.

Example 4. Consider the Troesch's problem [14].

$$y''(x) = \alpha\, \mathrm{Sinh}(\alpha y) + f(x) \quad , \quad 0 < x < 1$$

with the boundary conditions $y(0) = 0$ and $y(1) = \mathrm{Sinh}(1)$. In Table 4, the MAE presented in exact solution $y(x) = \mathrm{Sinh}(x)$, for $\alpha = .5, 1.0, 2.0$.

Table 1. Maximum absolute error $y(x) = (19x - 5x^2 - 36/x)/38$ *in for example 1*

Error		Step length (h)			
		4	8	16	32
(10)	MAE	.13332743(-3)	.88431334(-5)	.46868072(-6)	.62557199(-7)
Chawla's	MAE	.17962958(-5)	.26665235(-7)	.28091184(-7)	.67117142(-7)

Table 2. Maximum absolute error in $y(x) = 4/(1 + x)^2$ *for example 2*

Error		Step length (h)			
		8	16	32	64
(10)	MAE	.14692306(-3)	.65165545(-5)	.24017024(-6)	.47508550(-6)
Chawla's	MAE	.13908767(-2)	.33620209(-3)	.70991518(-4)	.28845704(-5)

Table 3. Maximum absolute error in $y(x) = -x - 1 + \frac{2}{2-x}$ *for example 3*

Error		Step length (h)			
		4	8	16	32
(10)	MAE	.24302900(-3)	.16158820(-4)	.76591971(-6)	.16093255(-7)
Chawla's	MAE	.50297973(-3)	.76657532(-4)	.10819662(-4)	.83681783(-6)

Table 4. Maximum absolute error in $y(x) = Sinh(x)$ *for example 4*

Error		Step length (h), $\alpha = .50$		Step length (h), $\alpha = 1.0$	
		4	8	4	8
(10)	MAE	.14901161(-7)	.29802322(-7)	.95367432(-6)	.59604645(-7)
Chawla's	MAE	.14901161(-7)	.29802322(-7)	.29802322(-7)	.59604645(-7)
Error		Step length (h), $\alpha = 2.0$			
		4	8	16	32
(10)	MAE	.34451485(-4)	.21457672(-5)	.11920929(-6)	No change
Chawla's	MAE	.73480606(-3)	.13351440(-4)	.23841858(-6)	.11920929(-6)

5. Conclusion

A new kind of finite difference scheme is presented for numerical solution of two-point boundary value problems in ordinary differential equation. The local truncation error in the method (10) presented and is of order six. However numerical results only confirm that method has order more than four. The computational accuracy of present method can be improved if $y_i^{(4)}$ replaced by high order difference formulae. However it is satisfactory to obtain a comparable results using only three function evaluation unlike [5]. Our future work will deal with the presence of higher order derivative term in method to obtain the best computational result, the method can offer. Though method developed for solving two point boundary value problems in ODEs, it has good potential for efficient application to many problems in PDEs. The extension of present method to different problem is crucial. Work in this specific direction is in progress.

References

[1] Baxley, J.V.: *Nonlinear Two Point Boundary Value Problems in Ordinary and Partial Differential Equations,* Ordinary and Partial Differential Equations, Lecture Notes in Mathematics vol. 846, pp. 46-54, 1981
DOI: 10.1007/BFb0089823

[2] Keller, H.B.: *Numerical Methods for Two Point Boundary Value Problems*, Blaisdell Publishing Co., Waltham, Massachusetts, 1968

[3] Numerov, B.V.: *A method of extrapolation of perturbations*, Monthly Notices of the Royal Astronomical Society, vol. 84, no. 8, pp. 592-601, 1924

[4] Chawla, M.M.: *A fourth order tridiagonal finite difference method for general nonlinear two point boundary value problems with mixed boundary conditions*, Journal of the Institute of Mathematics and its Applications, vol. 20, no. 2, pp. 83-93, 1977

[5] Chawla, M.M.: *A sixth order tridiagonal finite difference method for nonlinear two point boundary value problems*, BIT Numerical Mathematics, vol. 17, no. 2, pp. 128-133, 1977
DOI: 10.1007/BF01932284

[6] Mickens, R.E.: *Nonstandard Finite Difference Models of Differential equations*, World Scientific, Singapore, 1994

[7] Lambert, J.D.: *Numerical Methods for Ordinary Differential Systems*, John Wiley, Chichester, 1991

[8] Lambert, J.D., Shaw, B.: *On the numerical solution of* $y' = f(x, y)$ *by class of formulae based on rational approximation*, Mathematics and Computation, vol. 19, no. 91, pp. 456-462, 1965
DOI: 10.1090/S0025-5718-1965-0179947-7

[9] Van Niekerk, F.D.: *Non-linear one- step methods for initial value problems*, Computers & Mathematics with Applications, vol. 13, no. 4, pp. 367-371, 1987
DOI: 10.1016/0898-1221(87)90004-6

[10] Okosun, K.O., Ademiluyi, R.A.: *A Three Step Rational Methods for Integration of Differential Equations with Singularities*, Research Journal of Applied Sciences, vol. 2, no. 1, pp. 84-88, 2007
DOI: rjasci.2007.84.88

[11] Abelman, S., Eyre, D.: *A numerical Study of multistep methods based on continued fractions*, Computers & Mathematics with Applications, vol. 20, no. 8, pp. 51-60, 1990
DOI: 10.1016/0898-1221(90)90209-3
[12] Jain ,M.K.: *Numerical Solution of Differential Equations (2nd ed.),* Wiley Eastern Limited, New Delhi, 1984
[13] Stoer, J., Bulirsch, R.: *Introduction to Numerical Analysis (2nd ed.),* Springer-Verlag, Berlin Heidelberg, 1991
[14] Troesch, B.A.: *A simple approach to a sensitive two-point boundary value problem*, J. Comput. Phys, vol. 21, no. 3, pp. 279-290, 1976
DOI: 10.1016/0021-9991(76)90025-5

Acta
Technica
Jaurinensis

Vol. 7., No.2., pp. 114-122, 2014
DOI: 10.14513/actatechjaur.v7.n2.284
Available online at acta.sze.hu

An Observer-Based *PID* Regulator

L. Keviczky, Cs. Bányász

Széchenyi István University, Győr
Computer and Automation Research Institute and
MTA-BME Control Engineering Research Group
Hungarian Academy of Sciences
H-1111 Budapest, Kende u 13-17, HUNGARY
Phone: +361-466-5435; Fax: +361-466-7503
e-mail: keviczky@sztaki.hu; banyasz@sztaki.hu

Abstract: An equivalent transfer function representation (*TFR*) is used to study the state-feedback/observer (*SFO*) topologies of control systems. This approach is applied to combine this methodology with YOULA-parametrization (*YP*) introducing new classes of regulators. Then this method is used to introduce observer based *PID* regulators.

Keywords: Observer, state-feedback, model error, YOULA-parametrization, PID regulator

1. Introduction

In our previous paper [6] it was shown that in the classical state-feedback/observer (*SFO*) scheme the model error decreases by the sensitivity function of the observer feedback loop. An equivalent transfer function representation (*TFR*) was used to demonstrate this special feature of these regulators. It was also shown that this principle can be used to generalize for the YOULA-parametrized regulators, too. In this paper the demonstrated new approach is used to introduce further new class of regulators.

2. The Observer Based Youla-Regulator

For open-loop stable processes the all realizable stabilizing (*ARS*) model based regulator $\hat{C}$ is the *YOULA-parametrized* one:

$$\hat{C}\left(\hat{P}\right)=\left.\frac{Q}{1-Q\hat{P}}\right|_{\hat{P}\to P}=\frac{Q}{1-QP}=C\left(P\right), \tag{1}$$

where the "parameter" Q ranges over all proper ($Q\left(\omega=\infty\right)$ is finite), stable transfer functions [5], [7], see Fig. 1a.

It is important to know that the *Y-parametrized* closed-loop with the *ARS* regulator is equivalent to the well-known form of the so-called *Internal Model Control* (*IMC*)

principle [7] based structure shown in Fig. 1b.

Q is anyway the transfer function from r to u and the closed-loop transfer function (i.e., *CSF*) for $\hat{P} = P$, when $\ell \to 0$

$$\hat{T}_{\mathrm{ry}} = \frac{\hat{C}P}{1+\hat{C}P} = QP\frac{1+\ell}{1+(1-QP)\ell}\bigg|_{\ell\to 0} = QP = T_{\mathrm{ry}} \tag{2}$$

is linear (and hence convex) in Q.

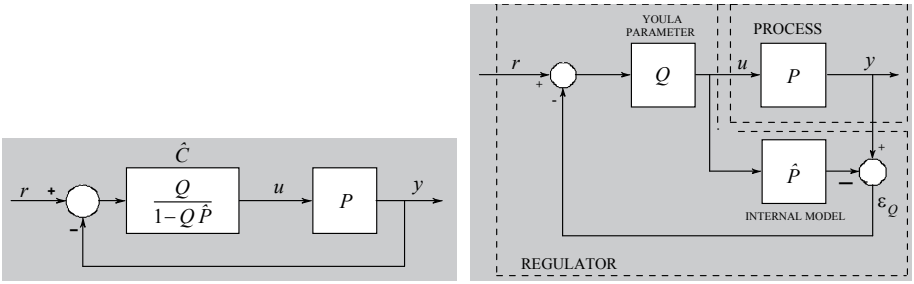


Figure 1. The equivalent IMC structure of an ARS regulator

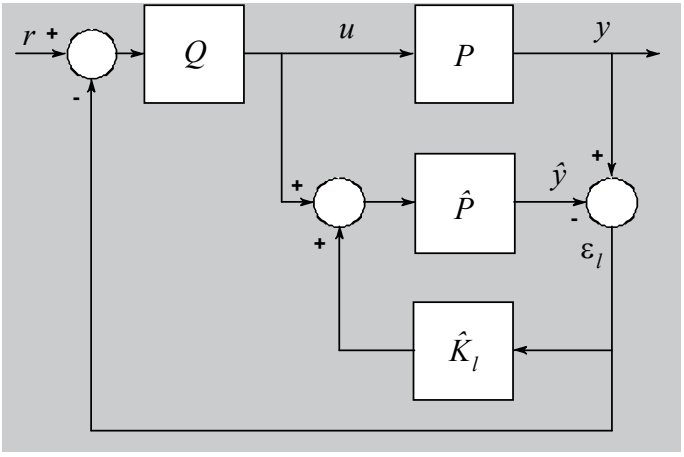


Figure 2. The observer-based IMC structure

It is interesting to compute the relative error ℓ_{T} of $\hat{T}_{\mathrm{ry}}$

$$\ell_{\mathrm{T}} = \frac{T_{\mathrm{ry}} - \hat{T}_{\mathrm{ry}}}{\hat{T}_{\mathrm{ry}}} = \frac{T_{\mathrm{ry}}}{\hat{T}_{\mathrm{ry}}} - 1 = \frac{QP}{\dfrac{QP}{1-Q\left(P-\hat{P}\right)}} - 1 = Q\left(P-\hat{P}\right) = QP\frac{\ell}{1+\ell} = T_{\mathrm{ry}}\frac{\ell}{1+\ell}.$$

(3)

The equivalent *IMC* structure performs the feedback from the model error ε_Q. Similarly to the *SFO* scheme it is possible to construct an internal closed-loop, which virtually reduces the model error to

$$\varepsilon_l = \frac{1}{1+\hat{K}_l\hat{P}}\left(y-\hat{P}u\right) = \frac{1}{1+\hat{K}_l\hat{P}}\varepsilon_Q = \frac{1}{1+\hat{L}_l}\varepsilon_Q = \hat{H}\,\varepsilon_Q\,; \qquad \hat{L}_l = \hat{K}_l\hat{P} \tag{4}$$

and performs the feedback from ε_l (see Fig. 2), where $\hat{L}_l$ is the internal loop transfer function. In this case the resulting closed-loop will change to the scheme shown in Fig. 3.

This means that the introduction of the observer feedback changes the *YOULA-parametrized* regulator to

$$\hat{C}'\left(\hat{P}'\right) = \frac{Q}{1-Q\dfrac{\hat{P}}{1+\hat{K}_l\hat{P}}} = \frac{Q\left(1+\hat{K}_l\hat{P}\right)}{1+\hat{K}_l\hat{P}-Q\hat{P}}\,. \tag{5}$$

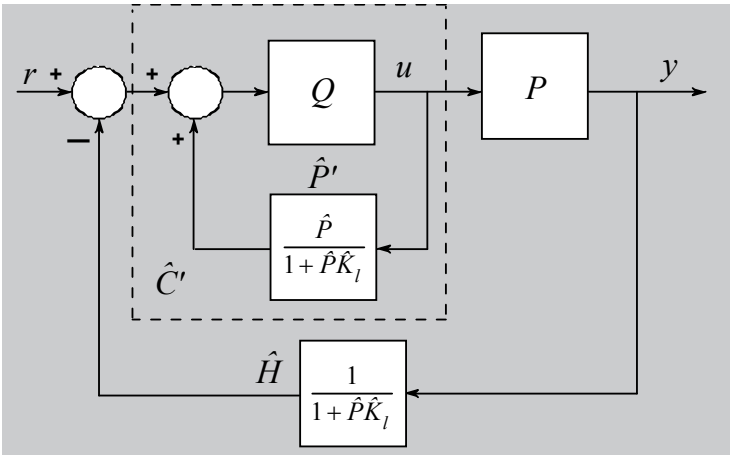


Figure 3. Equivalent closed-loop for the observer-based IMC structure

The form of $\hat{C}'$ shows that the regulator virtually controls a fictitious plant $\hat{P}'$, which is also demonstrated in Fig. 3. Here the fictitious plant is

$$\hat{P}' = \frac{\hat{P}}{1+\hat{K}_l\hat{P}} = \frac{\hat{P}}{1+\hat{L}_l}\,. \tag{6}$$

The closed-loop transfer function is now

$$T'_{\mathrm{ry}} = \frac{\hat{C}'P}{1+\hat{C}'P} = \frac{QP\left(1+\hat{K}_l\hat{P}\right)}{1+\hat{K}_l\hat{P}-Q\hat{P}+QP} = QP\left.\frac{1}{1+QP\dfrac{1}{1+\hat{K}_l\hat{P}}\dfrac{\ell}{1+\ell}}\right|_{\ell\to 0} = QP = T_{\mathrm{ry}}\,. \tag{7}$$

The relative error ℓ'_{T} of $\hat{T}'_{\mathrm{ry}}$ becomes

$$\ell'_{\mathrm{T}} = \frac{T_{\mathrm{ry}}-\hat{T}'_{\mathrm{ry}}}{\hat{T}'_{\mathrm{ry}}} = \frac{T_{\mathrm{ry}}}{\hat{T}'_{\mathrm{ry}}} - 1 = \frac{QP}{\dfrac{QP\left(1+\hat{K}_l\hat{P}\right)}{1+Q\left(P-\hat{P}\right)+\hat{K}_l\hat{P}}} - 1 = QP\frac{\ell}{1+\ell}\frac{1}{1+\hat{K}_l\hat{P}} = \ell_{\mathrm{T}}\frac{1}{1+\hat{L}_l} \tag{8}$$

which is smaller than ℓ_{T}. The reduction is by $\hat{H} = 1\Big/\left(1+\hat{L}_l\right)$.

3. An Observer Based PID-Regulator

The ideal form of a YOULA-regulator based on reference model design [5] is

$$C_{\mathrm{id}} = \frac{\left(R_{\mathrm{n}}P^{-1}\right)}{1-\left(R_{\mathrm{n}}P^{-1}\right)P} = \frac{Q}{1-QP} = \frac{R_{\mathrm{n}}}{1-R_{\mathrm{n}}}P^{-1}\,, \tag{9}$$

when the inverse of the process is realizable and stable. Here the operation of R_{n} can be considered a reference model (desired system dynamics). It is generally required that the reference model has to be strictly proper with unit static gain, i.e., $R_{\mathrm{n}}\left(\omega = 0\right) = 1$.

For a simple, but robust *PID* regulator design method assume that the process can be well approximated by its two major time constants, i.e.,

$$P \cong \frac{A}{\mathrm{A}_2}\,, \tag{10}$$

where

$$\mathrm{A}_2 = \left(1+sT_1\right)\left(1+sT_2\right). \tag{11}$$

According to (9) the ideal YOULA-regulator is

$$C_{\mathrm{id}} = \frac{R_{\mathrm{n}} P^{-1}}{1 - R_{\mathrm{n}}} = \frac{R_{\mathrm{n}} \left(1 + sT_1\right)\left(1 + sT_2\right)}{A\left(1 - R_{\mathrm{n}}\right)}; \quad T_1 > T_2 . \tag{12}$$

Let the reference model R_{n} be of first order

$$R_{\mathrm{n}} = \frac{1}{1 + sT_{\mathrm{n}}}, \tag{13}$$

which means that the first term of the regulator is an integrator

$$\frac{R_{\mathrm{n}}}{1 - R_{\mathrm{n}}} = \frac{\dfrac{1}{1 + sT_{\mathrm{n}}}}{1 - \dfrac{1}{1 + sT_{\mathrm{n}}}} = \frac{1}{1 + sT_{\mathrm{n}} - 1} = \frac{1}{sT_{\mathrm{n}}}, \tag{14}$$

whose integrating time is equal to the time constant of the reference model. Thus the resulting regulator corresponds to the design principle, i.e., it is an ideal *PID* regulator

$$C_{\mathrm{PID}} = A_{\mathrm{PID}} \frac{\left(1 + sT_{\mathrm{I}}\right)\left(1 + sT_{\mathrm{D}}\right)}{sT_{\mathrm{I}}} = A_{\mathrm{PID}} \frac{\left(1 + sT_1\right)\left(1 + sT_2\right)}{sT_1} \tag{15}$$

with

$$A_{\mathrm{PID}} = \frac{T_1}{A\,T_{\mathrm{n}}}; \quad T_{\mathrm{I}} = T_1; \quad T_{\mathrm{D}} = T_2 . \tag{16}$$

The YOULA-parameter Q in the ideal regulator is

$$Q = R_{\mathrm{n}} P^{-1} = \frac{1}{A} \frac{\left(1 + sT_1\right)\left(1 + sT_2\right)}{1 + sT_{\mathrm{n}}} . \tag{17}$$

It is not necessary, but desirable to ensure the realizability, i.e., it is reasonable to use

$$Q = R_{\mathrm{n}} P^{-1} = \frac{1}{A} \frac{\left(1 + sT_1\right)\left(1 + sT_2\right)}{\left(1 + sT_{\mathrm{n}}\right)\left(1 + sT\right)}, \tag{18}$$

where T can be considered as the time constant of the derivative action ($0.1T_{\mathrm{D}} \le T \le 0.5T_{\mathrm{D}}$). The regulator $\hat{C}'$ and the feedback term $\hat{H}$ must be always realizable. In the practice the *PID* regulator and the YOULA-parameter is always model-based, so

$$\hat{C}_{\mathrm{PID}}\left(\hat{P}\right) = \hat{A}_{\mathrm{PID}} \frac{\left(1 + s\hat{T}_1\right)\left(1 + s\hat{T}_2\right)}{s\hat{T}_1}; \quad \hat{A}_{\mathrm{PID}} = \frac{\hat{T}_1}{\hat{A}\,T_{\mathrm{n}}}, \tag{19}$$

$$\hat{Q} = R_{\mathrm{n}} \hat{P}^{-1} = \frac{1}{\hat{A}} \frac{\left(1 + s\hat{T}_1\right)\left(1 + s\hat{T}_2\right)}{1 + sT_{\mathrm{n}}} . \tag{20}$$

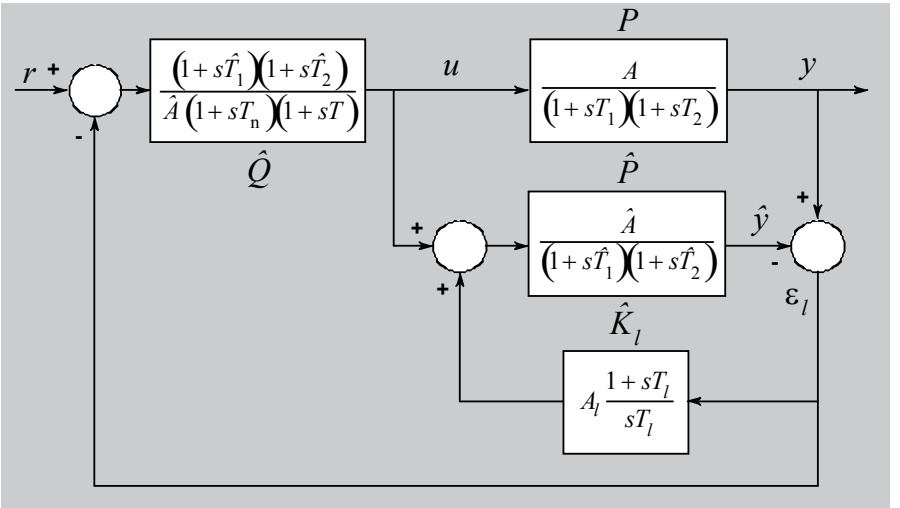


Figure 4. An observer based PID regulator

The scheme of the observer based *PID* regulator is shown in Fig. 4, where a simple *PI* regulator

$$\hat{K}_l = A_l \frac{1+sT_l}{sT_l} \tag{21}$$

is applied in the observer-loop. Here T_l must be in the range of T, i.e., considerably smaller than T_1 and T_2.

Note that the frequency characteristic of $\hat{H}$ cannot be easily designed to reach a proper error suppression. For example, it is almost impossible to design a good realizable high cut filter in this architecture. The high frequency domain is always more interesting to speed up a control loop, so the target of the future research is how to select $\hat{K}_l$ for the desired shape of $\hat{H}$.

4. Simulation Experiments

The simulation experiments were performed in using the observer based *PID* scheme shown in Fig. 4.

Example 1

The process parameters are: $T_1 = 20$, $T_2 = 10$ and $A = 1$. The model parameters are: $\hat{T}_1 = 25$, $\hat{T}_2 = 12$ and $\hat{A} = 1.2$. The purpose of the regulation is to speed up the basic step response by 4, i.e., $T_n = 5$ is selected in the first order R_n. In the observer loop a simple proportional regulator $\hat{K}_l = 0.01$ is applied. The ideal form of Q (17) was used. Figure 5 shows some step responses in the operation of the observer based *PID* regulator.

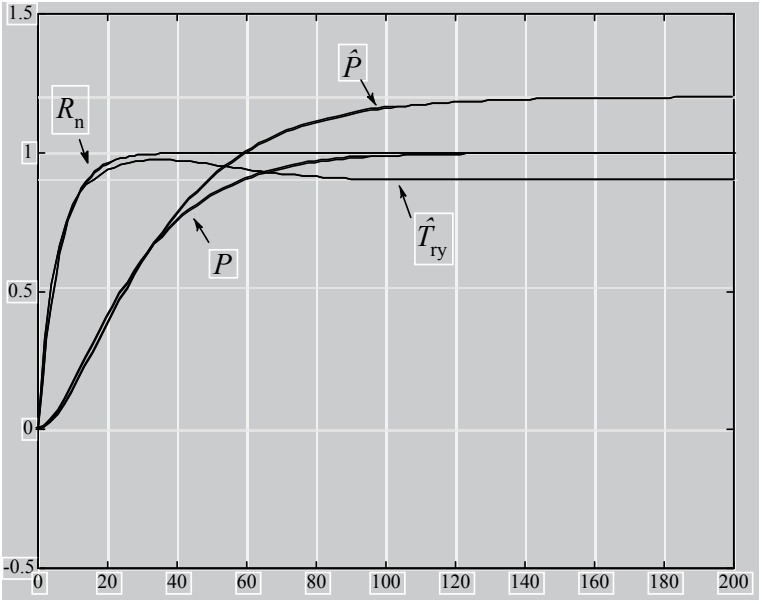


Figure 5. Step responses using the observer based PID regulator

It is easy to see that the $\hat{T}'_{ry}$ very well approximates R_n in the high frequencies (for small time values) in spite of the very bad model $\hat{P}$.

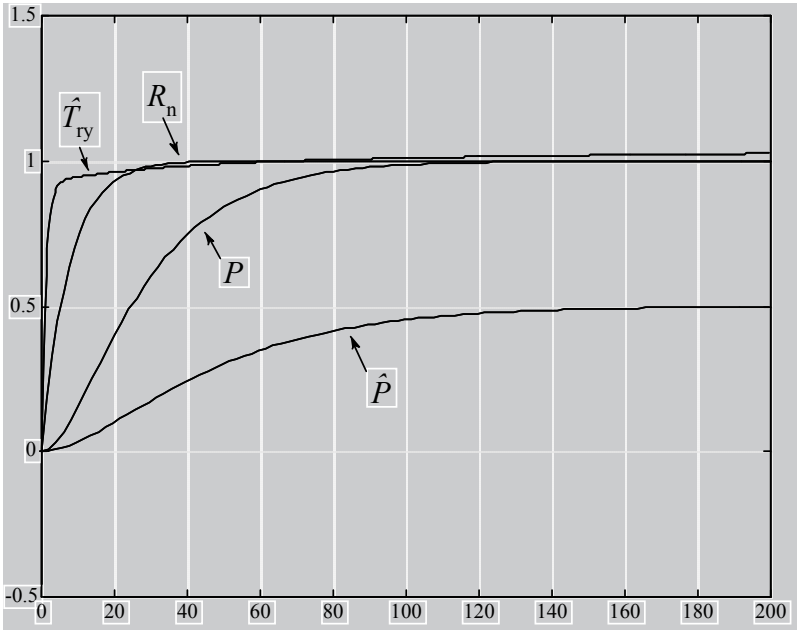


Figure 6. Step responses using the observer based PID regulator

Example 2

The process parameters and the selected first order R_n are the same as in the previous example. The model parameters are: $\hat{T}_1 = 30$, $\hat{T}_2 = 20$ and $\hat{A} = 0.5$. In the observer loop a *PI* regulator (67) is applied with $A_l = 0.001$ and $T_l = 2$. The ideal form of Q (17) was used. Figure 6 shows some step responses in the operation of the observer based *PID* regulator.

It is easy to see that the $\hat{T}'_{ry}$ well approximates R_n in the high frequencies (for small time values) in spite of the very bad model $\hat{P}$.

5. Conclusions

It was shown that the *SFO* methodology can be applied to the YOULA-parametrized regulators, too. This approach reduces the model error by the sensitivity function of the observer loop similarly to the *SFO* scheme.

Using this new approach an observer based *PID* regulator was also introduced. This regulator works well even in case of large model errors as some simulations showed.

Aknowledgement

This work was supported in part by the MTA-BME Control Engineering Research Group of the HAS, at the Budapest University of Technology and Economics and by the project TAMOP 4.2.2.A-11/1/KONV-2012-2012, at the Széchenyi István University of Győr.

References

[1] Åström K.J., B. Wittenmark: *Computer Controlled Systems*, Prentice-Hall, 1984
[2] Åström K.J.: *Control System Design Lecture Notes*, University of California, Santa Barbara, 2002
[3] Kailath T.: *Linear Systems*, Prentice Hall, 1980
[4] Keviczky L.: *Combined identification and control: another way,* Control Engineering Practice, vol. 4, no. 5, pp. 685-698, 1996
DOI: 10.1016/0967-0661(96)00052-4
[5] Keviczky L., Cs. Bányász: *Iterative identification and control design using K-B parametrization*, In: Control of Complex Systems, Eds: K.J. Åström, P. Albertos, M. Blanke, A. Isidori, W. Schaufelberger, R. Sanz, Springer, pp. 101-121, 2001
[6] Keviczky L., Cs. Bányász: *Attenuation of the Model Error in Observer-Based State-Feedback Regulators*, Acta Technica Jaurinensis, vol. 7, vo. 1, pp. 46-61, 2014
DOI: 10.14513/actatechjaur.v7.n1.256
[7] Maciejowski J.M.: *Multivariable Feedback Design*, Addison Wesley, 1989

Acta
Technica
Jaurinensis

Vol. 7., No.2., pp. 123-129, 2014
DOI: 10.14513/actatechjaur.v7.n2.075
Available online at acta.sze.hu

Pricing systems analysis of DRT systems

T. Andrejszki, Á. Török

BME Department of Transport Technology and Economics, 1111 Budapest, Stoczek utca 2. building St
phone: +361 4631054
e-mail: tamas.andrejszki@mail.bme.hu, atorok@kgazd.bme.hu

Abstract: Flexible Transport Systems are becoming more and more popular in the sector of public transportation. The reason for this is mainly the economic sustainability of these systems which is reached by minimising the empty routes and using the capacities in a highly effective way. FTS systems use modern minibuses which have better conditions in emission and environmental pollution than the old vehicles of operating Hungarian public transport services. The aim of this study is to examine the opportunities of pricing regimes used in FTS systems. Firstly the conventional pricing methods such as average pricing will be reviewed than the new systems such as and marginal cost based pricing. Later on their mixing is going to be analysed. In the study the structure of FTSs were similar to "dial-a-ride" systems which aimed to handle the highest level of flexibility.

Keywords: Flexible Transport Systems, pricing regimes, cost function

1. Introduction

In the competition for technological, economical and social efficiency the professionals and researchers of each scientific field are working on the escalation of new results, in order to improve welfare in the society. The transport economist – having an interdisciplinary research field – are no exception of this: they have the simultaneous task of providing economical and social welfare, of providing efficiency and equity and of providing technological optimum for transport processes A characteristic topic of this theme is to configure the cost calculation system and pricing method of the transportation companies: in this case the researcher cannot avoid the basic technical processes that are described by financial ones, and have special attributes, constraints and barriers. At the same time, attention has to be paid to the instructions of the economic theory, in order to help the company to contribute to the increase of social welfare. One of the currently consorting optimisation theories in economics is the neoclassical doctrine with its marginal costs and marginal revenues. [1] The neo-classical school was founded by Walras, and by his mathematical model that describes the usage of marginal costs and marginal revenues [2]. First application of this was evolved at the beginning of the 20th century with the welfare theories of Marshal [3] and Pigou [4]. As a result of their work, a new, prospective challenge was revealed before theorists and practitioners of the

economics: the mathematical equations that were resulted from the research don't only provide the profit-maximum of the company, but, at the same time they show the way to the optimal capital-allocation and efficiency among companies for the society. Some paper even aim to compare some conservative public transport policy options [5]. The aim of this article is to investigate different FTS related pricing regimes.

2. Static average cost based method

The average cost (AC) based method proves to be the simplest pricing method. This pricing uses the calculated flat rate based on previously carried out traffic surveys [6]. This price can ensure the return (or a specified part of the return) on investment if the number of passengers measure up to the surveys. The sum of the fee depends greatly on the market situation of the transportation companies. Therefore, the pricing of the competitive market and the state pricing used should be separated due to their monopoly position. State pricing is usually justified when due to certain reasons the supply-demand market mechanism and the economic competition cannot prevail. The highest price should be set as to cover for necessary expenditures of efficient operation and profit [7]. As to the lowest price, it should be fixed as to cover the expenditures of efficient operation. In both cases several funding cuts and subsidies have to be taken into consideration.

At a static average cost based pricing system, the travelling fare (€/passenger) is defined by the proportion of the operator's all fixed and variable costs to the estimated travel needs. In this case one journey's fare is influenced by the following:

- Transportation needs (e.g. number of passengers);
- The available transportation capacity;
- Profit margin;
- Return time.

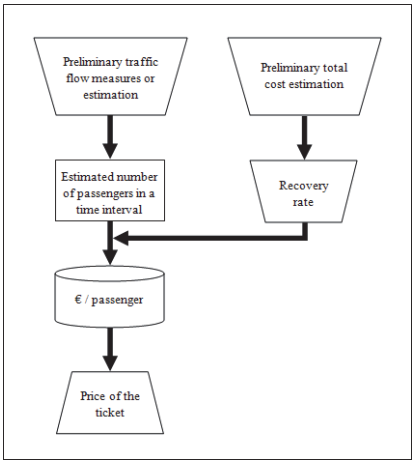


Figure 1. The process of static average cost based pricing

3. Dynamic average cost based method

Among the average cost based systems, certain pricing systems can be listed which determine the average specific cost based on one factor of the variable cost [8]. For such purpose, regarding passenger transportation, the passenger-kilometre can be applied for instance. In the case of such pricing systems the cost depends on the extent the passengers use the service.

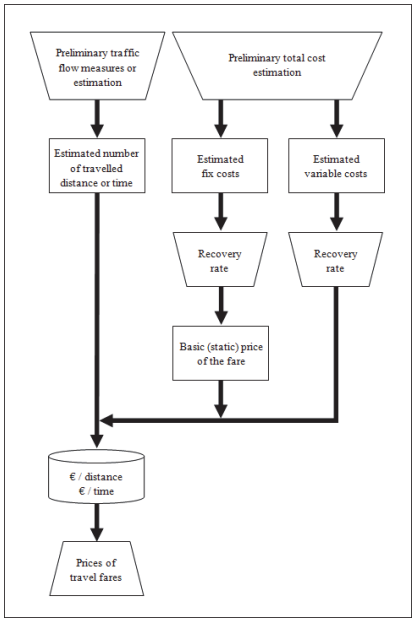


Figure 2. The process of dynamic average cost based pricing

So if the travel fare is previously not determined, but a specific price is used based on some measurable cost factors of an average route, in that case the pricing system can be defined as an average variable cost based system. In such a system the specific travel fare (e.g. €/km) is determined by the proportion of the total fixed and variable costs to the total estimated performance. In this case, the price of the travel fare is altered, besides the already listed factors of the average cost based system, by the following factors:

- Number of travelled stops;
- Travel time;
- Travel distance.

Regarding the average variable cost based system it is worth examining a certain case, in which a given passenger can be only transported through a longer route than the shortest between the departure and the arrival point [9]. Is it reasonable in such a case that this passenger pays more than the optimal price? If the passenger has to pay more than it is clear that the others have caused some sort of 'damage' to him or her. This can be adjusted e.g. if we determine the shortest route between the departure and the arrival point and calculate the difference between the shortest and the travelled actual route and

average these numbers. In proportion to the average deviation we adjust the fees to make certain that the "caused damages" are eventually even. However by using the method above, it will not be in the interest of the passengers for others to use this service.

Another question arises concerning the equity of the average variable cost based system when a passenger books a longer ride close from the centre, while a another one books a short one far from the centre. In such a case, based on the travelled kilometres the first passenger pays several times more than the second one, although the vacant bus route is due to the second passenger's admission to the system.

4. Marginal cost based method

Compared to the average cost based pricing systems, marginal cost pricing is based on a certain principle which states that all new users should pay for the additive expenses caused by their person. The prerequisite of marginal cost pricing is to define relevant factors of marginal costs, as well as to calculate their significance. Along with this, the categorization of the different types of costs can only be evaluated if we identify the interested parties who create these costs. The next step is the interpretation of the mechanism between the different types of marginal costs and their causes, which in the long-term demands the identification of the relevant behaviour dimensions [10].

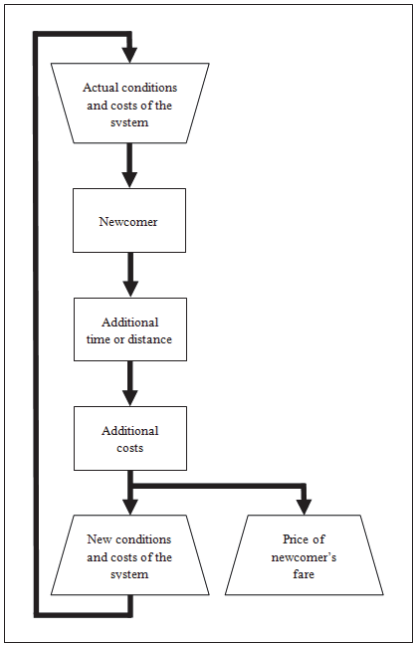


Figure 3. The process of marginal cost based pricing

The marginal cost term used in economics can be seen, in the field of flexible transportation as an extra cost by means of the route alternations due to a new registration. Thus, when the marginal cost based pricing is used all passengers pay the sum with which their admission to the system increased the costs of the operator.

$$MC = \frac{dTC}{dQ} \tag{1}$$

where:

- MC - marginal cost [€/passenger]
- TC - total cost [€]
- Q - quantity [passenger]

In case of a public transport service the quantity can be for example the number of passengers (and *dQ* is a new passenger incrementally admitted to the system – so after a new booking *dQ*=1). By this, if there is a service round consisting of n journeys which are inserted with one new booking, then the journey's marginal cost is the deviation between the service round with n + 1 journeys and the costs of the service round with n journeys.

In this system, users are interested in late bookings, since the first passenger is charged as if he was the only passenger to be transported, which seems to be rather expensive. To the contrary, if someone makes a booking just before the trip and the new passenger's route exactly corresponds with the set route then his cost can be zero. This characteristic of marginal cost based pricing does not coincide with the operator's interest.

Beside this, another case needs to be examined: when a new booking causes more extra operational cost (marginal cost) for the existing journey than the cost of separating him (so make 2 journeys with 2 vehicles). Though in the case of overcapacity, it can be solved by adding another vehicle to the system, however if surplus capacity characterizes the network then the above mentioned problem may turn out to be a determining point of conflict. This consideration would lead to the reconsideration of conservative vehicle and crew scheduling and reorganisation of operative scheduling as well [11], [12].

The contradictions of the pricing systems can be resolved by developing a mixed pricing system.

5. Mixed pricing

The new pricing system can be basically traced back to the average variable cost based pricing system, however in this case the base of the average cost is given by the loop cost. Individual loop cost refers to the cost of the shortest loop that starts from the headquarter, goes to the passenger's origin, then his final destination, and finally back to the headquarter.

So according to the new pricing method the costs of a complete route are divided in the proportion of the individual loop costs. Problems mentioned above -except one- are all resolved. Costs of vacant routes are distributed since the individual loops include the vacant routes. Passengers are motivated to suggest the service to other passengers as their expenses might decrease with the newcomers.

However if the marginal cost of the complete route (collecting loop) due to the new user is bigger than the given passenger's individual loop cost, than the other passengers'

expenses can even increase. To solve this problem, further correction needs to be applied. It is clear that every passenger causes some amount of externality to the others, since with more passengers the travel time and the level of comfort differs from travelling alone.

So the reason why the new system is called "Mixed" is the following. At every newcomer the system counts out these prices (based on individual loops) and compares them with the prices before the newcomer's appearance. If any price of it increases then the newcomer's price will be his marginal cost and for the others it will be the same as before. If there are no increases the new prices will be the individual loop-based prices.

It can be recognised that the mixed pricing method motivates passengers to order the service as early as possible because if they are in the first few volunteers the probability of being a passenger who should pay marginal cost is less. They can always count with that every newcomer might decrease their price. And at this point the system reached another important advantage: Passengers are motivated to advertise the DRT system and invite other peoples to use this service because it's economically better for them.

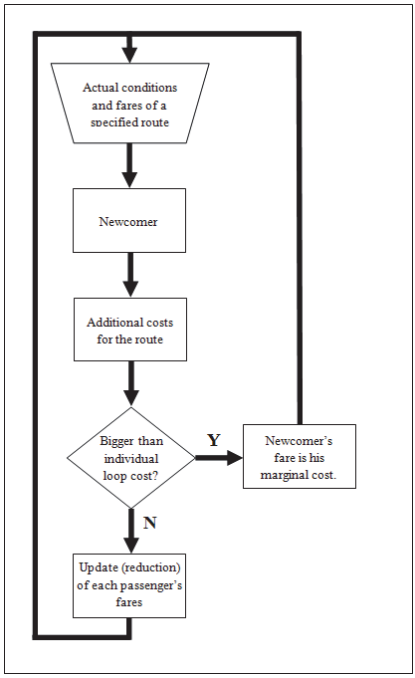


Figure 4. The process of mixed pricing

6. Conclusion

Different types of pricing regimes have been investigated in this article. Each of one has advantages and disadvantages.

On one hand average cots based pricing has simple formulas, it is clear and it can be calculated in a short time but it does not reflect the reality. On the other hand marginal

cost based pricing has a good reflection of the reality however the calculations take more energy and time to be done.

The mixed pricing system tries to compare the advantages of each one. Firstly in lot of cases mixed pricing is simpler than marginal cost based pricing. Secondly the mixed system is closer to the reality than the average cost based one. According to this it can be said that mixed pricing creates bridge or a golden mean between the two poles.

In theoretical way mixed pricing is worth to be used at transport companies but yet we have not had practical examples. Other studies are needed to examine the opportunities of putting this method into practise and then to evaluate the experiences for further development of this pricing system.

References

[1] Ronai P.: *Marginal cost based pricing model in the railway transport*, Ph.D. Dissertation, Budapest University of Technology and Economics, 2003

[2] Walras, L.: *Elements of the Theoretical Economics*; Lausanne, Switzerland, 1887

[3] Marshall, A.: *Principles of Economics*; Macmillan London, 1916

[4] Pigou, A.: *The Economics of Welfare*; Macmillan London, 1924

[5] Kumar M, Sarkar P, Madhu E.: *Development of fuzzy logic based mode choice model considering various public transport policy options*, International Journal for Traffic and Transport Engineering, Vol. 3, No. 4, pp408-425, 2013 DOI: 10.7708/ijtte.2013.3(4).05

[6] Horváth B., Horváth R., Gaál B.: *Estimation of Passenger Demand in Urban Public Transport*, Acta Technica Jaurinensis Vol. 6. No. 3., pp. 64-73., 2013

[7] Nelson, J. D., Phonphitakchai, T: *An evaluation of the user characteristics of an open access DRT service*; Research in Transportation Economics, 2012

[8] Palmer, K., Dessouky, M., Zhou, Z.: *Factors influencing productivity and operating cost of demand responsive transit*; Elsevier Transportation Research Part A, 2008

[9] Horváth B.: *A Simple Method to Forecast Travel Demand in Urban Public Transport*, Acta Polytechnica Hungarica Vol. 9, No. 4., pp. 165-176., 2012

[10] Milne, D., Niskanen, E. and Verhoef, E: *Operationalisation of marginal cost pricing within urban transport*; VATT-Research Reports, 2000

[11] Horváth B., Prileszky I.: *New ways in vehicle and crew scheduling*, Acta Technica Jaurinensis Vol. 4., No. 2., pp. 297-304., 2011

[12] Nagy E., Csiszár Cs.: *Research on Automation of Operative Scheduling in Urban Public Transportation*, Acta Technica Jaurinensis Vol. 6. No. 3., pp. 94-109., 2013

Acta
Technica
Jaurinensis

Vol.7., No.2., pp. 130–145, 2014
DOI: 10.14513/actatechjaur.v7.n2.260
Available online at acta.sze.hu

Application of Finite Element Method in High Frequency Simulations

G. Friedl

Széchenyi István University
9026, Győr, Hungary
E-mail: friedl@maxwell.sze.hu

Abstract: This paper is written in the frame of the usage of finite element method at high frequencies. Simulations in time and frequency domain have their own methodology, for example the setting of the excitations and boundary conditions. The paper presents the weak formulations of both domains derived from Maxwell's equations and from the Sommerfeld radiation condition, simulations in isotropic and linear media, and if it is possible, comparisons between the simulations and analytical results. The first part of the paper is about the theory of the electromagnetics, the form of the wave equations in time and frequency domain, and about the nodal finite element method. The second chapter is about deriving the weak forms, and about their implementation method. The final chapter shows the implementation of the finite element method, and some results simulated in time and frequency domain with different geometry and boundary conditions.

Keywords: *nodal finite element method, radio frequency simulation, time domain, frequency domain, Newmark method*

1. Introduction

1.1. Electromagnetics at high frequencies

The basics of the theory of electromagnetism are defined by Maxwell's equations [1], that describe the behavior of the electromagnetic field in every moment and at every point in space for all frequencies. The equations have a form in time and frequency domain as well. The differential forms of Maxwell's equations in time domain for linear media [2–7] are the following:

$$\nabla \times \boldsymbol{H}(\boldsymbol{r},t) = \boldsymbol{J}(\boldsymbol{r},t) + \sigma \boldsymbol{E}(\boldsymbol{r},t) + \varepsilon \frac{\partial \boldsymbol{E}(\boldsymbol{r},t)}{\partial t} \tag{1}$$

$$\nabla \times \boldsymbol{E}(\boldsymbol{r},t) = -\mu \frac{\partial \boldsymbol{H}(\boldsymbol{r},t)}{\partial t}, \tag{2}$$

$$\nabla \cdot \boldsymbol{B}(\boldsymbol{r},t) = 0, \tag{3}$$

$$\nabla \cdot \boldsymbol{D}(\boldsymbol{r},t) = \rho. \tag{4}$$

Here, the field quantities are depending on space $\boldsymbol{r}$ and time t $\boldsymbol{H} = \boldsymbol{H}(\boldsymbol{r},t)\left[\frac{\mathrm{A}}{\mathrm{m}}\right]$ indicates the magnetic field intensity, $\boldsymbol{J} = \boldsymbol{J}(\boldsymbol{r},t)\left[\frac{\mathrm{A}}{\mathrm{m}^2}\right]$ is the current density, $\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{r},t)\left[\frac{\mathrm{C}}{\mathrm{m}^2}\right]$ is the electric displacement vector, $\boldsymbol{E} = \boldsymbol{E}(\boldsymbol{r},t)\left[\frac{\mathrm{V}}{\mathrm{m}}\right]$ is the electric field intensity, $\boldsymbol{B} = \boldsymbol{B}(\boldsymbol{r},t)\,[\mathrm{T}]$ indicates the magnetic flux density and $\rho\left[\frac{\mathrm{C}}{\mathrm{m}^3}\right]$ is the charge density, the permittivity of vacuum is $\varepsilon_0 = 8.854 \cdot 10^{-12}\left[\frac{\mathrm{F}}{\mathrm{m}}\right]$, and the permeability of vacuum is $\mu_0 = 4\pi \cdot 10^{-7}\left[\frac{\mathrm{H}}{\mathrm{m}}\right]$. The constitution relations in linear media have the form:

$$\boldsymbol{D} = \varepsilon_0 \varepsilon_r \boldsymbol{E}, \tag{5}$$

$$\boldsymbol{B} = \mu_0 \mu_r \boldsymbol{H}, \tag{6}$$

$$\boldsymbol{J} = \sigma(\boldsymbol{E} + \boldsymbol{E}_i). \tag{7}$$

Here, $\boldsymbol{E}_i$ denotes the so called impressed electric field. The first Maxwell's equation in (1) and the second one in (2) in the frequency domain under the condition $\sigma \ll \varepsilon\omega$ can be written [2–6] as:

$$\nabla \times \boldsymbol{H}(\boldsymbol{r},\omega) = \boldsymbol{J}(\boldsymbol{r},\omega) + \sigma \boldsymbol{E}(\boldsymbol{r},\omega) + \mathrm{j}\omega\varepsilon_0\varepsilon_r \boldsymbol{E}(\boldsymbol{r},\omega), \tag{8}$$

$$\nabla \times \boldsymbol{E}(\boldsymbol{r},\omega) = -\mathrm{j}\omega\mu_0\mu_r \boldsymbol{H}(\boldsymbol{r},\omega), \tag{9}$$

where ω is the angular frequency and j is the imaginary unit. If there are no surface currents, on the boundary of the two media, the tangential component of the electric and the magnetic field intensity are continuous:

$$\boldsymbol{n} \times (\boldsymbol{E}_2 - \boldsymbol{E}_1) = 0, \tag{10}$$

$$\boldsymbol{n} \times (\boldsymbol{H}_2 - \boldsymbol{H}_1) = 0. \tag{11}$$

Above, $\boldsymbol{n}$ denotes the normal vector of the surface. The electric field is perpendicular to any perfectly conducting plane. At infinity, the electric field and the magnetic field must vanish, the electric field and magnetic field satisfies the Sommerfeld radiation condition [8]:

$$\lim_{r\to\infty} r\left[\nabla \times \boldsymbol{E} + \mathrm{j}k_0 \boldsymbol{n} \times \boldsymbol{E}\right] = \vec{0}, \tag{12}$$

where r is the distance from the source of the electromagnetic field, and k_0 is the wave number in free space. The wave number is defined as:

$$k_0 = \omega\sqrt{\mu_0\varepsilon_0}. \tag{13}$$

1.1.1. Wave equation in time domain

The wave equation for the electric field can be derived from (1) and (2) in time domain [5, 8] by eliminating the magnetic field intensity as:

$$\varepsilon \frac{\partial^2 \boldsymbol{E}}{\partial t^2} + \sigma \frac{\partial \boldsymbol{E}}{\partial t} + \nabla \times \left(\frac{1}{\mu} \nabla \times \boldsymbol{E} \right) = -\frac{\partial \boldsymbol{J}}{\partial t}. \tag{14}$$

Simulations are performed in a finite space, the boundaries are not at the infinity, i.e. the Sommerfeld radiation condition is not usable in the form of (12). On the surface of a finite space, the radiation condition in time domain [8] can be approximated as:

$$\boldsymbol{n} \times \left(\frac{1}{\mu_0} \nabla \times \boldsymbol{E} \right) + Y_0 \boldsymbol{n} \times \left(\boldsymbol{n} \times \frac{\partial \boldsymbol{E}}{\partial t} \right) \approx 0, \tag{15}$$

where Y_0 is the wave admittance of free space, and

$$Y_0 = \sqrt{\frac{\varepsilon_0}{\mu_0}} = \frac{1}{120\pi}[S]. \tag{16}$$

1.1.2. Wave equation in frequency domain

The wave equation for the electric field can be derived from (8) and (9) in frequency domain [5, 8] by eliminating the magnetic field intensity as:

$$\nabla \times \left(\frac{1}{\mu_r} \nabla \times \boldsymbol{E} \right) - k_0^2 \varepsilon_r \boldsymbol{E} = -\mathrm{j} k_0 Z_0 \boldsymbol{J}, \tag{17}$$

where Z_0 is the wave impedance of the free space,

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 120\pi[\Omega]. \tag{18}$$

On the surface of a finite space, the Sommerfeld radiation condition in frequency domain [8] can be approximated as:

$$\boldsymbol{n} \times \nabla \times \boldsymbol{E} + \mathrm{j} k_0 \boldsymbol{n} \times (\boldsymbol{n} \times \boldsymbol{E}) \approx 0. \tag{19}$$

1.2. Basics of nodal finite element method

Finite element method [6,8–13] is a numerical technique developed for finding approximate solutions of differential equations. It is used to approximate analytically unsolvable or hardly solvable problems, such as electromagnetic, thermal, fluid, etc. problems. The method has four main steps [6]:

- *Specification of the model*: defining the geometry, and the physics of the problem;
- *Preprocessing*: discretization of the geometry, finite element mesh;
- *Calculation*: finite element assembly, solving the equation system;
- *Postprocessing*: evaluation of the results.

In this paper, the *preprocessing* step was made in the frame of `Gmsh` [14, 15], the *calculation* and the *postprocessing* steps were realized by `Matlab` [16] functions.

To solve the approximation of the problem with the weighted residual method, the equations are multiplied by arbitrary weighting functions, but in *Galerkin*'s method [17], the weighting functions are selected as the approximate shape functions. In this paper the nodal shape functions are used, because every example are defined with a current density perpendicular to the plane of the geometry. There are defined shape functions on each node of each element. A shape function interpolates the solution between the mesh nodes. A shape function on a node has the value of 1, and 0 on all other nodes. In 1-dimension, the elements are straight sections. The 1-dimensional shape functions can be defined [6] as:

$$N_i(x) = \prod_{j=1, j\neq i}^{m} \frac{x - x_j}{x_i - x_j}, \tag{20}$$

where N is the shape function, m is the number of the nodes of the element, x_i and x_j are the coordinates of the nodes of the element. The first and the third order shape functions can be seen on Fig. 1.

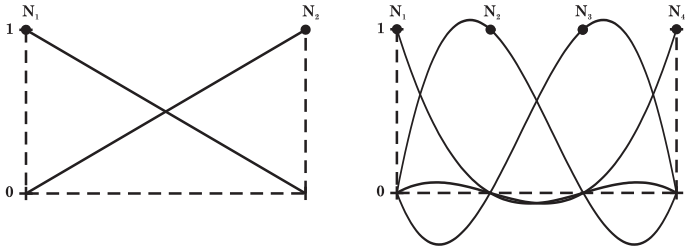


Figure 1. 1-dimensional shape functions

The 2-dimensional linear shape functions on triangle elements can be introduced [6] by using barycentric coordinates. The area of triangles can be calculated by determinants as:

$$\Delta = \frac{1}{2} \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}. \tag{21}$$

An arbitrary point in the triangle with coordinates (x,y) splits it onto three different triangles (see in Figure 2). The area of the triangles can be calculated, as above, as:

$$\Delta_1 = \frac{1}{2}\begin{vmatrix} 1 & x & y \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}, \Delta_2 = \frac{1}{2}\begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x & y \\ 1 & x_3 & y_3 \end{vmatrix}, \Delta_3 = \frac{1}{2}\begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x & y \end{vmatrix}. \quad (22)$$

Now, the 2-dimensional linear shape functions can be defined as:

$$N_i = \frac{\Delta_i}{\Delta}. \quad (23)$$

The nodal finite element method can be used only, if the unknown vector space is perpendicular to the 2-dimensional geometry, or the unknown field is scalar.

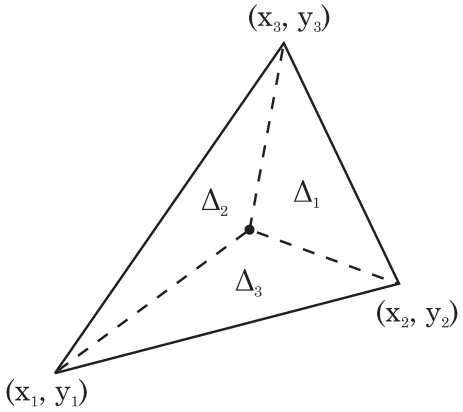


Figure 2. To understand the barycentric coordinates

2. Weak formulations, and their implementation method

2.1. Equations in time domain

To find the approximation of the solution of the time domain problem, the wave equation (14) can be multiplied with an arbitrary vector weighting function $\boldsymbol{T}$ and integrated over the whole volume V:

$$\int_V \boldsymbol{T} \cdot \nabla \times \left(\frac{1}{\mu}\nabla \times \boldsymbol{E}\right) \,\mathrm{d}V + \int_V \boldsymbol{T} \cdot \varepsilon \frac{\partial^2 \boldsymbol{E}}{\partial t^2} \,\mathrm{d}V + \int_V \boldsymbol{T} \cdot \sigma \frac{\partial \boldsymbol{E}}{\partial t} \,\mathrm{d}V = - \int \boldsymbol{T} \cdot \frac{\partial \boldsymbol{J}}{\partial t} \,\mathrm{d}V. \quad (24)$$

This equation is the so called direct formulation, that contains second order derivatives. It can be eliminated by invoking the vector algebraic identity [9]:

$$\nabla \cdot (\boldsymbol{x} \times \boldsymbol{y}) = \boldsymbol{y} \cdot \nabla \times \boldsymbol{x} - \boldsymbol{x} \cdot \nabla \times \boldsymbol{y}, \tag{25}$$

with $\boldsymbol{x} = \boldsymbol{T}$ and $\boldsymbol{y} = \mu^{-1}\nabla \times \boldsymbol{E}$, and the approximation of the Sommerfeld radiation condition (15), the weak form of the time domain analysis [8] can be represented as:

$$\int_V (\nabla \times \boldsymbol{T}) \cdot \frac{1}{\mu}(\nabla \times \boldsymbol{E})\,\mathrm{d}V + \int_V \boldsymbol{T}\varepsilon\frac{\partial^2 \boldsymbol{E}}{\partial t^2}\,\mathrm{d}V + \int_V \boldsymbol{T}\sigma\frac{\partial \boldsymbol{E}}{\partial t}\,\mathrm{d}V$$
$$+Y_0 \oint_A (\boldsymbol{n} \times \boldsymbol{T}) \cdot \left(\boldsymbol{n} \times \frac{\partial \boldsymbol{E}}{\partial t}\right)\,\mathrm{d}A = -\int \boldsymbol{T} \cdot \frac{\partial \boldsymbol{J}}{\partial t}\,\mathrm{d}V, \tag{26}$$

where S is the integration variable on surface. The finite element method use the weak formulation with *Galerkin*'s-method when the weighting function and the shape function are the same. The solution of the electric field within each element using the shape functions can be expressed [6, 8] as:

$$E(r) = \sum_{i=1}^{m} N_i(r)E_i. \tag{27}$$

Rewriting (27) into (26) results in a second order, inhomogeneous differential equation system [8]:

$$\mathbf{Q}\frac{\partial^2 \mathbf{e}}{\partial t^2} + \mathbf{R}\frac{\partial \mathbf{e}}{\partial t} + \mathbf{S}\,\mathbf{e} = \mathbf{f}, \tag{28}$$

where:

$$Q_{i,j} = \int_V \boldsymbol{T}_i \cdot \varepsilon \boldsymbol{T}_j\,\mathrm{d}V, \tag{29}$$

$$R_{i,j} = \int_V \boldsymbol{T}_i \cdot \sigma \boldsymbol{T}_j\,\mathrm{d}V + Y_0 \oint_A (\boldsymbol{n} \times \boldsymbol{T}_i) \cdot (\boldsymbol{n} \times \boldsymbol{T}_j)\,\mathrm{d}A, \tag{30}$$

$$S_{i,j} = \int_V (\nabla \times \boldsymbol{T}_i)\frac{1}{\mu}(\nabla \times \boldsymbol{T}_j)\,\mathrm{d}V, \tag{31}$$

$$f_i = -\int_V \boldsymbol{T}_i \cdot \frac{\partial \boldsymbol{J}}{\partial t}\,\mathrm{d}V, \tag{32}$$

and **e** is the columnvector of the unknown electric field. An unconditionally stable method to solve numerically a second order differential equation is the so called *Newmark-*β method [19]. Using the central difference form for the first and the second order time derivatives:

$$\frac{\partial u(t)}{\partial t} \approx \frac{u(t)^{n+1} - u(t)^{n-1}}{2\Delta t}, \tag{33}$$

$$\frac{\partial^2 u(t)}{\partial t^2} \approx \frac{u(t)^{n+1} - 2u(t)^n + u(t)^{n-1}}{\Delta t^2}, \tag{34}$$

and the weighted average for undifferentiated quantities:

$$\mathbf{e} \approx \beta \mathbf{e}^{n+1} + (1-2\beta)\mathbf{e}^{n} + \beta \mathbf{e}^{n-1}, \tag{35}$$

$$\mathbf{f} \approx \beta \mathbf{f}^{n+1} + (1-2\beta)\mathbf{f}^{n} + \beta \mathbf{f}^{n-1}, \tag{36}$$

where β is a parameter between 0 and 1, substituting these into (28), the form of the time-stepping equation system is:

$$\left(\mathbf{Q}\frac{1}{\Delta t^2} + \mathbf{R}\frac{1}{2\Delta t} + \beta\mathbf{S}\right)\mathbf{e}^{n+1} = \left(\frac{2}{\Delta t^2}\mathbf{Q} - (1-2\beta)\mathbf{S}\right)\mathbf{e}^{n}$$
$$-\left(\mathbf{Q}\frac{1}{\Delta t^2} - \mathbf{R}\frac{1}{2\Delta t} + \beta\mathbf{S}\right)\mathbf{e}^{n-1} + \beta\mathbf{f}^{n+1} + (1-2\beta)\mathbf{f}^{n} + \beta\mathbf{f}^{n-1}. \tag{37}$$

It is easy to see, that (37) is now a solvable equation system, where $\mathbf{e}^{n+1}$ is the unknown vector, and the right hand side of the equation is known from the time steps earlier. This method is unconditionally stable, if $\beta \geq \frac{1}{4}$ [8].

2.2. Equations in frequency domain

Multiplied the wave equation in frequency domain (17) with an arbitrary vector weighting function $\boldsymbol{T}$ and integrated over the whole volume V:

$$\int_V \boldsymbol{T} \cdot \left(\nabla \times \frac{1}{\mu_r} \times \boldsymbol{E}\right) \mathrm{d}V - \int_V \boldsymbol{T} \cdot k_0^2 \epsilon_r \boldsymbol{E}\, \mathrm{d}V = -\int_V \boldsymbol{T} \cdot \mathrm{j}\omega\mu_0 \boldsymbol{J}\, \mathrm{d}V. \tag{38}$$

This equation contains second order derivatives. It can be eliminated by using the vector algebraic identity in (25) with $\boldsymbol{x} = \boldsymbol{T}$ and $\boldsymbol{y} = \mu_r^{-1}\nabla \times \boldsymbol{E}$, then the Gauss-Ostrogradskij theorem [6]:

$$\oint_A \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{A} = \int_V \nabla \cdot \boldsymbol{v}\, \mathrm{d}V, \tag{39}$$

with $\boldsymbol{v} = \boldsymbol{T} \times \left(\mu_r^{-1}\nabla \times \boldsymbol{E}\right)$ and the approximation of the Sommerfeld radiation condition(19), the weak formulation of the frequency domain analysis is:

$$\int_V (\nabla \times \boldsymbol{T}) \cdot \frac{1}{\mu_r}(\nabla \times \boldsymbol{E})\, \mathrm{d}V - \int_V \boldsymbol{T} \cdot k_0^2 \epsilon_r \boldsymbol{E}\, \mathrm{d}V$$
$$+\mathrm{j}k_0 \oint_A (\boldsymbol{n} \times \boldsymbol{T}) \cdot (\nabla \times \boldsymbol{E})\, \mathrm{d}A = \int_V \boldsymbol{T} \cdot \mathrm{j}k_0 Z_0 \boldsymbol{J}\, \mathrm{d}V, \tag{40}$$

Substituting (27) into the weak form of (40), the result is a simple equation system in the form:

$$\mathbf{K}\mathbf{e} = \mathbf{b}, \tag{41}$$

where

$$K_{i,j} = \int_V \left[(\nabla \times \boldsymbol{T}_i) \cdot \frac{1}{\mu_r} (\nabla \times \boldsymbol{T}_j) - k_0^2 \boldsymbol{T}_i \cdot \epsilon_r \boldsymbol{T}_j \right] \mathrm{d}V + \mathrm{j}k_0 \oint_A (\boldsymbol{n} \times \boldsymbol{T}_i) \cdot (\boldsymbol{n} \times \boldsymbol{T}_j) \, \mathrm{d}A, \tag{42}$$

and

$$b_i = \int_V \boldsymbol{T}_i \cdot \mathrm{j}k_0 Z_0 \boldsymbol{J} \, \mathrm{d}V. \tag{43}$$

The simulations can be evaluated for only one frequency at one case. Along the boundaries, which are not in vacuum/air, the wave number of the surface integral in (42) should be the wave number calculated from the parameters of the boundary material.

3. Examples

The following simulations will show the advantages and disadvantages of the time and frequency domain simulations. If it is possible, the accuracy of the models will be shown compared to analytical results.

3.1. Simulations in time domain

3.1.1. Perfect electrically conducting wall

A perfect electrically conducting wall reflects the electromagnetic wave with a reflexion coefficient -1. The first simulated geometry is a 3 meter long line, an excitation is placed at 1 meter, and the boundary on the right side is a perfectly conducting wall. The conducting wall is implemented as a simple *Dirichlet* boundary, where the electric field is set to be zero. A *Dirichlet*-type boundary condition defines the exact value of the solution. The left boundary must be defined as an absorbing boundary to grant the wave propagation without reflexion on the wall. This is a Neumann type boundary condition, which means, during the assembly, to the value $R_{i,i}$ (See in (30)), where i is the number of the node, one has to add the value of Y_0. A *Neumann*-type boundary condition specifies the derivative of the solution on a boundary. Two simulations were made on one frequency and one simulation was made with a combined signal (See in Figure 3). The results in Fig. 3 are the maximal values on each element after the transient. If the applied frequency is 150 MHz, the distance between the source and the wall is the half of the wavelength, so the sum of the incident and the reflected wave is zero, but at 225 MHz, the distance is three-quarter of the wavelength, so the incident wave meets the reflected wave at the antenna in equal phase.

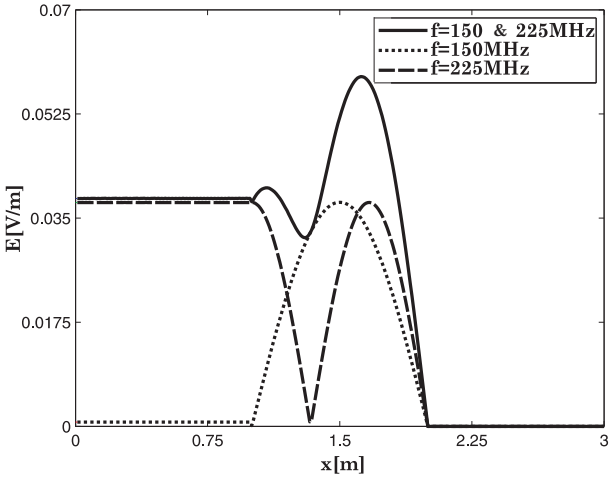


Figure 3. Results with perfect electrically conducting wall

3.1.2. Reflexion on the boundary of two different media

On the boundary of two media, if the wave impedances differ from each other, reflexion occurs, what can be calculated as:

$$\Gamma = \frac{Z_{02} - Z_{01}}{Z_{02} + Z_{01}}, \tag{44}$$

where Γ is the reflexion coefficient, Z_{02} and Z_{01} are the wave impedance of the two media. The geometry is a 5 meter long line, which is separated into two part (See in Figure 4). In time domain, the reflexion can be calculated from the incident and the reflected wave,

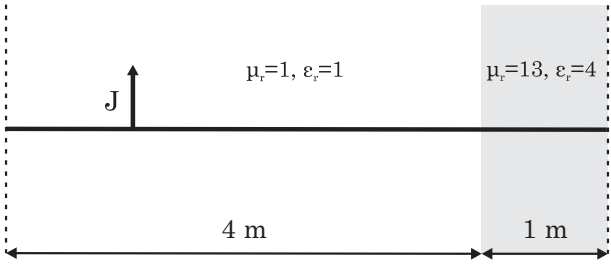


Figure 4. Applied geometry for the reflexion simulation

or from the standing wave ratio. The left and the right boundary are both absorbing boundaries, but on the right side it has different wave admittance. The result can be seen

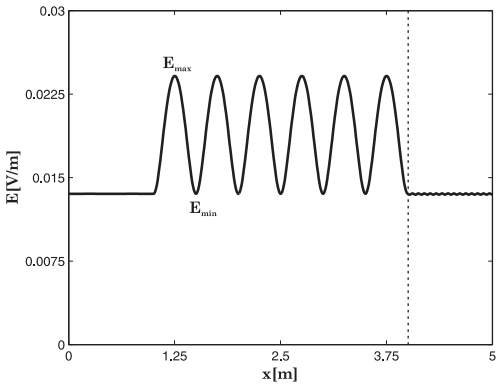


Figure 5. Results with two different media

on Fig. 5. The maximal and the minimal value of the electric field can be read from the results. The calculated reflexion coefficient is 0.286, and the simulated is 0.285.

3.1.3. Rectangular cavity rezonator

Analyzing resonators and their quality factors requires time domain simulations, because transients can be analyzed only in the time domain. The transient analysis is needed, because the oscillation is the transient response of the resonators. A resonator with perfectly conducting walls, used on the resonance frequency, has an infinite quality factor. The geometry of the applied problem is a simple square, witch size of the sides are 0.1 meter (See in Figure 6). The resonance frequency of this resonator of the mode TE_{101} [5] is 1.0606 GHz. The excitation is one period at the resonance frequency. Figure 7 shows the results of an ideal rectangle cavity and a rectangle cavity with badly conducting walls. The boundary of the rectangle cavity with the loss has the properties σ=10 $[S/m]$, and $\mu_r = \varepsilon_r = 2$. The excitation is perpendicular to the plane of the geometry. The wall of

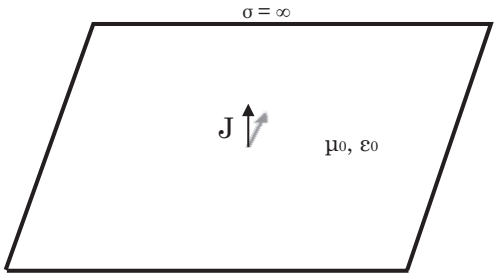


Figure 6. Model of the rectangular cavity

the resonator is set as a perfect electrically conducting surface, the electric field must be perpendicular to the perfectly conducting surface. In the 2-dimensional simulations, where the excitation is perpendicular to the simulation plane, the electric field is parallel to the source current, so the electric field is zero on the conducting wall in this case. In Fig. 7, the dotted line shows the case with perfeclty conducting walls, and the continuous line is for the electric field in the cavity with badly conducting walls. The quality factor of the

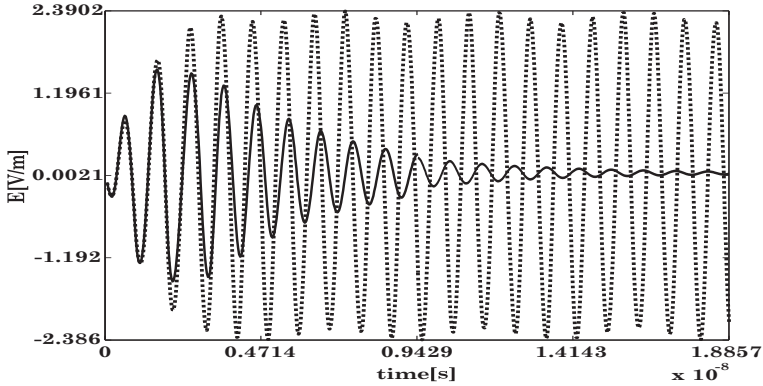


Figure 7. Electric field in the center of rectangular cavity resonator

rectangular cavity can be defined as the number of period till the amplitude of the electric field reaches the value of the original amplitude divided by the Euler number *e*. Here the value of quality factor is approximately about 6.

Time domain simulations are very time consuming, because every time step must be calculated. The transient analysis is only possible in time domain, but for the most of the high frequency problems, it is not needed to calculate the transient behavior.

3.2. Simulations in frequency domain

3.2.1. Perfect electrically conducting wall

The following problem, the geometry and the excitation is the same as in the section 3.1.1.The method of implementing of the conducting and absorbing wall is the same as in time domain. The result of a frequency domain simulation is a complex value. The absolute value and the phase of the result can be seen on Fig. 8 and Fig. 9. The figures above show the result without the conducting wall. The absolute values of the results are the same as in time domain. The phase is constant between the excitation and the wall, which means, that standing waves are formed there. The linear parts of the phase plot mean simple wave propagating.

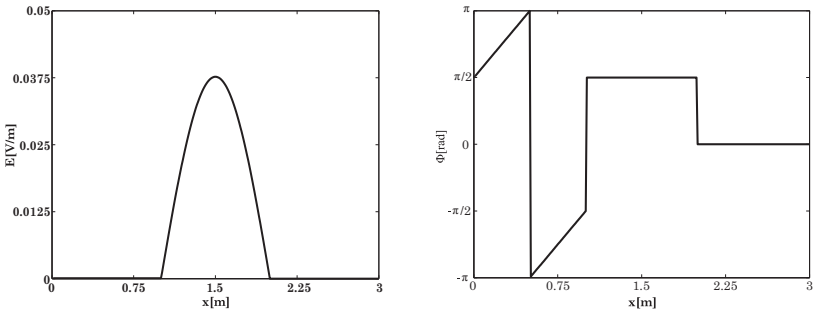


Figure 8. The absolute value and the phase of the result at 150 MHz

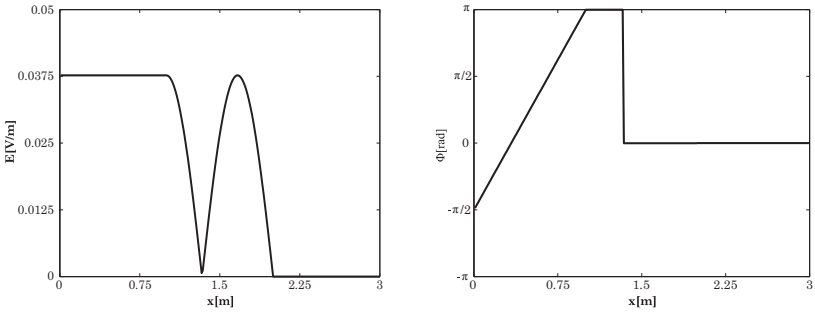


Figure 9. The absolute value and the phase of the result at 225 MHz

3.2.2. Rectangular waveguide

This simulation is a about a simple rectangular waveguide in TE_{10} mode. The model of the geometry can be seen in Fig. 10. The width of the waveguide is 0.1 m, so the

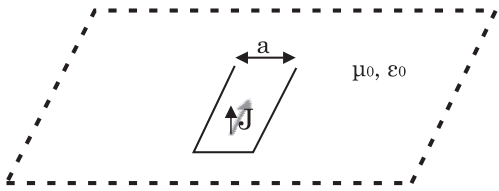


Figure 10. The model of the simulated rectangular waveguide

cutoff frequency of this arrangement is 1.5 GHz. The frequency of the excitation is chosen as 2 GHz, from the value of the cutoff frequency and the applied frequency the inguide wavelength van be calculated, witch is 0.2268 m. Fig. 11 shows the absolute value and the phase of the result of the simulation. On Fig. 11 it can be seen that the open ended

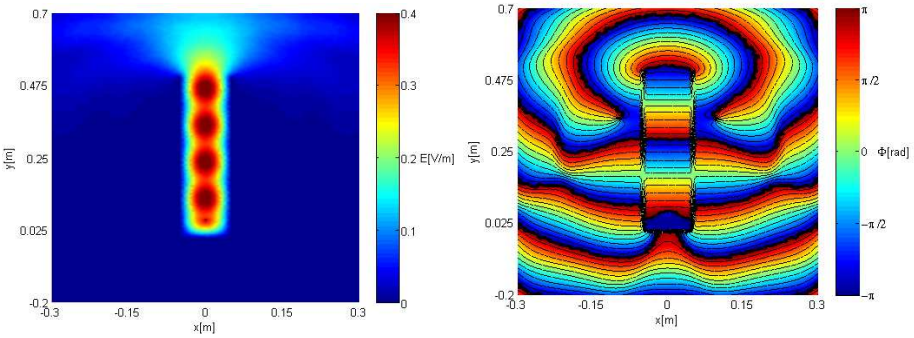


Figure 11. The absolute value and the phase of the result

rectangular waveguide radiates backward. The guide wavelength can be measured from the phase plot, in this case it is approximately 0.228 m. The error is in the order of magnitude of the size of the elements in the waveguide.

3.2.3. Parabolic antenna

Parabolic antennas are widely used by satellite communication systems owing to their polar pattern. Due to the special geometry of the parabolic antennas, plane waves are propagating parallel to their axis. The geometry and the mesh are shown on Fig. 12. The

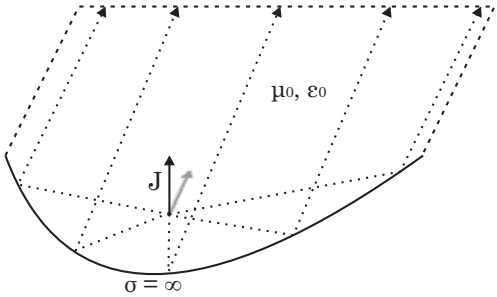


Figure 12. The model of the simulated parabolic antenna

frequency of the simulation was 300 MHz. In practice, the size of the mirror is much larger, than the wavelength of the signal, the simulation was made on lower frequency for the better visibility, and to decrease the number of unknowns. The mirror is set as a perfect electrically conducting surface. Fig. 13 shows the absolute value and the phase of the result. It can be seen on the plot of the phase, that the wave propagates parallel to the axis of the mirror as an almost perfect plane wave. The real parabolic antennas work on higher frequencies, and have a better polar pattern.

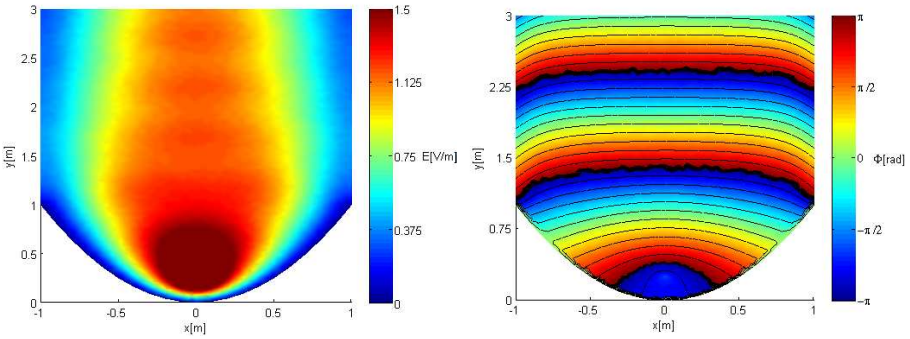


Figure 13. The absolute value and the phase of the result

The simulations in frequency domain give fast result for the problems, but only at one frequency at a time. In practice, the most important question is the behavior in, and out of the applied frequency range of any device.

4. Conclusions

In this paper, simple 1-D and 2-D electromagnetic field problems were presented in time and frequency domain. The first chapter is a short theoretical summary of the wave equations expressed to the electric field and of the nodal finite element method as well. In the second chapter the deriving of the applied weak forms in time and frequency domain and their application methods were discussed. These methods were presented via some simple examples. The simulations represent the physical behavior well. Time domain simulation is very effective to find the result with an appropriate input signal and to calculate the transients phenomena, however, it can be very time consuming, because the field must be calculated in every time step. In contrast, a frequency domain simulation can be done only at one frequency at a time, but the implementation of it is much easier, it gives a fast result, and in practice, the behavior of the devices in frequency domain is much more important. In the future, the edge elements will be implemented to simulate complex problems, where the excitation is in the plane of the simulation.

Acknowledgement

This research was supported by the **European Union and the State of Hungary, co-financed by the European Social Fund** in the framework of TÁMOP-4.2.4.A/ 2-11/1-2012-0001 'National Excellence Program'.

References

[1] Maxwell, J. C.: *A Dynamical Theory of the Electromagnetic Field*, Philosophical Transactions of the Royal Society, London, 1865

[2] Simonyi, K., Zombory, L.: *Theoretycal Electricity (in Hungarian)*, Műszaki Könyvkiadó, Budapest, 2000

[3] Fodor, Gy.: *Electromagnetic Fields (in Hungarian)*, Műegyetemi Kiadó, Budapest, 1996

[4] Standeisky, I.: *Electrodynamics (in Hungarian)*, Universitas, Győr, 2006

[5] Pozar, D. M.: *Microwave Engineering*, John Wiley and Sons, New York, 2012

[6] Kuczmann, M., Iványi, A.: *The Finite Element Method in Magnetics*, Academic Press, Budapest, 2008

[7] Jackson, J. D.: *Classical Electrodynamics*, John Wiley and Sons, New York, 1999

[8] Jin, J.: *Finite Element Analysis of Antennas and Arrays*, John Wiley and Sons, New York, 2009

[9] Jin, J.: *The Finite Element Method in Electromagnetics*, John Wiley and Sons, New York, 2002

[10] Gustrau, F., Manteuffel, D.: *EM Modeling of Antennas and RF Components for Wireless Communication Systems*, Springer, Berlin, 2006

[11] Pepper, D. W., Heinrich, J. C.: *The Finite Element Method*, Taylor and Francis, New York, 2006

[12] Kuczmann, M.: *Vector Preisach Hysteresis Modeling: Measurement, Identification and Application*, Physica B - Condensed Matter, vol. 406, no. 8, pp. 1403-1409, 2011 DOI: 10.1016/j.physb.2011.01.037

[13] Kuczmann, M., Budai, T., Kovács, G., Marcsa, D., Friedl, G., Prukner, P., Unger, T., Tomozi, Gy.: *Application of PETSC and other useful packages in finite element simulation*, Pollack Periodica, vol. 8, no. 2, pp. 141-148, 2013 DOI: 10.1556/Pollack.8.2013.2.15

[14] Gmsh, http://geuz.org/gmsh/, (last visited: 05 February 2014)

[15] Geuzaine, C., Remacle, J. F.: *Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities*, International Journal for Numerical Methods in Engineering, vol. 79, no. 11, pp. 1309-1331, 2009 DOI: 10.1002/nme.2579

[16] Matlab, www.mathworks.com, (last visited: 05 February 2014)

[17] Iványi, A.: *Continous and Discrete Simulations in Electrodynamics (in Hungarian)*, Akadémiai Kiadó, Budapest, 2003

[18] Bronshtein, I. N., Semendyayev, K. A., Musiol, G., Muehling, H.: *Handbook of Mathematics*. Typotex, Budapest, 2009

[19] Newmark, N. M. *A method of computation for structural dynamics*, Journal of Engineering Mechanics, ASCE, 1959

Acta
Technica
Jaurinensis

Vol. 7., No.2., pp. 146-155, 2014
DOI: 10.14513/actatechjaur.v7.n2.271
Available online at acta.sze.hu

Levels of Selected Metals in Ambient Air PM10 in an Urban Site of Győr, Hungary

A. Szabó Nagy, Zs. Csanádi, J. Szabó

Széchenyi István University, Physics and Chemistry Department
Egyetem square 1, 9026 Győr, Hungary
E-mail: nszaboa@sze.hu

Abstract: The aim of this study was to assess the ambient concentrations of some selected trace metals (Pb, Cd and Ni) and the metalloid As in the PM10 aerosol fraction in an urban site of Győr, Hungary in 2011. The results show that the study area has excellent air quality with respect to metal(loid)s during the four sampling periods. However, the annual average concentration of PM10 indicates a polluted area. The PM10 concentrations were higher in heating season than in summer time. The levels of PM10 and trace elements determined were compared with other cities located in Hungary and around the world.

Keywords: PM10, trace metals, ambient air, air quality

1. Introduction

Air quality is an important issue for public health, the economy and the environment. Poor air quality as a result of air pollution is a major environmental health risk, contributing to respiratory disease, cardiovascular disease, and lung cancer. In addition to the health effects, air pollution has considerable economic impacts, cutting short lives, increasing medical costs, and reducing productivity through lost working days across the economy. Particulate matter with an aerodynamic diameter smaller than 10 μm (PM10) is considered to be a reliable indicator for possible health effects due to fine particles in ambient air. It consists of a complex mixture of solid and liquid particles of organic and inorganic substances suspended in the air. Although metal-bearing aerosols constitute a small fraction of the PM10 mass, the exceeding concentration and/or long-term exposure to metals could cause severe toxic effects on human health. Thus, measurement of metal concentration levels in inhalable particles is important in determining their potential impacts on human health [1-3].

Metal-bearing aerosols in the ambient atmosphere are produced by various anthropogenic and natural sources. Combustion of fossil fuels and wood, exhaust emission from vehicles, industrial activities, energy production, construction and waste incineration are known to be anthropogenic sources, while volcanic activity, wind-eroded soil dusts, forest fires, and sea salt spray may contribute to natural metal-bearing aerosols [1, 3]. Vehicle traffic is a potential source of Cd, Pb, Mn and Ni as a result of

fuel combustion and the wearing of brakes, tires, and other components [4]. The Pb emission has decreased due to the introduction of unleaded fuels. However, increasing environmental concentrations of platinum group elements (Pt, Pd, Rh) from catalytic converters has been reported worldwide. Emissions from coal burning power plants are a potential source for As, Cd, Cr, Pb and Ni while local ceramic and metal-processing facilities could emit a wide variety of trace metals and metalloids [5].

The aim of this work was to give an overview about the levels of some selected trace metals (Pb, Cd and Ni) and the metalloid As in the PM10 aerosol fraction in an urban area of Győr (Hungary) with high traffic density during the year 2011. The levels of PM10 and trace elements determined in our study were compared with air quality standards and with some published data of other cities.

2. Experimental

2.1. Study area

Győr (47°41′02″N, 17°38′06″E) is the most important city in the northwest area of Hungary – halfway between Wien, Bratislava and Budapest – situated on one of the important roads of Central Europe. The city is the sixth largest in Hungary, and one of the seven main regional centres of the country. The number of inhabitants is about 128,500. Győr is a dynamically developing city due to its good geographic situation and as an emphasized centre in automotive industry. It has become one of the largest economic, industrial and traffic areas of Hungary. The monitoring site is located at the junction of Tihanyi Árpád street and Ifjúság boulevard (Fig. 1).

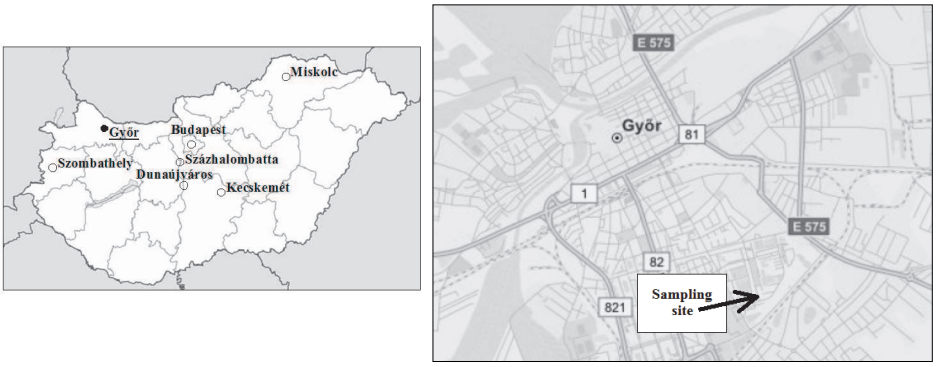


Figure 1. The location of Győr, Hungary and the sampling site in Győr. Some selected other Hungarian stations are also marked, about which this study includes some concentration data

2.2. Sampling and chemical analysis

The concentrations of PM10 aerosol fractions and PM10 bound trace metal(loid)s were measured and samples were collected in every third month at 14 day intervals, continuously for 24 hours in 2011 at the monitoring site of Győr. A Digitel High Volume DHA80 (Digitel Elektronik AG, Switzerland) sampler [6] was used for

collection ambient aerosol particles, which were chemically analysed later. This equipment is considered to be equivalent to the requirements of the European Standard (EN 12341) for sampling PM10 matter [7]. Samples are taken onto high purity Advantec QR-100 quartz fibre filters (size: 150 mm diameter) for a period of 24 h at a flow rate of 30 m^3/h into a container with about 700 m^3 volume capacity. The air flow rate can be established with a pump and controlled with an air flow meter.

Before and after sampling, the filters were conditioned during 48 h at 20±1 °C and 50±5% relative humidity. The particle total mass was determined by weighing of the sampling filters before and after sampling and the PM10 concentration calculated from the weighted mass on the filter and the sampling volume. This DHA-80 equipment has a container of 15 filters mounted in filter holders and they are changed automatically to the flow position. After the sampling, the filters were wrapped in aluminium foil separately and stored cooled until chemical analysis.

The concentrations of trace metal(loid)s (Pb, Cd, As and Ni) in the PM10 aerosol fraction were measured by graphite furnace atomic absorption spectroscopy (SOLAAR MQZ, Unicam Ltd., Cambridge, UK) equipped with Zeeman and deuterium background correctors, a graphite furnace GF9 and an autosampler. One half of the filter was cut by a ceramic scissor and the sample was treated with 15 mL aqua regia and digested at temperatures up to 210 ºC for 20 min using a CEM Mars 5 microwave. The resulting solution was filtered and diluted to 100 mL with distilled water. A 20-µL volume of the sample was injected into the graphite tube. The sample analysis was conducted in accordance with the MSZ EN 14902:2006 Hungarian standard method procedure [8].

3. Results and discussion

3.1. Particulate matter concentrations

Table 1 gives an overview of the mass concentrations of PM10 as well as the measured trace elements in PM10 at the urban site of Győr in 2011. Fig. 2 shows the average concentrations and standard deviations of PM10 during the different sampling periods. The PM10 concentrations ranged from 11 to 119.14 µg/m^3. The highest concentration was detected in the sampling of November. Concentrations of PM10 exhibit large variability during the sampling periods. The PM10 concentrations were higher in heating season than in summer time.

Table 1. Concentrations of PM10 and the measured trace metal(loid)s at the urban site of Győr during different sampling periods in 2011

		Trace elements ng/m^3			
Sampling period	***PM 10 µg/m^3***	***Pb***	***Cd***	***As***	***Ni***
I. 16 February – 1 March	46.59–110.68	11.11–40.75	0.2–0.73	0.25–1.21	1.03–19.53
II. 4–17 May	15.39–29.44	2.95–17	0.28–1.34	0.28–1.04	2.22–5.86
III. 1–14 August	11–29.1	1.34–7.08	0.05–0.53	0.26–6.64	3.13–8.64
IV. 1–15 November	42.6–119.14	7.53–24.81	0.19–0.57	0.24–1.32	1.23–5.91

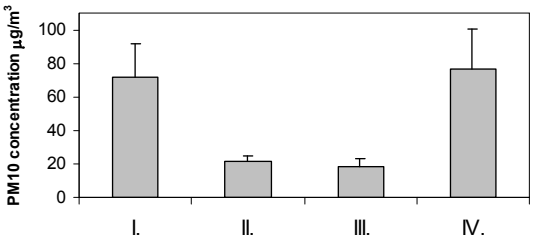


Figure 2. PM10 concentrations at the urban site of Győr during the four sampling periods in 2011

3.2. Trace metal and arsenic concentrations

The trend of trace metal(loid) levels at the urban site of Győr in 2011 was found in decreasing order of Pb > Ni > As > Cd. The concentration of Pb showed the maximum value of 40.75 ng/m^3 whereas Cd showed the lowest concentration of 0.05 ng/m^3 (Table 1). Overall, the measured trace elements typically amount to about 0.04 % of the total mass of the PM10 fraction in the Győr atmosphere. Fig. 3 illustrates the temporal trend of trace element concentrations during the four sampling periods.

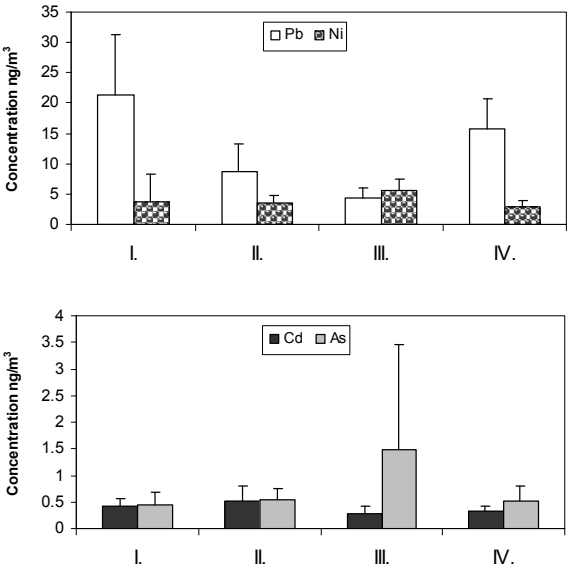


Figure 3. Lead, nickel, cadmium and arsenic concentrations in PM10 at the urban site of Győr during during the four sampling periods in 2011

Higher average mass concentrations of Pb were found in winter samples than in samples collected in non-heating seasons. The As concentrations were higher in summer due to the two highest concentrations (6.64 and 4.96 ng/m^3) detected in the third sampling period. However, no significant temporal trend was observed for Ni and Cd.

The wide range of obtained data for concentration of Ni in the first sampling period can be explained by the result of only one sample. Excluding the maximum value of 19.53 ng/m^3, the Ni level in the first sampling period was less than 5 ng/m^3.

3.3. Comparison with air quality standards

Humans can be adversely affected by exposure to air pollutants in ambient air. The European Union (EU) and the Hungarian air quality standards as well as the Hungarian air quality index (AQI) used in this study are summarised in Table 2.

The annual average mass concentration of PM10 (47.43 μg/m^3) slightly exceeded the EU and the equivalent Hungarian guideline value (40 μg/m^3). This limit is 2 times higher than the World Health Organization (WHO) air quality guideline value for PM10 (20 μg/m^3) [2]. The annual average concentrations of each trace element (Pb, Cd, As and Ni) observed for Győr were below the permitted levels. The AQI results show that the study area has excellent air quality with respect to metal(loid)s. However, the annual AQI for PM10 indicate a polluted area. The comparison of the 24-hour PM10 concentrations with the 24-hour AQI values show acceptable, polluted or heavily polluted results in heating season, while indicate excellent or good air quality in the samplings of spring and summer.

Table 2. Air quality standards and index for PM10 and the measured trace elements

Pollutant	***EU standard [3]***	***Hungarian standard [9]***	***Quality index [10]***				
			1. Excellent	***2. Good***	***3. Acceptable***	***4. Polluted***	***5. Heavily polluted***
PM10 (24-hour average) μg/m^3	50	50	0–20	20–40	40–50	50–90	90–
PM10 (annual average) μg/m^3	40	40	0–16	16–32	32–40	40–80	80–
Trace elements[a] (annual average) ng/m^3							
Pb	500	300	0–120	120–240	240–300	300–600	600–
Cd	5[b]	5	0–2	2–4	4–5	5–10	10–
As	6[b]	10	0–4	4–8	8–10	10–20	20–
Ni	20[b]	25	0–10	10–20	20–25	25–50	50–

[a]: Measured as contents in PM10 [b]: Target value

3.4. Comparison with other cities

Figs. 4 and 5 compare Győr PM10 and metal(loid) concentrations with results from other Hungarian cities (see also Fig. 1) [10]. The annual average PM10 concentrations exceeded the EU limit in some Hungarian cities in 2011. The exposure limit excess is probably due to the traffic and domestic heating, collectively [3]. The highest 24-hour concentration (183.3 μg/m^3) was detected in an urban/traffic area of Budapest [10]. The air quality for PM10 in the different Hungarian cities was acceptable or polluted. However, the AQI for trace elements were excellent in all Hungarian cites. The Pb, Cd and As levels were the highest in the industrial area of Miskolc. The annual average

concentration of Ni was the highest in Győr compared to the other Hungarian cities, which may be explained by the local emission sources.

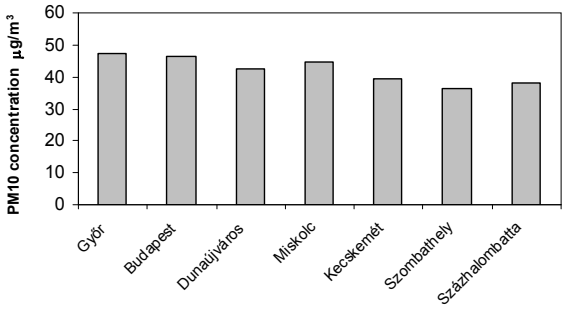


Figure 4. PM10 annual average concentrations observed for Győr with other Hungarian cities in 2011

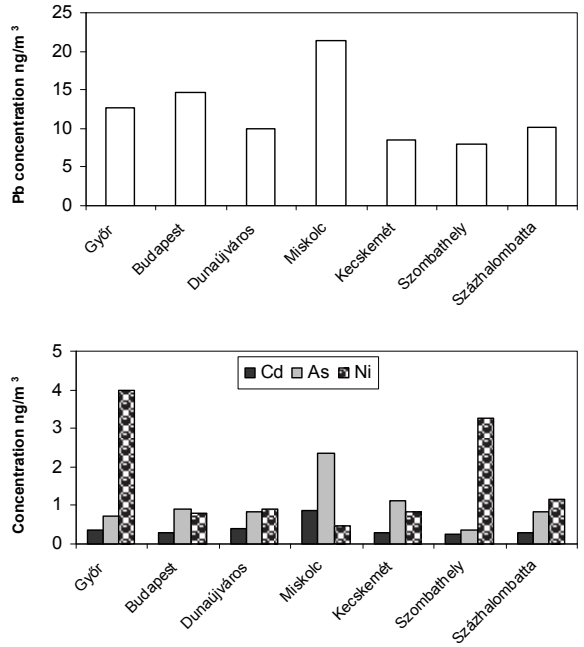


Figure 5. Trace metal(loid) annual average concentrations observed for Győr with other Hungarian cities in 2011

Although a direct comparison of literature data is difficult due to the analytical methods used, the different detection limits, the different sampling years and several other different factors such as urban or suburban population, traffic density, industrial activities, energy production, meteorological and atmospheric conditions, the concentrations of PM10 and trace elements in the urban area of Győr were also compared to the concentrations observed for published data of other cities around the

world. The European cities listed in Table 3 have good air quality for PM10 compared to the other cities located in Asia or central and southern America. Győr has lower ambient metal(loid) concentrations than several other cities around the world listed in Table 3. The comparison of the data between the cities indicates that the pollution concentrations are strongly dependent on size of urban area and population, and economic potential, which are in a different order of magnitude.

Table 3. Comparison of the PM10 and measured trace metal(loid) concentrations of Győr with results obtained in other cities around the world

Sampling site	***Type of site***	***Sampling year***	***PM10 μg/m³***	***Trace elements ng/m³***				***Reference***
				Pb	***Cd***	***As***	***Ni***	
Europe								
Győr, Hungary	Urban	2011	11–119.14 47.43±31.65	1.34–40.75 12.61±8.77	0.05–1.34 0.38±0.2	0.24–6.64 0.74±1.07	1.03–19.53 3.97±2.71	This study
Wien, Austria	Urban	2004	8–124 33±18	2–77 11±10	0.1–1.5 0.5±0.3	0.2–16 1.2±1.7	0.2–51 9.9±6.4	[11]
Barcelona, Spain	Urban	1999-2000	20–119 50	22–467 149	ND	ND	1–38 7	[12]
Palermo, Italy	Suburban	2005	11–83 25	1.8–32 9.8	ND	0.02–7.4 1.8	0.1–16 4.6	[13]
Venice, Murano, Italy	Industrial	2001-2003	1.3–216 45	2.3–523 161	1.0–326.3 170	0.3–424 181	2.4–179 74	[14]
Basel, Switzerland	Suburban	1998-1999	24.8	21	0.03	0.44	2.2	[15]
Athens, Greece	Urban	Summer / winter 2003	13–25 19±3 / 16–109 56±25	3.9–9 6.4±1.5 / 4.6–36.7 19.5±9.8	0.05–1.07 0.5±0.3 / 0.04–1.4 0.3±0.4	ND / 8.5–38.1 14.7±7.3	3.5–45.1 14.7±11.4 2.2–14 8.2±4	[16]
Thessaloniki, Greece	Urban and industrial	Winter 2006-2007	29–237 63±69	7.65–163 39.4	0–84.8 13.1	0–35.9 6.6	0.43–67.5 14.8	[17]
Edinburgh, UK	Urban	1999-2000	7.3–29.1 14.2	1.28–130 14.1	0–10.1 0.34	0.13–1.49 0.37	0.89–37.9 3.43	[18]
Belgrade, Serbia	Urban	2003-2005	2.8–333.8 68.4±46.4	0.5–152.5 46.5±128.5	0–17.7 1.4±2.2	ND	0.4–107.7 17.7±17.7	[19]
America								
Rio de Janeiro, Brazil	Suburban	2004-2005	71–312 169±42	0–69 15.9	0–1.6 0.4	ND	0–7.6 2.1	[20]
Puebla City, Mexico	Urban	2008	199.2	38.4	2.76	22.5	13.3	[21]
Costa Rica, Central America	Urban	2010-2011	55±15	11.5±3.9	ND	ND	2.1±0.8	[22]
Tuscon, Arizona, USA	Urban and industrial	2008-2009	1.5–152.9 25.5±15.5	0–12.1 2.8±1.6	0–4.8 0.1±0.5	0–6.2 0.3±0.8	0–23.4 0.7±1.9	[23]
Asia								
Beijing, China	Urban	Summer / winter 2002-2003	24–462 172±102 / 29–446 184±131	ND 110±90 / ND 370±370	ND 2.4±2.8 / ND 15.2±20.6	ND 20±20 / ND 60±70	ND 40±30 / ND 110±110	[24]
Taj Mahal, Agra, India	Suburban and industrial	2007-2008	31–362 155±78	0–109 26±22	0–52.9 23.5±11.9	0–220 50±40	0–150 50±30	[25]
Seoul, Korea	Urban and industrial	2002	18–279 79±60	39–401 200±97	2.2–30.4 5.5±6.7	ND	0.3–313 46±87	[26]

ND: No Data

A summary report on air quality in Europe [3] has highlighted that the EU limit values for PM10 were exceeded widely in 2011 according to the data of the European air quality database. The annual limit value for PM10 was exceeded most often in Poland, Italy, Slovakia, the Balkan region, Turkey and also in several urban regions. The daily limit value was exceeded in other cities in those countries, as well as in many other countries in central, western and southern Europe. Cities in Latvia, Sweden and the United Kingdom also exceeded the daily limit value for PM10. However, the human exposure to Pb, Cd, As, and Ni ambient air concentrations above the limit or target values is a local problem, and typically caused by specific industrial plants.

4. Conclusions

The concentrations of PM10 and the most monitored and regulated airborne particulate trace metal(loid)s (Pb, Cd, As and Ni) were determined in an urban site of Győr during four sampling periods in 2011. The annual average concentration of PM10 indicates a polluted area. However, the concentrations of PM10 exhibit large variability during the sampling periods. The PM10 levels indicate excellent or good air quality in the sampling periods without heating. Levels of airborne metal(loid)s in Győr area were relatively low, similar to the values reported for not polluted cities. The annual average concentrations of each trace element were below the EU and the Hungarian air quality standards.

Acknowledgement

We are indebted to József Erdős, István Vass, Bálint Kauker, Zsuzsanna Károly Némethné, Tünde Takács Kovácsné, Lajosné Bakódy and Péter Lautner (North Transdanubian Regional Environmental Protection and Nature Conservation Inspectorate Laboratory, Hungary) for chemical analyses, data and site information. We also thank József Erdős for field support.

References

[1] WHO: *Air quality guidelines for Europe*, World Health Organization, Regional Office for Europe, Copenhagen, 2000

[2] WHO: *WHO air quality guidelines*, *global update 2005*, World Health Organization; Regional Office for Europe, Copenhagen, 2005

[3] EEA: *Air quality in Europe - 2013 report*, European Environment Agency, Luxembourg, 2013

[4] Johansson, C., Norman, C., Burnan, L.: *Road traffic emission factors for heavy metals*, Atmospheric Environment, vol. 43, no. 31, pp. 4681–4688, 2009
DOI: 10.2181/036.043.0202

[5] Gaffney, J.S., Marley, N.A.: *The impacts of combustion emissions on air quality and climate - from coal to biofuels and beyond*, Atmospheric Environment, vol. 43, no. 1, pp. 23–36, 2009
DOI: 10.1016/j.atmosenv.2008.09.016

[6] Digitel Electronic AG: *Digitel high volume aerosol sampler*, Manual, Hegnau, Switzerland, 2010

[7] EN 12341: *Air quality - Determination of the PM10 fraction of suspended particulate matter - Reference method and field test procedure to demonstrate reference equivalence of measurement methods,* 1998

[8] MSZ EN 14902:2006: *Ambient air quality. Standard method for the measurement of Pb, Cd, As and Ni in the PM10 fraction of suspended particulate matter* (in Hungarian)

[9] 4/2011 (I.14.) VM.: *Guidelines for the air load levels and the stationary point source emissions* (in Hungarian)

[10] OMSZ ÉLFO: *Summary of the OLM PM10 sampling program in 2011*, Reference Centre for Air Quality Protection, 2012 (in Hungarian)

[11] Limbeck, A., Handler, M., Puls, C., Zbiral, J., Bauer,Puxbaum, H.: *Impact of mineral components and selected trace metals on ambient PM10 concentrations*, Atmospheric Environment, vol. 43, pp. 530–538, 2009
DOI: 10.1016/j.atmosenv.2008.10.012

[12] Querol, X., Alastuey, A., Rodriguez, S., Plana, F., Ruiz, C.R., Cots, N., Massague, G., Puig, O.: *PM10 and PM2.5 source apportionment in the Barcelona Metropolitan area, Catalonia, Spain*, Atmospheric Environment, vol. 35, no. 36, pp. 6407–6419, 2001
DOI: 10.1016/S1352-2310(01)00361-2

[13] Dongarra, G., Manno, E., Varrica, D., Vultaggio, M.: *Mass levels, crustal component and trace elements in PM10 in Palermo, Italy*, Atmospheric Environment, vol. 41, pp. 7977–7986, 2007
DOI: 10.1016/j.atmosenv.2007.09.015

[14] Rossini, P., Matteucci, G., Guerzoni, S.: *Atmospheric fall-out of metals around the Murano glass-making district (Venice, Italy)*, Environmental Science and Pollution Research, vol. 17, no. 1, pp. 40–48, 2010
DOI: 10.1007/s11356-009-0122-8

[15] Hueglin, C., Gehrig, R., Baltensperger, U., Gysel, M., Monn, C., Vonmont, H.: *Chemical characterisation of PM2.5, PM10 and coarse particles at urban, near-city and rural sites in Switzerland*, Atmospheric Environment, vol. 39, pp. 637–651, 2005
DOI: 10.1016/j.atmosenv.2004.10.027

[16] Vassilakos, C., Veros, D., Michopoulos, J., Maggos, T., O' Connor, C.M.: *Estimation of selected heavy metals and arsenic in PM10 aerosols in the ambient air of the Greater Athens Area, Greece*, Journal of Hazardous Materials, vol. 140, no. 1-2, pp. 389–398, 2007
DOI: 10.1016/j.jhazmat.2006.11.002

[17] Terzi, E., Argyropoulos, G., Bougatioti, A., Mihalopoulos, N., Nikolaou, K., Samara, C.: *Chemical composition and mass closure of ambient PM10 at urban sites,* Atmospheric Environment, vol. 44, no.18, pp. 2231–2239, 2010
DOI: 10.1016/j.atmosenv.2010.02.019

[18] Heal, M.R., Hibbs, L.R., Agius, R.M., Beverland, I.J.: *Total and water soluble trace metal content of urban background PM10, PM2.5 and Black Smoke in Edinburgh, UK*, Atmospheric Environment, vol. 39, pp. 1417–1430, 2005
DOI: 10.1016/j.atmosenv.2004.11.026

[19] Rajšić, S., Mijić, Z., Tasić, M., Radenković, M., Joksić, J.: *Evaluation of the levels and sources of trace elements in urban particulate matter*, Environmental Chemistry Letters, vol. 6, no. 2, pp. 95–100, 2008
DOI: 10.1007/s10311-007-0115-0

[20] Toledo, V.E., Almeida, P.B., Quiterio, S.L., Arbilla, G., Moreira, A., Escaleira, V., Moreira, J.C.: *Evaluation of levels, sources and distribution of toxic elements in PM10 in a suburban industrial region, Rio de Janeiro, Brazil*, Environmental Monitoring and Assessment, vol. 139, no. 1-3, pp. 49–59, 2008
DOI: 10.1007/s10661-007-9815-y

[21] Morales-Garcia, S. S., Rodriguez-Espinosa, P. F., Jonathan, M. P., Navarrete-Lopez, M., Herrera-Garcia, M. A., Munoz-Sevilla, N. P.: *Characterization of As and trace metals*

embedded in PM10 particles in Puebla City, Mexico, Environmental Monitoring and Assessment, vol. 186, no. 1, pp. 55–67, 2014
DOI: 10.1007/s10661-013-3355-4

[22] Murillo, J. H., Roman, S. R., Marin, J. F. R., Ramos, A. C., Jimenez, S. B., Gonzales, B. C., Baumgardner, D. G.: *Chemical characterization and source apportionment of PM10 and PM2.5 in the metropolitan area of Costa Rica, Central America*, Atmospheric Pollution Research, vol. 4, no. 2, pp. 181–190, 2013
DOI: 10.5094/APR.2013.018

[23] Foley, T., Betterton, E. A., Wolf, A.: *Ambient PM10 and Metal Concentrations Measured in the Sunnyside Unified School District, Tucson, Arizona*, Journal of the Arizona-Nevada Academy of Science, vol. 43, no. 2, pp. 67–76, 2012
DOI: 10.2181/036.043.0202

[24] Sun, Y., Zhuang, G., Wang, Y., Han, L., Guo, J., Dan, M., Zhang, W., Hao, Z.: *The air-borne particulate pollution in Beijing - concentration, composition, distribution and sources*, Atmospheric Environment, vol. 38, pp. 5991–6004, 2004
DOI: 10.1016/j.atmosenv.2004.07.009

[25] Singh, R., Sharma, B.S.: *Composition, seasonal variation and sources of PM10 from world wide heritage site Taj Mahal, Agra*, Environmental Monitoring and Assessment, vol. 184, no. 1-3, pp. 5945–5956, 2012
DOI: 10.1007/s10661-011-2392-0

[26] Kim, K.H., Mishra, V.K., Kang, C.H., Choi, K.C., Kim, Y.J., Kim, D.S., Youn, Y.H., Lee, J.H.: *The metallic composition of aerosols at three monitoring sites in Korea during winter 2002*, Environmental Monitoring and Assessment, vol. 121, no. 1-3, pp. 381–399, 2005
DOI: 10.1007/s10661-005-9136-y

Acta
Technica
Jaurinensis

Vol.7., No.2., pp. 156–171, 2014
DOI: 10.14513/actatechjaur.v7.n2.272
Available online at acta.sze.hu

Benchmarking the LGO Solver Suite within the COCO Test Environment

M. F. Hatwágner[1], J. D. Pintér[2]

[1]**Széchenyi István University**
Dept. of Information Technology, Győr, Hungary
[2]**Pintér Consulting Services, Inc., Halifax, Nova Scotia, Canada**

Abstract: Optimization problems often arise in the context of scientific-engineering research and practice. In many situations that require optimization, there is no need to develop new, highly customized software for a new problem, because there are readily usable optimization packages to choose from. Benchmarking and testing software environments can greatly assist the process of choosing an appropriate optimization tool. The benchmarking environments typically include a substantial collection of well-known and widely used test functions; they also offer a properly defined methodology to compare the solvers considered. One of these benchmarking environments is called COCO (an abbreviation that stands for "COmparing Continuous Optimizers"). COCO has been used at the annual BBOB (Black-Box Optimization Benchmarking) workshops. COCO assesses the capabilities of optimization solvers based on a given collection of test problems and evaluation criteria, under identical and fully reproducible circumstances. The goal of our present study is to benchmark the LGO (Lipschitz Global Optimizer) solver suite and to compare it to several other solvers using COCO.

Keywords: Benchmarking optimization Software, LGO Solver Suite for Nonlinear Optimization, COCO Test Environment, Black-Box optimization Benchmarking.

1. Introduction

In this article the performance of the LGO solver suite is examined and compared to several other solvers using the COCO benchmarking environment. First, the main features of COCO are reviewed. This is followed by a concise description of the key features of LGO. Next, we describe the process of solver parameterization, and conclude with presenting the results based on the COCO test problem suite and its benchmarking process.

2. A Brief Description of the COCO Benchmarking Environment

GECCO, the GEnetic and Evolutionary Computation Conference [11]) has been held since 2009 in every year: the BBOB workshop [3] has been part of these conferences (with the exception of 2011). During the BBOB workshop the COCO [5] test environment has been used to compare various global optimization methods, and corresponding software. The yearly workshops always have a specific aim and emphasis, but there has not been a substantial change in the test environment or the overall objectives.

The comparison of the solvers is done applying a quantitative and completely reproducible process. The COCO test environment uses 24 real-valued test functions selected by a group of researchers on the basis of their mathematical properties: subsequently, these functions have been used to construct both the noiseless and noisy test model suites. The detailed description of these functions can be found in [8, 9]. Every model in the test suite has only a single objective: therefore the testing of multi-objective optimization capabilities and performance is currently not possible. The COCO environment allows testing of solvers written in various programming languages including C/C++, Java, Matlab, Python and R. The current version of COCO can be downloaded from the project website [6].

After linking a given solver to the test environment, the solver can be examined using two criteria set by COCO. The first criterion is the number of objective function evaluations required to solve each test problem. The second measure used is the runtime required to solve the same set of test problems. Based on practical reasons (with a view towards real-world applications), the first of these criteria is given a stronger emphasis.

The benchmarking results are generated automatically, and then stored in formatted output files: the latter can be downloaded from the annual BBOB websites (e.g. [28, 29]) for subsequent usage and comparisons. Python scripts are also included in the COCO test environment, to support the generation of figures and tables based on the output files. These objects can be then used to prepare topical reports and articles very easily, based on pre-assembled LaTeX templates.

For more detailed information on COCO and BBOB, see the BBOB set-up instructions [12] or the COCO user documentation [10].

3. A Brief Introduction to LGO

Consider the following general non-linear programming (NLP) problem:

$$\begin{gathered} \min f(x) \\ g(x) \leq 0 \\ x_l \leq x \leq x_u \end{gathered}$$

Here x is the real n-vector of the model function arguments; x_l and x_u are correspondingly defined (given) finite lower and upper bound n-vectors; $g(x)$ is the m-component vector of the general model constraints, and f is the objective function. The vector inequalities shown in the model formulation are (evidently) interpreted component-wise. The feasible set of the NLP problem will be denoted by D, is a proper subset of R^n.

If D is non-empty and all model functions f, g are continuous then – by the classical Bolzano-Weierstrass theorem – the NLP model has a non-empty set of (globally) optimal solutions. Notice at the same time that the NLP model may well be multi-modal: that is, initiating a suitable local optimization procedure from given starting points in D can lead to results of different quality. The objective of global optimization is to find the best possible solution - which gives the global optimum – under such circumstances.

The LGO solver suite has been designed to solve(numerically) instances of the NLP model stated above. LGO seamlessly integrates a suite of global and local scope algorithms (solvers) to handle the NLP in a robust and efficient manner. Let us point out that robustness and efficiency are partially conflicting criteria: LGO's overall development strategy is admittedly biased towards robustness, i.e. the ability to handle (also) possibly very tough optimization problems. NLP problems are tackled by LGO applying an integrated sequence of component solvers which, as a rule, find solutions to difficult optimization challenges – given a sufficient number of function evaluations and program execution time. All LGO solver components apply a direct (derivative-free) approach and proceed based only on model function evaluations. This feature may be advantageous in numerous "black box" type applications in which model function values may result from a numerical procedure such as simulation, or the solution of a system of differential equations and so on. Here we assume that the numerical procedure in question depends on (i.e., the corresponding model function is parametrized by) the decision variables x of the NLP.

We also note that LGO can handle general constrained nonlinear optimization problems – as opposed to merely box-constrained ones in which the general model constraints $g(x)$ are absent. Obviously, this feature adds complexity to LGO's overall design and it also influences its efficiency when the latter is measured by the number of function evaluations or program execution speed, without the need for measuring constraint satisfaction quality. As an example, consider the solution of a highly nonlinear system of equations, say in tens of variables – a tough challenge for many population-based heuristic methods, while being a routine task for LGO and similar scope global constrained optimization engines.

The in-depth technical discussion of LGO is outside of the scope of the present article. The theoretical foundations leading to LGO and some key implementation aspects are summarized in [18]. Platform-specific software implementation versions, numerous illustrative applications and case studies are discussed e.g. in [18–23].

The core LGO implementation is available for Windows / Linux / Mac / Unix OS based computer platforms, and it can be used to solve optimization problems formulated using

the Fortran, C and C# programming languages. LGO is also available as a solver option in different optimization modelling and technical computing environments [22, 23].

Since LGO and COCO both support the use of models developed using the C language, this version was selected for the tests described here. Fig. 1 summarizes the basic structure of LGO and its key components.

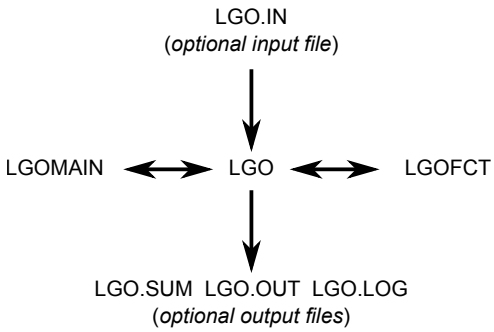


Figure 1. Structure of the LGO program.

The core LGO solver suite can be configured to use a specific input file and to generate result files, but these files are not needed in the COCO benchmarking context. The COCO environment itself generates randomized test data and also collects the results of the solver engine computations.

In the figure LGOMAIN is the main program (driver) file that calls LGO. After setting all static (i.e. once defined, then unchanging) optimization model information and a list of key solver parameters, LGOMAIN launches LGO. In the core implementation, LGO then iteratively calls LGOFCT: the latter serves to define all model functions in the NLP formulation.

Due to the setup of the COCO benchmarking framework, LGOMAIN's content has been embedded in the *MY_OPTIMIZER.c* file (the latter being a component of COCO). Furthermore, although in the core LGO implementation LGOFCT defines the model functions, in the COCO framework it is only a "placeholder" to redirect the objective function calls of the test environment.

Some of the key LGO solver settings were changed or tuned during the tests described here, therefore they are of particular interest.

One of these parameters is *opmode*, the operation mode selector. It has four possible settings:

0: Local Search (LS) only;

1: Global Branch-and-Bound (BB), followed by LS;

2: Global Adaptive Randomized Search (GARS), followed by LS;

3: Global Multi-start Randomized Search (MS), here each global search cycle is followed by LS.

For practical reasons, LGO starts with a LS phase which precedes all global search options. This is followed first by a regularly spaced sampling (RSS) presolver mode and another quick global presolver: each of these searches is followed by a LS phase. The RSS solver component is described in [24]. If further search is implied by *opmode* and it is supported by LGO's calling parameters and its internal resource allocation strategy, then one of BB + LS, GARS + LS, and MS + a set of LS phases follow.

For further details, we refer again to [18–20, 22–24].

4. Tuning of LGO's Solver Parameters

After completing the linking of LGO to the COCO test environment, some systematic tuning of the solver parameters was conducted to analyze the resulting benchmark results. All test functions were solved with the same preset parameter values during the benchmarking stages.

Let us note here that – according to the rules of COCO/BBOB studies – the modification of settings during runtime are forbidden: "On all functions the very same parameter setting must be used (which might well depend on the dimensionality, see Section Input to the Algorithm and Initialization). That means, a priori use of function-dependent parameter settings is prohibited (since 2012). In other words, the function ID or any function characteristics (like separability, multi-modality, ...) cannot be considered as input parameter to the algorithm. Instead, we encourage benchmarking different parameter settings as 'different algorithms' on the entire testbed." [17]

The above cited "rule" serves to simulate most real-life optimization situations in which in-depth problem-based insight typically cannot be used to effect solver parametrization.

4.1. The Effect of Different LGO Operational Modes

First of all, the impact of changing the value of *opmode* was examined. The recommended default setting is 3 corresponding to MS + LS [23]. We only advise the setting *opmode* = 0, if we want only a typically quick LS started from a given initial point.

After each of the four possible values were tested, it was concluded that there is no significant difference between the different operation mode based results, when solving the COCO test models . This finding is in line with LGO's overall development philospophy;

it is also supported by the fact that LGO does substantial search before switching to one of its BB + LS, GARS + LS or MS + LS modes. As a rule, the most resource-intensive global multi-start search and local search (MS + LS) mode produced the best results almost in all cases considering the proportion of the solved test function instances. The results are shown by Fig. 2.

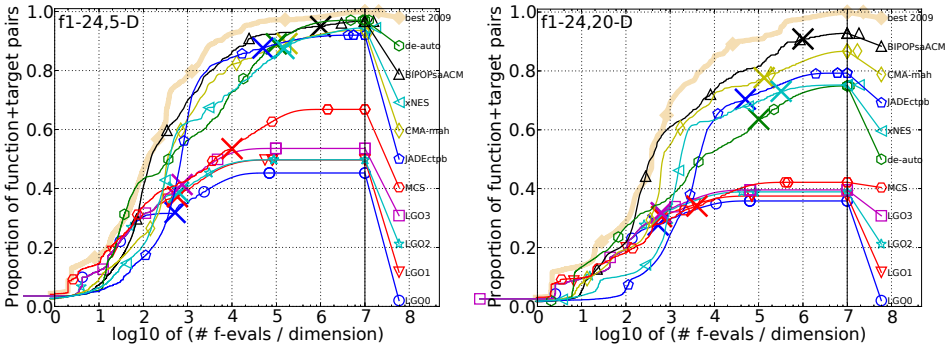


Figure 2. Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension for 50 targets in $10^{[-8..2]}$ for all functions and subgroups in 5-D and 20-D. The "best 2009" line corresponds to the best ERT observed during BBOB 2009 for each single target. The number after "LGO" represents the selected operation mode.

The performance measure used in the figure is ERT (the Expected value of Running Time based on randomized starting points). Directly citing the COCO developers' statement: ERT is quantitative, well-interpretable, relevant with respect to the "real world" and as simple as possible. The details of it [27] and the advantage of using it [1] are documented in the corresponding papers. The ERT computes a single measurement from a data sample set. Bootstrapping [7] can provide a dispersion measure for this aggregated measurement. COCO computes the running time of the single sample as the sum of function evaluations in the drawn trials (for the last trial up to where the target function value was reached) [1, 2].

The results of LGO were compared with some other prominent solvers in Fig. 2. The selection criteria for the solvers included were the following:

- MCS [14] has been included, because it can achieve similarly good performance to LGO (based on the in-depth performance study of [30] in which LGO and MCS were among the best derivative-free solvers among two dozen solvers when solving more than 500 test problems.
- The other solvers have been selected as the most successful ones in the BBOB 2012 competition, based on solving the COCO set of test models.

This way, 1+5 solvers were selected and indicated in the figure in addition to LGO (although it is sometimes not easy to differentiate among the plots belonging to the different solvers).

4.2. Experimenting with the Number of Objective Function Evaluations

As the figures indicate, at the beginning of the optimization process LGO produces nearly the same results as MCS, and achieves better results than many competitors such as BIPOPsaACM, CMA-mah or JADEctpb. However, in the later search phase – in the given resource / time frame and on the given test function set, with a given definition of required presicion for successful solution – MCS and most of the selected competitors surpass LGO.

Let us point out here again that the BBOB solvers have been specifically developed to solve merely box-constrained problems which make up the COCO test suite. More importantly – unlike many popular heuristic algorithms – LGO has a proper theoretical convergence guarantee for each of its global scope search methods: however, in theory the number of model function evaluations can (even should) go to infinity...

From this time on, the operation mode of LGO was fixed to multi-start random search and local search (MS + LS), and the other parameters were simply selected (pre-tuned). Specifically, we studied settings for the following LGO solver parameters: *g_maxfct* and *l_maxfct* define the maximal, total number of function evaluations in the selected global search phase and each local search phase, respectively. Moreover, *max_nosuc* determines the maximal number of function evaluations in the selected global search phase without improvement, before switching to the subsequent local search phase.

Several benchmarking trials were conducted using the same values for *g_maxfct, l_maxfct*, and *max_nosuc*. The default values of the parameter settings are actually dependent on the model size. We then heuristically applied the multiplier values 200, 400, 800, and 1600 to a simple estimate of model complexity, [23].

The increasing values of the parameters aided LGO to raise the proportion of the solved test functions. In 5D, it increased the proportion with approx. 5%, in 20D with approx. 3-4%. The latter value is practically the same as the value of MCS. (See Fig. 3).

4.3. Different Number of Objective Function Evaluations Allowed in the Various Optimization Phases

To decrease the number of objective function evaluations and thereby make the solver more effective, the value of *l_maxfct* was modified to 800 and 400 while the values of *g_maxfct* and *max_nosuc* remained the same: 1600. However, this did not result in a better overall performance as measured within COCO, see Fig. 4.

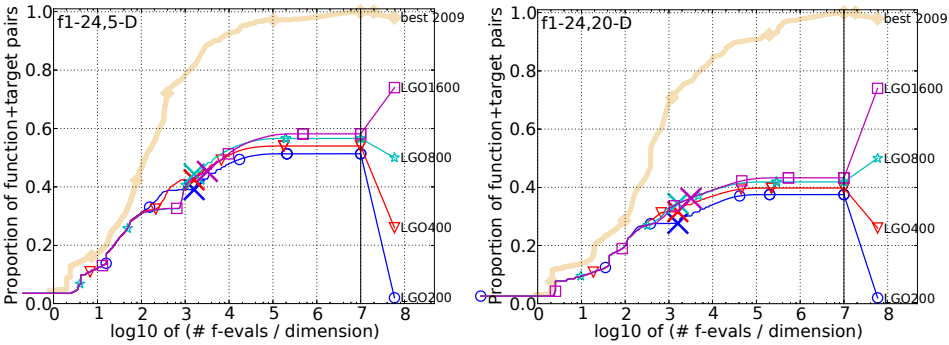


Figure 3. Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension for 50 targets in $10^{[-8..2]}$ *for all functions and subgroups in 5-D and 20-D. The "best 2009" line corresponds to the best ERT observed during BBOB 2009 for each single target. The number after "LGO" represents the values of the examined solver parameters (*g_maxfct, l_maxfct, max_nosuc*).*

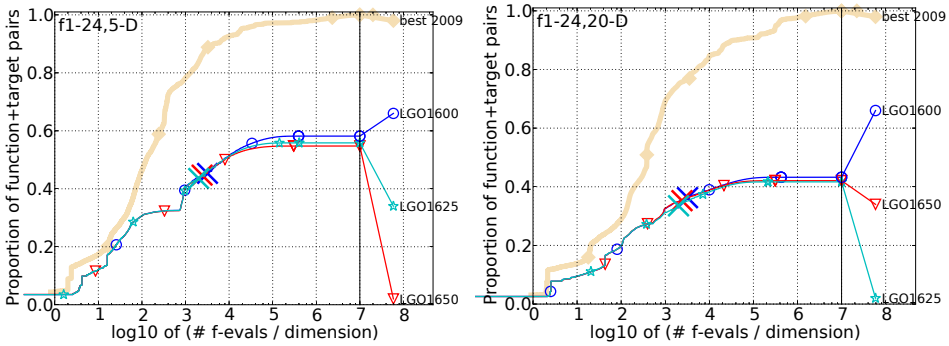


Figure 4. Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension for 50 targets in $10^{[-8..2]}$ *for all functions and subgroups in 5-D and 20-D. The "best 2009" line corresponds to the best ERT observed during BBOB 2009 for each single target. The value of* g_maxfct *and* max_nosuc *was 1600 in all cases. The values of* l_maxfct *was 1600 in plot "LGO1600", 800 in plot "LGO1650" and only 400 in plot "LGO1625".*

4.4. Comparing LGO with GLOBAL, MCS and NEWUOA

Next, the same value (1600) was used again for *g_maxfct, l_maxfct* and *max_nosuc*. In Fig. 5 and 6 the results of LGO are compared with the results of GLOBAL [4, 16], MCS [13, 14] and NEWUOA [26, 31].

GLOBAL is a multistart type stochastic method that involves a combination of sampling, clustering, and local search. It reduces the number of local search steps using clustering. The Multilevel Coordinate Search (MCS) algorithm was inspired by the DIRECT method [15]. MCS adaptively splits the search space into smaller boxes. The partitioning procedure is aimed at those boxes in which lower function values are expected to be found. By starting a local search from selected good points, an improved result is often obtained. Finally, NEWUOA is an iterative method that uses a quadratic model which is used in a trust region procedure for adjusting its variables and to approximate the value of the objective function.

If models with separable functions are to be solved, then MCS is the clear winner in 5D. The other three algorithms perform rather similarly: the difference in the proportion of the solved problems is less than 5%. In 20D, MCS is still in winning position, but the differences of the final success rates are less than about 6%.

LGO solved approx. 73% of the moderate test function instances (within the given time frame), but in this case the competitors had better performance results. The efficiency of the other three algorithms was very similar. In case of 20D NEWUOA is the winner, while LGO is the second best with a success rate of approx. 65% (again and again, within the given time frame. . .).

One can notice that in the initial phase of optimizing the ill-conditioned COCO test functions in 5D LGO shows strong performance, but later its convergence speed decreased and it achieved only the third place. MCS was the loser of this kind of benchmark in 5D and 20D as well.

LGO performed best if the goal was to solve multi-modal functions. In 5D, the differences between the results of LGO, MCS and GLOBAL are minimal, while NEWUOA – due to its admittedly more local search scope – seems to have a serious drawback. LGO is the winner in 20D, too. Let us recall here the important fact global optimization is the key intended target application area for using LGO.

If the functions are weakly structured multi-modal, LGO is only the third, and the convergence is slow in 5D. It is very suprising, that the winner in 5D (MCS) is only the third in 20D. In the later case, NEWUOA is the best and LGO is the last (approx. 34% of instances solved).

In general, MCS proved to be the best algorithm in 5D cases. LGO was the last and it was able to solve about 58% of the function instances. It is a bit strange, but MCS was the

worst algorithm in 20D. NEWUOA was the best algorithm, and LGO placed third. LGO solved approx. 42% of the test function instances.

It can be seen, that LGO performs well if the goal is to determine the optima of multi-modal functions, while NEWUOA and GLOBAL are able to solve more of the COCO tests using the same number of function evaluations. It should be pointed out again that this finding essentially depends on the COCO test model suite, since the far more detailed study [30] places each of LGO, MCS, and NEWUOA among its "winning solvers". We will cite the summary "verdict" of the latter study in our Conclusions below.

5. Conclusions

Benchmarking optimization software can be very helpful for users to select appropriate tools to handle their models – especially in the context of this article when the model class considered is extremely broad, covering from relatively simple problems to some hard global optimization challenges.

In the most general terms, the objective of the LGO software is to provide an overall robust and efficient suite of solver modules which help to obtain a high-quality feasible solution, across a significant range of general constrained nonlinear – global and local – optimization problems. Recall that this is done in a derivative-free setting: i.e. LGO does not require model function gradients or higher order information. This feature makes it applicable even to completely "black box" models – as long as these are defined by continuous functions.

Given this broadly defined scope mandate, it is not too surprising that on a small set of test models chosen (or perhaps even "fabricated") by other researchers LGO sometimes does better, but at other times does not, than other good quality solvers such as GLOBAL, MCS or NEWUOA. LGO, MCS and NEWUOA are among the most successful bound-constrained solvers, as per the recent in-depth numerical performance study of Rios and Sahinidis [30]. This study is based on a far more extensive set of test models than the COCO test model collection: from among more than 20 solvers LGO, MCS, and NEWUOA are the best performers with some other solver not discussed here. Below we cite the study of Rios and Sahinidis [30]:

"This paper addresses the solution of bound-constrained optimization problems using algorithms that require only the availability of objective function values but no derivative information... The paper presents a review of derivative-free algorithms, followed by a systematic comparison of 22 related implementations using a test set of 502 problems. The test bed includes convex and nonconvex problems, smooth as well as nonsmooth problems. The algorithms were tested under the same conditions and ranked under several criteria, including their ability to find near-global solutions to nonconvex problems, improve a given starting point, and refine a near-optimal solution. A total of 112,448 problem instances were solved. We find that the ability of all these solvers to obtain good solutions diminishes

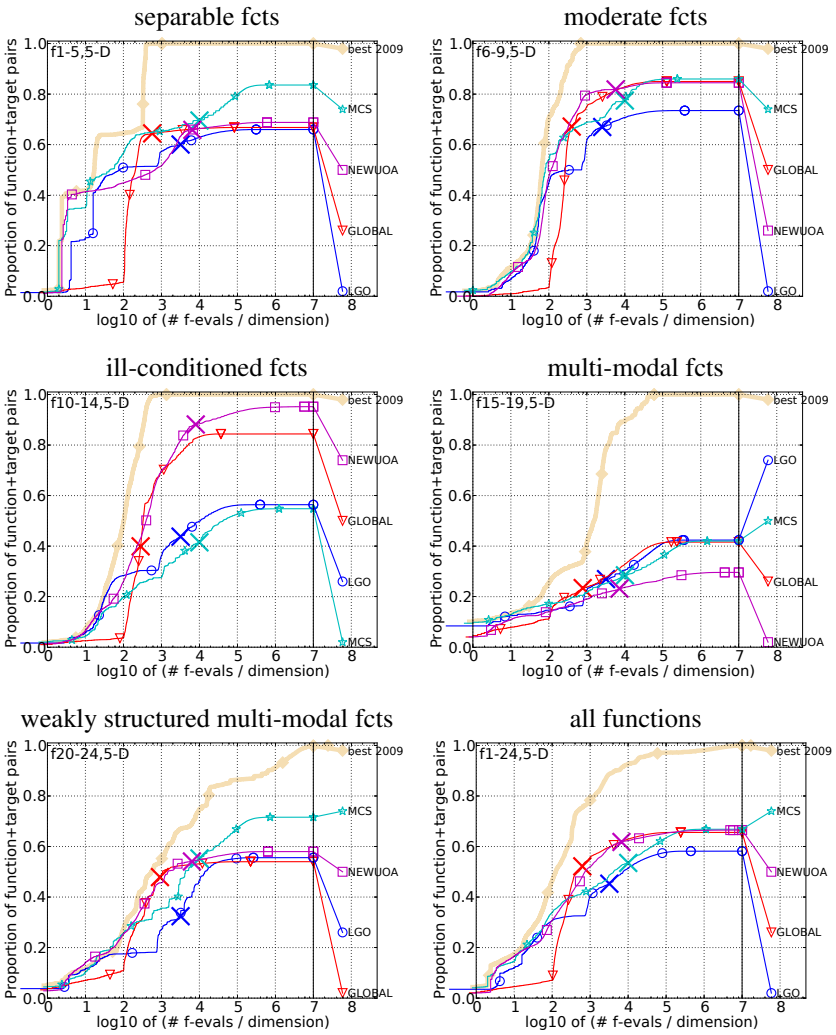


Figure 5. Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension for 50 targets in $10^{[-8..2]}$ for all functions and subgroups in 5-D. The "best 2009" line corresponds to the best ERT observed during BBOB 2009 for each single target.

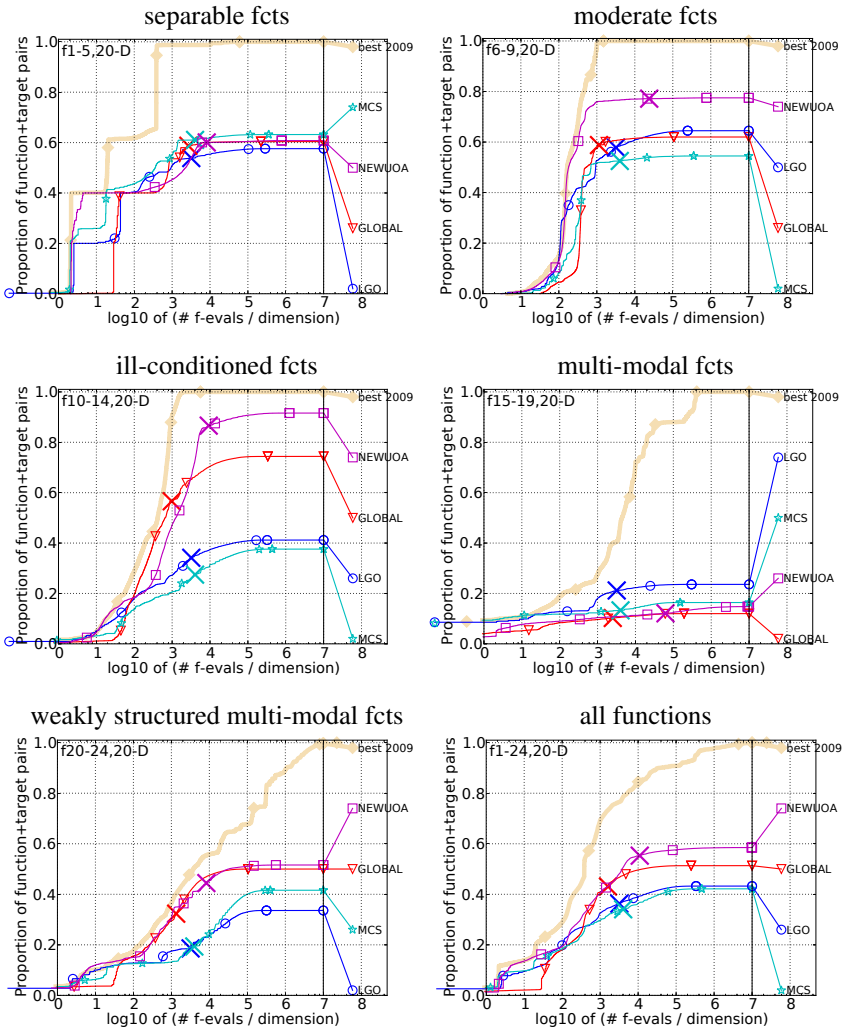


Figure 6. Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension for 50 targets in $10^{[-8..2]}$ for all functions and subgroups in 20-D. The "best 2009" line corresponds to the best ERT observed during BBOB 2009 for each single target.

with increasing problem size. For the problems used in this study, TOMLAB/MULTIMIN, TOMLAB/GLCCLUSTER, MCS and TOMLAB/LGO are better, on average, than other derivative-free solvers in terms of solution quality within 2500 function evaluations, while TOMLAB/OQNLP, NEWUOA and TOMLAB/MULTIMIN show superior performance in terms of refining a near-optimal solution."

(The TOMLAB/LGO implementation mentioned above is based on a several year old core LGO solver.)

Cf. also the recent study by Pintér and Kampas [25] which is based on a markedly different – and again more diverse – collection of models than the COCO test suite.

To conclude, our present study summarizes the numerical experiments conducted in the COCO benchmarking environment, highlights their key findings. A set of heuristically tested ("optimized") LGO solver parameter values was chosen and used. We also compared LGO with several other prominent solvers on the COCO test function set. We found that if the goal is to optimize highly multi-modal functions then LGO performed well; in other (arguably, simpler) cases the "competitors" often outperformed LGO to some extent.

We plan to expand on our own benchmarking studies, and we will continue to report results for more comprehensive test problem suites. Among our collected test problems we have included also some truly difficult scalable model-classes that – in addition to their real-life background and relevance – pose a challenge to the best state-of-the art nonlinear solvers. We also include tests on general constrained (i.e., not only box-constrained) models which can greatly surpass in difficulty their box-constrained (embedded) sub-models. The results of these tests will appear in forthcoming work.

Acknowledgement

The Project leading to this study has been supported by the Hungarian Government and co-financed by the European Social Fund: TÁMOP-4.2.2.A-11/1/KONV-2012-0012: Basic research for the development of hybrid and electric vehicles.

References

[1] Auger, A., Hansen, N.: *Performance evaluation of an advanced local search evolutionary algorithm*, in Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2005), pp. 1777–1784, 2005

[2] Auger, A., Ros, R.: *Benchmarking the pure random search on the BBOB-2009 testbed*, in GECCO (Companion) ACM, Editor: Franz Rothlauf, pp. 2479–2484, 2009

[3] Black-Box Optimisation Benchmarking homepage, *http://coco.gforge.inria.fr/doku.php?id=bbob-2013*

[4] Csendes, T., Pál, L., Sendin, J. O. H., and Banga, J. R.: *The GLOBAL Optimization Method Revisited*, Optimization Letters, Vol. 2, pp. 445–454, 2008

[5] Documentation site of COCO (COmparing Continuous Optimisers), *http://coco.gforge.inria.fr/doku.php*

[6] Download page of COCO, *http://coco.lri.fr/downloads/download13.09/bboball13.09.tar.gz*

[7] Efron, B., Tibshirani, R.: *An introduction to the bootstrap*, Chapman & Hall/CRC, 1993

[8] Finck, S., Hansen, N., Ros, R., Auger, A.: *Real-Parameter Black-Box Optimisation Benchmarking 2010: Presentation of the Noiseless Functions*, Working Paper 2009/20, compiled April 13, 2013, *http://coco.lri.fr/downloads/download13.09/bbobdocfunctions.pdf*

[9] Finck, S., Hansen, N., Ros, R., Auger, A.: *Real-Parameter Black-Box Optimisation Benchmarking 2010: Presentation of the Noisy Functions*, Working Paper 2009/21, compiled April 13, 2013, *http://coco.lri.fr/downloads/download13.09/bbobdocnoisyfunctions.pdf*

[10] Finck, S., Ros, R.: *COCO (COmparing Continuous Optimisers) Software: User Documentation*, compiled April 13, 2013, *http://coco.lri.fr/downloads/download13.09/bbobdocsoftware.pdf*

[11] Genetic and Evolutionary Computation Conference homepage, *http://www.sigevo.org/gecco-2013/*

[12] Hansen, N., Auger, A., Finck, S., Ros, R.: *Real-Parameter Black-Box Optimisation Benchmarking: Experimental Setup*, compiled April 13, 2013, *http://coco.lri.fr/downloads/download13.09/bbobdocexperiment.pdf*

[13] Huyer, W., Neumaier, A.: *Benchmarking of MCS on the Noiseless Function Testbed*, Manuscript, 2009, *http://www.mat.univie.ac.at/ neum/ms/mcs_exact.pdf*

[14] Huyer, W., Neumaier, A.: *Global Optimization by Multilevel Coordinate Search*, Journal of Global Optimization, Vol. 14, pp. 331–355, 1999

[15] Jones, D.R., Perttunen, C.D., Stuckman, B.E.: *Lipschitzian optimization without the Lipschitz Constant*, Journal of Optimization Theory and Applications, Vol. 79, pp. 157–181, 1993

[16] Pál, L., Csendes, T., Markót, M. C., and Neumaier, A.: *Black-box optimization benchmarking of the GLOBAL method*, Evolutionary Computation, Vol. 20, pp. 609–639, 2012

[17] Parameter Setting and Tuning of Algorithms, part of BBOB 13.09 documentation, *http://coco.lri.fr/COCOdoc/bbo_experiment.html#sec-tuning*

[18] Pintér, J. D.: *Global Optimization in Action*, Kluwer Academic Publishers, Dordrecht, 1996

[19] Pintér, J. D.: *LGO: A program system for continuous and Lipschitz optimization*, in Developments in Global Optimization, Editors: Bomze, I.M., Csendes, T., Horst, R. and Pardalos, P.M., Kluwer Academic Publishers, Dordrecht, pp. 183–197, 1997

[20] Pintér, J. D.: *Global optimization: Software, test problems, and applications*, in Handbook of Global Optimization, Volume 2, Editors: Pardalos, P. M. and Romeijn, H. E., Kluwer Academic Publishers, Dordrecht, pp. 515–569, 2002

[21] Pintér, J. D., Editor: *Global Optimization – Scientific and Engineering Case Studies*, Springer Science + Business Media, New York, 2006

[22] Pintér, J. D.: *Software development for global optimization*, in Global Optimization: Methods and Applications, Fields Institute Communications Volume 55, Editors: Pardalos, P.M. and Coleman, T. F., American Mathematical Society, Providence, RI, pp. 183–204, 2009

[23] Pintér, J. D.: *LGO – A Model Development and Solver System for Global-Local Nonlinear Optimization*, User's Guide (Current version: June 2013), Pintér Consulting Services, Inc. Canada; www.pinterconsulting.com, 2013

[24] Pintér, J. D., Horváth, Z. *Integrated experimental design and nonlinear optimization to handle computationally expensive models under resource constraints*, Journal of Global Optimization, Vol. 57, pp. 191–215, 2013

[25] Pintér, J. D., Kampas, F. J.: *Benchmarking nonlinear optimization software in technical computing environments: Global optimization in Mathematica with MathOptimizer Professional*, TOP (An official journal of the Spanish Society of Statistics and Operations Research), Vol. 21, pp. 133–162, 2013

[26] Powell, M. J. D.: *The NEWUOA software for unconstrained optimization without derivatives*, in Large Scale Nonlinear Optimization, Editors: Di Pillo, G. and Roma, M., Springer Science + Business Media, New York, pp. 255–297, 2006

[27] Price, K.: *Differential evolution vs. the functions of the second ICEO*, in Proceedings of the IEEE International Congress on Evolutionary Computation, pp. 153–157, 1997

[28] Raw data files of BBOB 2012,
http://coco.lri.fr/BBOB2012/rawdata/

[29] Raw data files of BBOB 2013,
http://coco.lri.fr/BBOB2013/rawdata/

[30] Rios, L. M., Sahinidis, N. V.: *Derivative-free optimization: a review of algorithms and comparison of software implementations*, Journal of Global Optimization, Vol. 56, pp. 1247–1293, 2013, DOI: 10.1007/s10898-012-9951-y

[31] Ros, R.: *Benchmarking the NEWUOA on the BBOB-2009 function testbed*, in: Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, ACM, pp. 2421–2428, 2009, ISBN: 978-1-60558-505-5, DOI: 10.1145/1570256.1570338

Acta
Technica
Jaurinensis

Vol. 7., No.2., pp. 172-182, 2014
DOI: 10.14513/actatechjaur.v7.n2.281
Available online at acta.sze.hu

Up-to-Date Finite Element Based Simulation of Permanent Magnet

G. Kovács

Széchenyi István University, Department of Automation
Egyetem tér 1, H-9026 Győr, Hungary
e-mail: kovacsg@maxwell.sze.hu

Abstract: The paper presents a two dimensional finite element based solver for static magnetic field problems. The application has been written in C programming language. Magnetic vector potential of some reference models have been calculated by the help of two environments. The first one is the MATLAB environment and the second one is the C programming language based finite element code. The simulation results of the two environments were compared to each other focusing on the magnetic vector potential and the simulation time. Differences of the simulation results are showed in this paper, as well.

Keywords: Finite Element Method, C programming language, MATLAB, permanent magnet

1. Introduction

The computer-aided design is one of the important parts of the electric engine development. Electric engines have been improved at the Széchenyi István University, as well and one of the parts of this development is to design and optimization a brushless DC (direct current) motor family, which are will be applied with bicycles and smaller motors. There are also two projects where these motors will be applied. One of them is the development of hybrid E-VAN, which is an adapted Ford-truck. The second one is an electric car which is an individual development by the students and teachers of the university. The main aspect of this PMS motor (Permanent Magnet Synchronous motor) development is to reduce the weight and the size of the engine but the torque and losses of the motor should not decrease. There are different ways to design the PMS motors. This paper shows a C programming language based finite element simulation of the permanent magnets which are used in permanent magnet synchronous motors.

2. Structure of the development environment

The new finite element package consists of two main parts which can be seen in Fig. 1. The first part is the GMSH software [1] which is a two and three-dimensional finite element mesh generator with built-in pre- and post-processing facilities. The second part of the finite element development environment is the C language based package under

Linux operating system, where partial differential equations and matrix operations can be solved. The finite element based solver has been written by the help of this package.

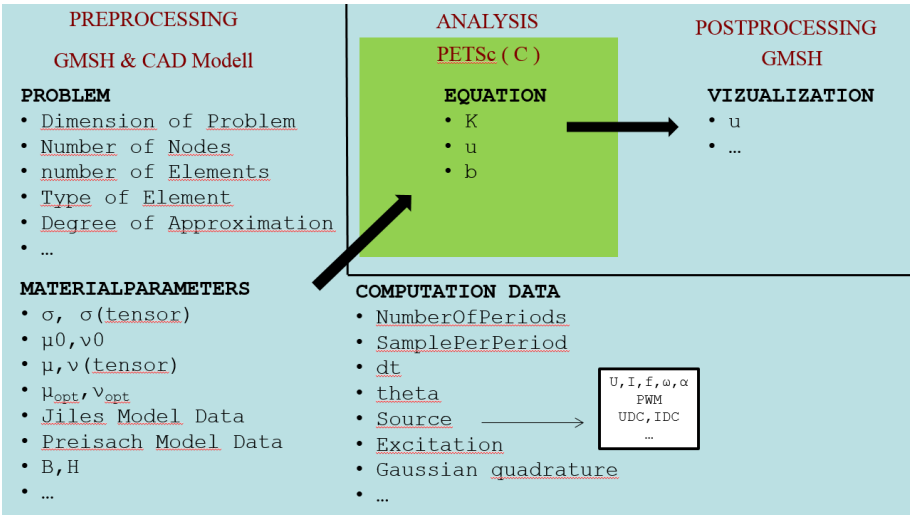


Figure 1. Structure of the development environment

Many data have to be used in the FEM (Finite Element Method) structure. For example number of nodes, number of elements, material parameters, etc., which are given from the geometry and the mesh in the pre-procession step? These parameters and data are used in the Analysis step to solve the equations of partial differential equations. And finally the data of the simulation results are shown in the post-procession step.

Fig. 2. shows the graphical user interface of the finite element based package which consists of two parts.

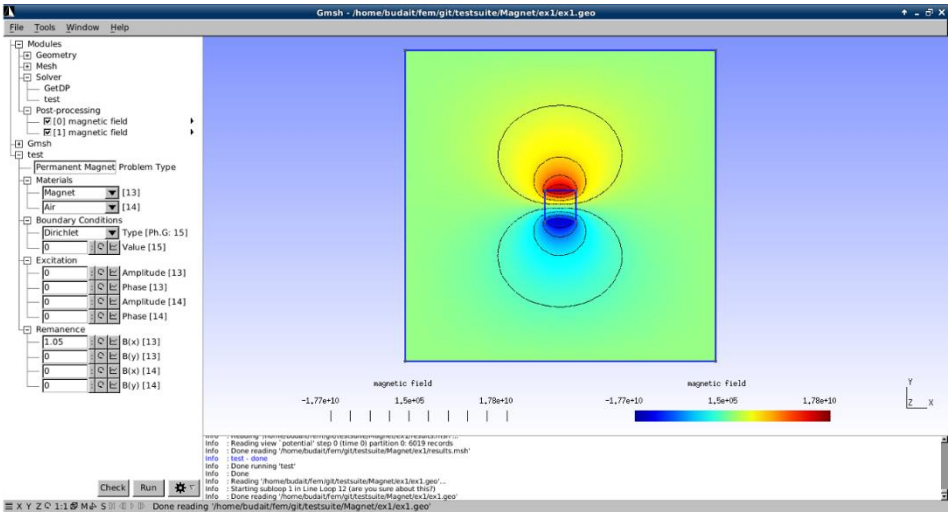


Figure 2. Graphical User Interface

On the left side the mesh, material parameters and boundary conditions can be set. And on the right side the model and the simulation results can be seen which a permanent magnet is.

3. Governing Equations

The basic model consists of a permanent magnet, a ferrite core and an excited single coil which can be seen in Fig. 3.

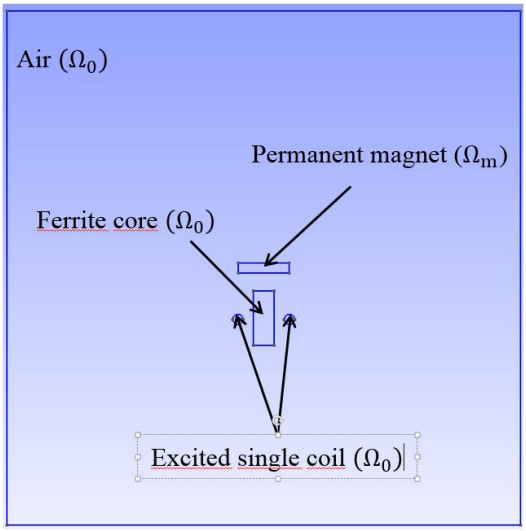


Figure 3. Scheme of the basic model

From the basic model five simple different arrangements had been created. The different simulated arrangements are as follows:

- One turned excited coil;
- One turned excited coil with ferrite core;
- Permanent magnet;
- Permanent magnet with ferrite core;
- The basic model which contains the permanent magnet, the ferrite core and the excited coil, as well.

These models have been calculated in the C programing language based the finite element environment and in MATLAB environment.

The simulated problem has been modelled as a static magnetic field problem, where the following Maxwell's equations can be used [2]-[5]:

$$\nabla \times \boldsymbol{H} = \boldsymbol{J}_0, \text{in } \Omega_0 \cup \Omega_m, \tag{1}$$

$$\nabla \cdot \boldsymbol{B} = 0, \text{in } \Omega_0 \cup \Omega_m. \tag{2}$$

Here $\boldsymbol{H}$ is the magnetic field intensity, $\boldsymbol{J_0}$ is the source current density, $\boldsymbol{B}$ is the magnetic flux density. The $\boldsymbol{H}$ magnetic field intensity can be expressed as

$$\boldsymbol{H} = \begin{cases} \nu_0 \boldsymbol{B}, \text{in } \Omega_0 \\ \nu_0 \nu_\mathrm{r} \boldsymbol{B} \text{ in } \Omega_\mathrm{m} \end{cases}. \tag{3}$$

Here ν_0 is the reluctivity of vacuum and ν_r is the relative reluctivity of magnet. The air region is denoted by Ω_0 and the magnetic region is denoted by Ω_m. The $\boldsymbol{A}$ magnetic flux density can be expressed as

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}, \tag{4}$$

where $\boldsymbol{A}$ is the magnetic vector potential [2]-[5]. This expression satisfies (2), because of the identity $\nabla \cdot \nabla \times \boldsymbol{v}$ for any vector function $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{r})$.

When the domain contains permanent magnets, their magnetic characteristic are given by [3]

$$\boldsymbol{B} = \mu \boldsymbol{H} + \boldsymbol{B}_0, \tag{5}$$

where $\boldsymbol{B}_0$ is the remanent flux density. Substituting (1) and (4) to (5) and using the constitutive relations IS (3) the following partial differential equation can be obtained:

$$\nabla \times \frac{1}{\mu} \nabla \times \boldsymbol{A} - \nabla \times \frac{1}{\mu} \boldsymbol{B}_0 = \boldsymbol{J}_0. \tag{6}$$

The divergence of the magnetic vector potential can be selected according to Coulomb's gauge [5,7],

$$\nabla \cdot \boldsymbol{A} = 0,$$

which is satisfied automatically in two dimensional problems [4], [5]. Using some mathematical identity and using some formulations [4], [5], the following weak equation can be obtained

$$\int_\Omega \frac{1}{\mu} \nabla \times \boldsymbol{W} \cdot \nabla \times \boldsymbol{A} \, \mathrm{d}\Omega - \int_\Omega \frac{1}{\mu} \nabla \times \boldsymbol{W} \cdot \boldsymbol{B}_0 \, \mathrm{d}\Omega = \int_\Omega \boldsymbol{W} \cdot \boldsymbol{J}_0 \, \mathrm{d}\Omega \,, \tag{7}$$

which solution results in the approximation of the magnetic vector potential.

4. Simulation results of the models

Five simple different arrangements had been created from the basic model. These models are calculated in the C programing language based on the finite element environment and in MATLAB environment. The simulation results were compared each other focusing the time and the accurate of the magnetic vector potential. In both environments the unknowns are the same, or closely the same which is 12790.

4.1. Comparison of the computation time of the models

Computation times of two different development environment are compared with each other. The sum up of the computation times of the simulations can be seen in Table 1.

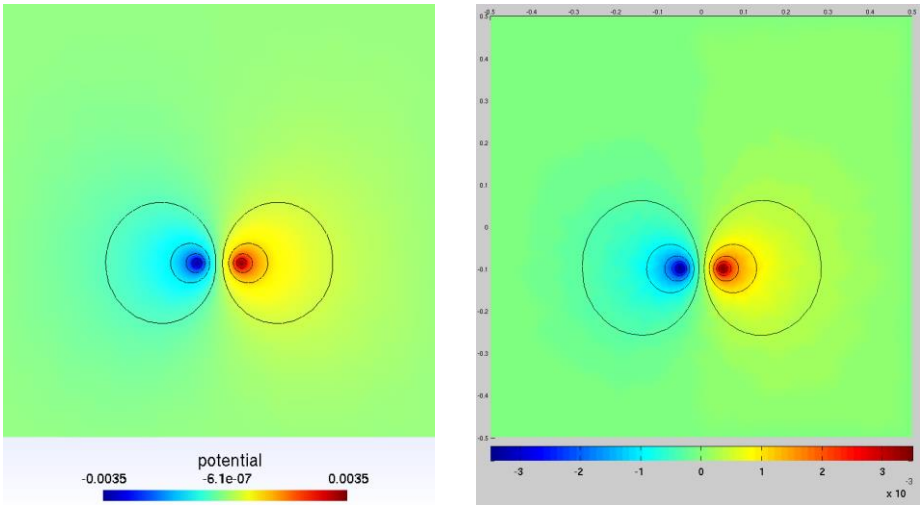
Table 1. Computation times of models in two different environments

model	***MATLAB***	***C***
excited coil	60sec	13sec
ferrite core and excited coil	60sec	13sec
magnet	60sec	12sec
magnet and ferrite core	59sec	10sec
magnet, ferrite core and excited coil	63sec	11sec

On the left side the computation times in MATLAB environment are shown, and on the right side the computation times in C environment are shown in Table 1. In case of every arrangement the problem was calculated almost five times faster by the help of the C environment than with MATLAB.

4.2. Comparison of the magnetic vector potential of the models

Simulation results were also compared with each other focusing on the accuracy of the magnetic vector potential. Fig. 4. shows the simulation result of the magnetic potential in the case of excited one turned coil model. Fig 4a shows the simulation results in GMSH environment. Fig. 4b shows the simulation results in MATLAB environment.

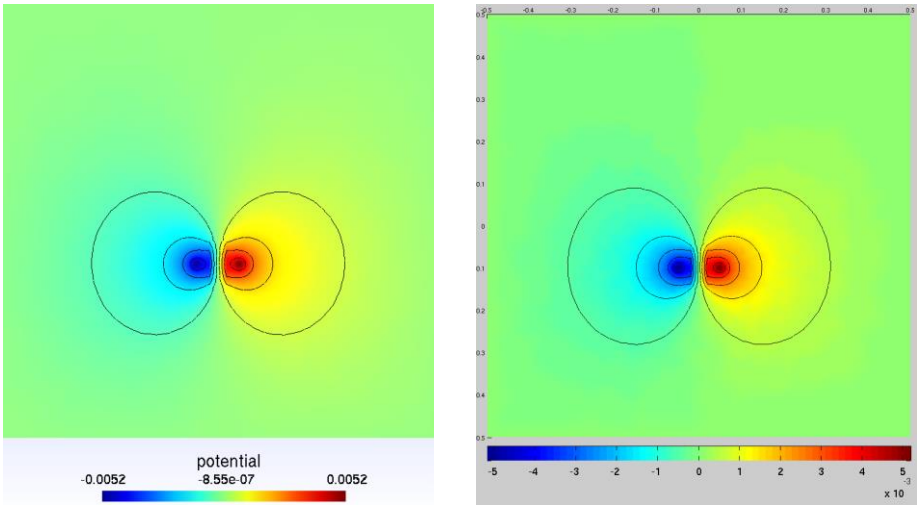


a) *b)*

Figure 4. Simulation results of excited coil

Comparing Fig. 4a and 4b, it can be seen that the magnitude of the magnetic vector potential in the model are similar.

Fig. 5. shows the simulation result of the magnetic potential, where the model consists an excited one turned coil with ferrite core. Fig. 5a shows the simulation results in GMSH environment. Fig. 5b shows the simulation results in MATLAB environment.

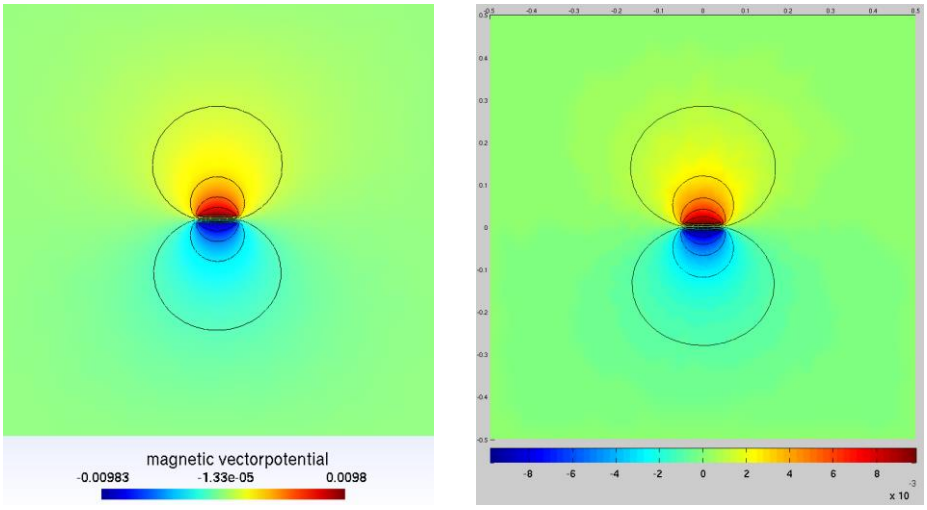


a) *b)*

Figure 5. Simulation results of excited coil with ferrite core

Comparing the Fig. 5a and 5b figures with each other, it can be seen that the magnitude of the magnetic vector potential in the model are similar.

Fig. 6. shows the simulation result of the magnetic potential, where the model consists a permanent magnet. Fig. 6a shows the simulation results in GMSH environment. Fig. 6b shows the simulation results in MATLAB environment.

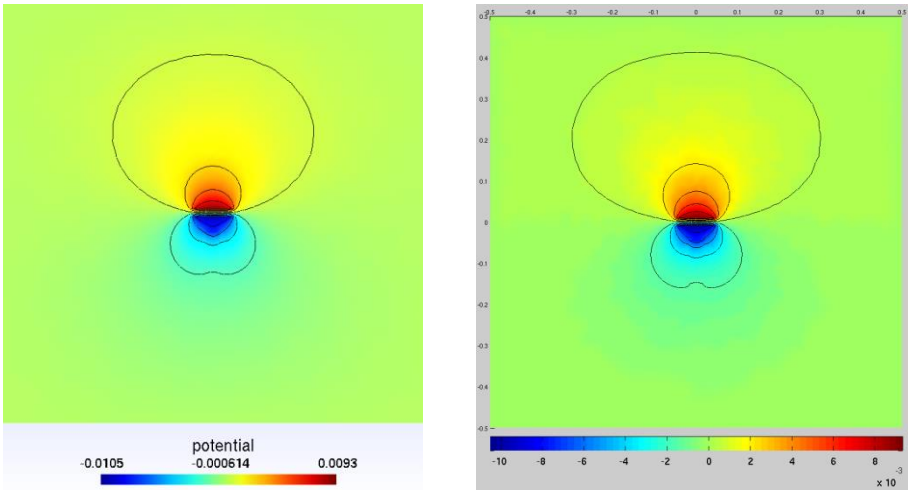


a) *b)*

Figure 6. Simulation results of magnet

Comparing the Fig. 6a and 6b figures with each other, it can be seen that the magnitude of the magnetic vector potential in the model are similar.

Fig. 7. shows the simulation result of the magnetic potential, where the model consists a permanent magnet with ferrite core. Fig. 7a shows the simulation results in GMSH environment. Fig. 7b shows the simulation results in MATLAB environment.

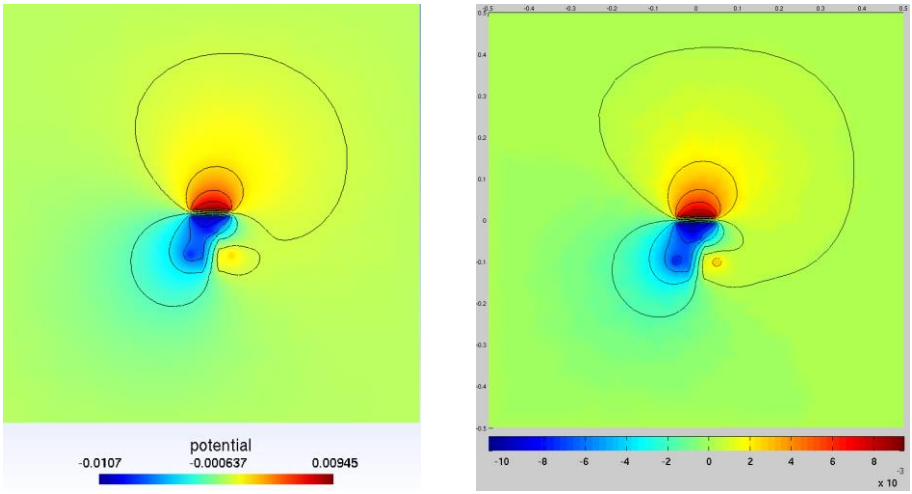


a) *b)*

Figure 7. Simulation results of magnet with ferrite core

Comparing the Fig. 7a and 7b, it can be seen that the magnitude of the magnetic vector potential in the model are similar.

Fig. 8. shows the simulation result of the magnetic potential, where the model consists of a permanent magnet with ferrite core and one turned excited coil. Fig. 8a shows the simulation results in GMSH environment. Fig. 8b shows the simulation results in MATLAB environment.



a) *b)*

Figure 8. Simulation results of magnet, ferrite core and excited coil

Comparing the Fig. 8a and 8b, it can be seen that the magnitude of the magnetic vector potential in the model are similar.

4.3. Differences of the simulation results

Simulation results were also compared with each other focusing on the magnitude of the differences of models.

Fig. 9. shows the differences of the simulation results of the two different development environment in the case of the model of one turned excited coil.

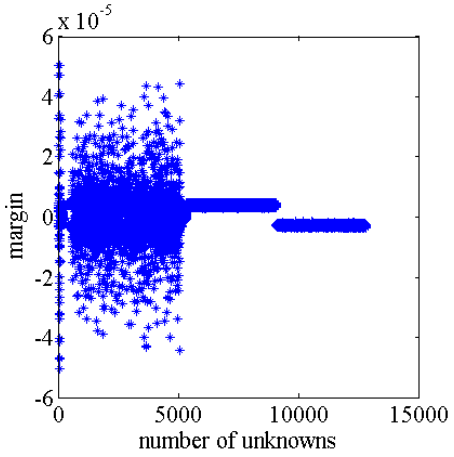


Figure 9. Margin between the simulation results, model: excited coil

The magnitude of the difference of the simulation results is about 10^{-5}.

Fig. 10. shows the differences of the simulation results of the two different development environment in the case of the model of one turned excited coil with ferrite core.

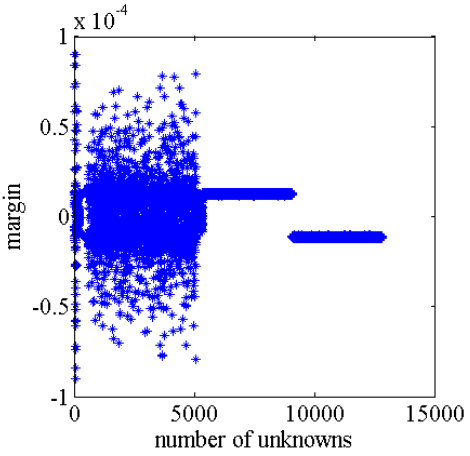


Figure 10. Margin between the simulation results, model: excited coil with ferrite core

The magnitude of the difference of the simulation results is about 10^{-4}.

Fig. 11. shows the differences of the simulation results of the two different development environment in the case of the model of permanent magnet.

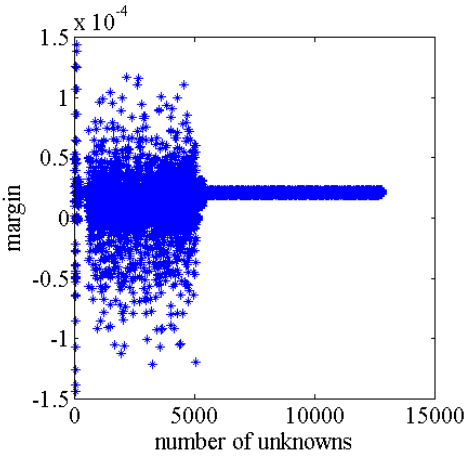


Figure 11. Margin between the simulation results, model: magnet

The magnitude of the difference of the simulation results is about 10^{-4}.

Fig. 12. shows the differences of the simulation results of the two different development environment in the case of the model of permanent magnet with ferrite core.

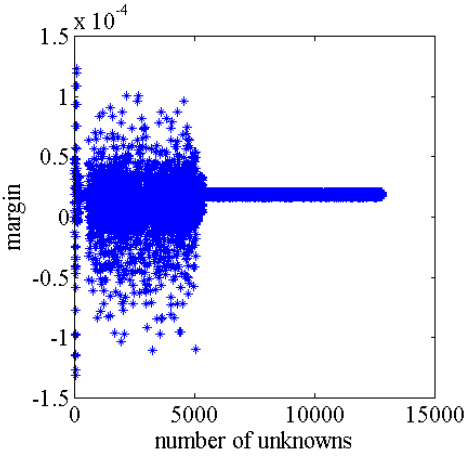


Figure 12. Margin between the simulation results,model: magnet with ferrite core

The magnitude of the difference of the simulation results is about 10^{-4}.

Fig. 13. shows the differences of the simulation results of the two different development environment in the case of the model of permanent magnet with ferrite core and one turned excited coil.

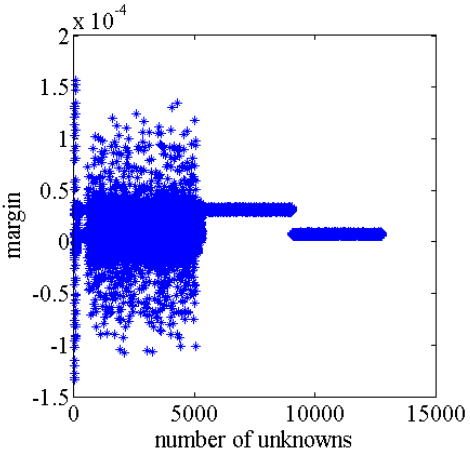


Figure 13. Margin between the simulation results, model: magnet, ferrite core and excited coil

The magnitude of the difference of the simulation results is about 10^{-4}. It can be seen in every case that the differences of the simulation results are very small.

5. Conclusion

The new, finite element package based on C programming language is faster than the MATLAB solver, but the accuracy of the simulation results is adequate.

In the future work this C programming language based finite element package will be used for simulation and optimization of permanent magnet synchronous motors in a more fast and accurate way. This solution will have used in a research work where arrangements of the magnets and air gaps of the rotor of the BLDC motor to develop more energy efficient BLDC motors will have investigated.

Acknowledgement

This research was realized in the frames of TÁMOP 4.2.4. A/2-11-1-2012-0001 „National Excellence Program – Elaborating and operating an inland student and researcher personal support system convergence program" The project was subsidized by the European Union and co-financed by the European Social Fund.

References

[1] Geuzaine, C., Remacle, J.F.: *Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities*, International Journal for Numerical Methods in Engineering, vol. 79, no. 11, pp. 1309-1331, 2009 DOI: 10.1002/nme.2579

[2] Bianchi, N.: *Electrical Machine Analysis Using Finite Elements*, CRC Press, Boca Raton, 2005

[3] Bastos, J.P.A., Sadowski, N.: *Electromagnetic Modeling by Finite Element Methods*, Marcel Dekker Inc., New York, Basel, 2003
[4] Gieras, J.F., Wang, R-J., Kamper, M.J.: *Axial Flux Permanent Magnet Brushless Machines*, (2nd Edition) Springer, 2008
[5] Kuczmann, M., Iványi, A.: *The Finite Element Method in Magnetics*, Akademic Press, Budapest, 2008
[6] M. A. Plonus: *Applied Electromagnetics*, McGraw-Hill, New York, 1978
[7] Bíró, O., Richter, K.R.: *CAD in Electromagnetism*, Advances in Electronics and Electron Physics, vol. 82, no. 1, pp. 1-91, 1991
DOI: 10.1016/S0065-2539(08)60911-7
[8] Kostaridis, A., Soras, C., Makios, V.: *Magnetostatic Analysis of a Brushless DC motor Using a Two-Dimensional Partial Differential Equation Solver*, Computer Applications in Engineering Education, vol. 9, no. 2, pp. 93–100, 2001
DOI: 10.1002/cae.1010
[9] Kumaravelu, U.D., Yakub, S.M.: *Simulation of outer rotor permanent magnet brushless DC motor using finite element method for torque improvement*, Journal Modelling and Simulation in Engineering, vol. 2012, article ID. 17, 6 pages, 2012
DOI: 10.1155/2012/961212
[10] Caihong, Z., Hongtao, Z.: *Magnet Field Finite Element Analysis of Permanent Magnet Brushless DC Motors based on ANSOFT*, Advanced Materials Research, vol. 904, pp. 504-507, 2014
DOI: 10.4028/www.scientific.net/AMR.904.504

Acta
Technica
Jaurinensis

Vol. 7., No.2., pp. 183-192, 2014
DOI: 10.14513/actatechjaur.v7.n2.282
Available online at acta.sze.hu

Microsoft Change Management Applying Comparison of Different Versions

I. Orosz, T. Orosz

Óbuda University, Székesfehérvár, Hungary
e-mail:orosz.istvan@arek.uni-obuda.hu, orosz.tamas@arek.uni-obuda.hu

Abstract: Microsoft Dynamics AX continuously extends functional and technical approaches. Former Axapta and Navision applications are being replaced with the modern and smart AX and NAV solutions, like the other Dynamics modules and components. A brand new MS Dynamics implementation follows standard methodologies of Sure Step. However, companies, where one can find former Microsoft ERP systems, like Axapta or Navision, face to several change management issues and problems. One can figure out that business requirements for Software change management depend on specific business functions, industry, technologies and so on. This paper introduces an agile way for business change management in Microsoft Dynamics environment. A comparison of different business requirements is made in different MS Dynamics business change issues.

Keywords: Microsoft Dynamics AX, change management, Navision

1. Introduction

Microsoft widely used in either in several core business areas, like logistics, financials and human resources or in the industry, such as automotive, pharmaceuticals, etc. Business deploying Dynamics to cover required functionalities needs continuous improvement and changes in functional and development framework of MS Dynamics System.

Dynamics AX [1-2], provides full solutions for small, middle and larger sizes of organizations. Multiple examples can be found either for replacing suddenly the entire IT System or for changing it slowly and gradually. However, we cannot either repair business processes which should be replaced sooner or later, nor replace them immediately due to sustainability reasons.

Dynamics AX, one of the main ERP software in the World was announced in 1998 as IBM Dynamics AX, it was developed by IBM and Damgaard. The developers later merged with Navision Software A/S, the combined company was annexed by Microsoft in 2002 [5]. After Microsoft handled over the product, it becomes a real enterprise resource planning system. The AX modifications is done via using MorphX, X++, AOS, AOT and other technologies. The AX 2012 version uses a new technique for custom modifications; Visual Studio 2010 has to be installed for developing clients.

The Dynamics AX software consists of four main parts [3-4]:

- Database Server, a database that stores the Dynamics AX data. This server is a Microsoft SQL server usually.
- File Server, a folder containing the Dynamics AX application files.
- Application Object Server(s) (AOS), a service that controls all aspects of Microsoft Dynamics AX's operation. There can be more AOS in case of scalability.
- Client(s), the actual user interface [13] into Microsoft Dynamics AX. Some of the business logic is programmed in clients, and some are implemented in the Application Object Server.
- The core paradigm of AX change management is innovation and sustainability, so the main focus of this paper is how these two aspects are present.
- Microsoft Dynamics AX change management

The implementation of Microsoft Dynamics AX 2012 according to Microsoft paradigm of change management is somehow different from the traditional aspects. This method describes the transition of an organization from a previous state to a planned state. This process is a managed one, which leads the whole organization along the change. In project management, this refers to a project management process, where the changes form a project and are officially introduced and approved [6]. The change management usually comes with organizational developments as well. The purpose of change management is to minimize the negative impacts on the organization and to avoid concerns. There are some different types of change management: changing the behavior of the personnel; technological changes; operational and structural changes; strategic and mission changes. The changes, which affect Dynamics AX, come from operational, structural and technological sides. The organizational change management should start with detailed description of the current situation [14], and after that focusing on the need of change and the ability of change. The objectives, content, and the process all should be part of the organizational change management plan. It uses the following techniques: performance metrics such as financial results, operational efficiency, leadership commitment, communication effectiveness. The perceived need for change to plan the appropriate strategies, in order to avoid change failures or solve troubled change projects.

Organizational change management is closer to a successful project ending, if these points are contained in the project:

1. The management of proceeds, the realization to define measurable participant aims, creating a business case for their achievement (which should be continuously updated), and monitor assumptions, risks, dependencies, costs, return on investment, disadvantages and cultural issues affecting the progress of the associated work.
2. Effective Communications that informs various stakeholders of the reasons for the change (why?), the benefits of successful implementation (what is in it for us, and you) as well as the details of the change (when? where? who is involved? how much will it cost? etc.).

3. Devise an effective education, training and/or skills upgrading scheme for the organization.
4. Counter resistance from the employees of companies and align them to overall strategic direction of the organization.
5. Provide personal coaching (if needed) to alleviate any change related fears.
6. Monitoring of the implementation and fine-tuning as required.[8]

It is also important to make a clear governance and organizational model of the company, which is suited and aligned to the expectations of the company. The consultation model has to take a closer look at the organisational culture of the company, the principles and the values along which the company is driven.

The future structure of the organization has to be designed with considering the following fields: segmentation, autonomy, integration. The flow of the changes which affects the technological side of Dynamics AX changes is shown on Fig.1.

In order to overcome the needs of BPR [10-12] which remarkably require a long process, Dynamics AX has developed an implementation methodology to speed-up the whole process by instantly implementing principals which has been underlined by Dynamics AX business process recipes.

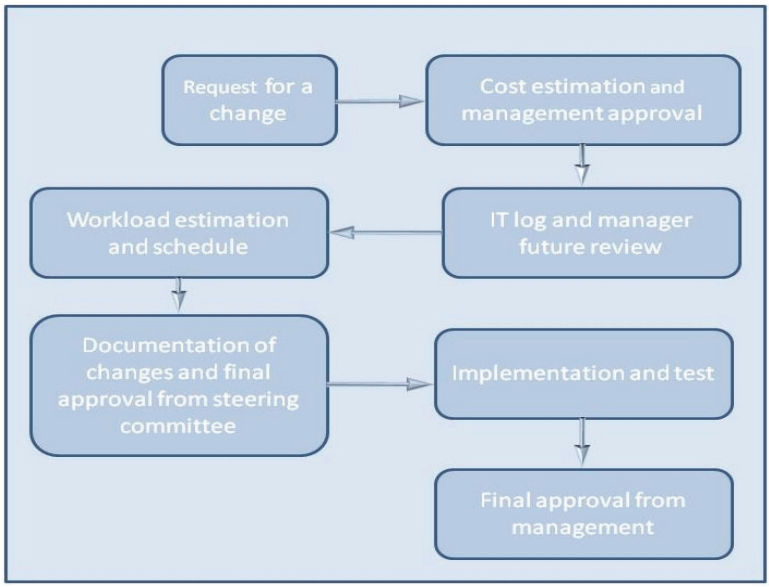


Figure 1. Flow of technological changes

Sustainability is one of the core factors of developing Dynamics AX 2012. One of the key aspects was to speed up the business and technological process of upgrading from the previous versions to AX 2012. Sure Step is the official software process for Dynamics AX. This is the tool of defining process phases, milestone roles, artifacts, cross phase processes, additional project management processes. Supports a wide range of software like Dynamics AX, NAV, GP, SL and CRM. The project types, which are

supported, are the following: full and rapid implementation, optimization and upgrade. The core steps of these implementations are: diagnostics, analysis, design, development, deployment, operation. Optimization and upgrade are different from these implementations in the way that an existing software environment has to be modified.

The process guide is an html solution combined with some ActiveX controller, which allows tailoring the process needs, as shown in Fig. 2.

Sure Step describes six main and phases and 2 additional phases for optimization and upgrade. It covers the complete lifecycle from project initiation, development, deployment, optimization and upgrade to the next version, where every phase ends at a milestone. A milestone is a sum of artifacts created or refined in the phase. If a milestone is executed with success, then the next step can be started. At the first look the phases look like a waterfall model, but naturally Sure Step isn't all about waterfall.

Figure 2. Microsoft Sure Step Editor

These phases are the following:

- Diagnostic: the diagnostic phase contains analysis of the customer process at a very high level. Focus of the diagnostic phase is the project initialization; setup a project plan, agree on an approach and scope definition.
- Analysis: in the analysis phase most of the business processes are identified and documented at a high level. If necessary an external specialist helps in executing this step. The goal of the analysis phase is to understand the customers business and processes performed. Modeling and documenting the customers business is important. Microsoft Provides a new tool for the modeling: Microsoft Sure Step Business Modeler
- Design: main purpose of the design phase is to find a way how the customers' processes and needs can be implemented with Dynamics. At this point there may be more than one solution for a topic. Need to identify the best strategy for the implementation phase. Although prototyping is not covered in sure step, this is the point where to build prototypes and do load testing on the prototype implementation.

- Development: most of the programming work is done in the development. Development covers the creation of new features and adaption of existing features as well as the data migration. All features and the data migration have to be tested. Beside feature and data migration testing it is necessary to do security testing. The phase ends when most of the specified features are developed and tested and the data migration is done.
- Deployment: goal of the deployment phase is to setup the operational dynamics environment at the customer. Beside installation, configuration and feature deployment, the phase focuses on testing at system level. So the specialists have to run the user acceptance tests, process tests, security tests and load tests as defined in the test plan. When the system runs and all tests pass the deployment phase is finished.
- Operation: this is the place for work in live environment.

Optimization and upgrade: these two often go together. The optimization steps usually lead to a new version upgrade.

2. Dynamics AX innovation and sustainability

There are some situations, where the original waterfall concept of Sure Step is not suitable. Waterfall is the classic implementation model staging from Analysis, to Design, to Deployment, and finally Operation it works for MS Dynamics Sure Step project types (Enterprise, Standard, Rapid) where those project types have different customization levels except Rapid, which has no or minimal customization, as it presents and out-of-the-box approach.

Typically the waterfall is a good method, shown at Fig. 3., when the duration of the project is shorter than the frequency of organizational change. In other words, if the organizational change is prevalent, and the project is longer than this period, then the project might fail, because the requirements might change when the project is running.

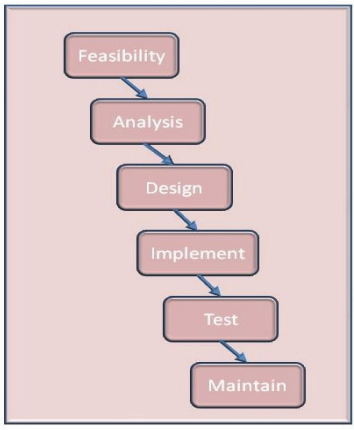


Figure 3. Waterfall method

Other issue of waterfall method is that project step has to be done in strict order, and there is no possibility to move back to the previous stage. This means some lack of innovation of the project, but with more focus on sustainability. Because all the requirements are analysed at the beginning, if any change occur in the needs, the whole process will be affected. At the end, all the development is tested and deployed at once, in a big step. The biggest risk of the waterfall method is that if there is a misunderstanding in the requirements at the beginning, there is no way to step back while the project runs, and this affects the project ending as well. These events can make the project more expensive, and the fix of the possible mistakes can take a long period of time. Typically one can say that waterfall method provides a structure, which suits best to traditional type of projects with little innovation. Projects, which need new improvements in technology, maybe not the best subject to a waterfall method.

3. Agile way for business change management in Dynamics AX

There is a different new way of implementing Dynamics AX, which is called Agile Implementation. Agile method is totally different from waterfall method. Agile implementation model is an iterative, incremental process for developing Microsoft Dynamics Solutions. This Project Type gives customers greater control over the final solution because they can quickly change the direction of solution development and implementation from one sprint cycle to the next. It focuses delivering the whole functionality in smaller chunks, within a series of smaller developing series. By delivering and accepting smaller parts of development, the risk of the customer will not get the solution needed is smaller. [9] If the developed solution does not fit into the needs of the customer, it can be easily and quickly fixed in the next chunk, which means that customers get involved into the project deeper. This also means, that customers get a deeper understanding of the project, resulting a higher quality of the product, and often resulting lower costs. The drawback of agile implementation is that it is useful only with strict material and time agreement.

Typically used at a single site requiring specific features and moderate to complex customizations [3]. The development phase in the agile execution operates in Scrum development model, life cycle is divided to 30 days sprint cycle and daily sprint cycle contains analysis, design, coding implementation and end up with solution test, and finalizing product specification.

The short description of the process is the following, Fig. 5.:

- Solution Backlog: the solution feature list is listed here.
- Release Backlog: identifying and prioritizing the set of features which are to be developed during the 30 day sprint, and determines all of the feature time estimation (story point).
- Sprint Backlog: it is the breakdown of release backlog compared to feature priority and estimation it could be 3-days to 30 days (daily sprint)
- User story: this is the place for descript feature business function, users (roles), and test script (less documentation).

- Defect Backlog: identifying and reporting system bugs.
- Stand-up daily meeting: what has been done today (status).

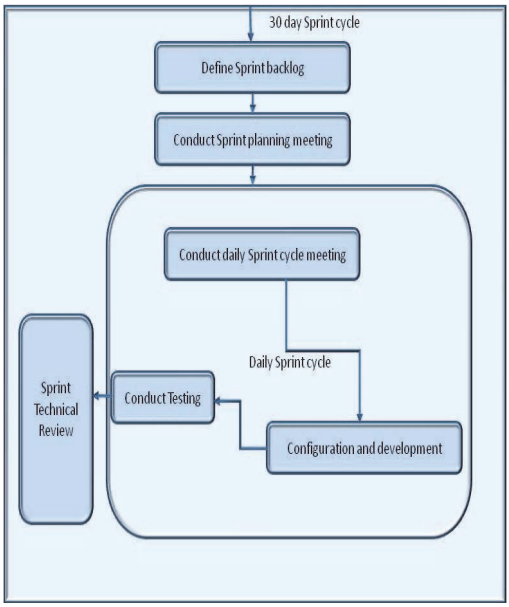


Figure 4. Sprint cycle

The agile implementation starts with a detailed business process analysis, and the declaration of high level requirements with fit-gap analysis. The output of the fit-gap analysis will become a main document of the project, it is called Solution Backlog. The Solution Backlog is a living and changing document, tracking the current business and project priorities along the project. Until this stage, the agile project looks like a usual project type.

At the stage of project execution the agile implementation is totally different. The traditional project is divided into two stages, Design and Development, the agile project is doing 30 day sprint cycles, shown in Fig. 5.

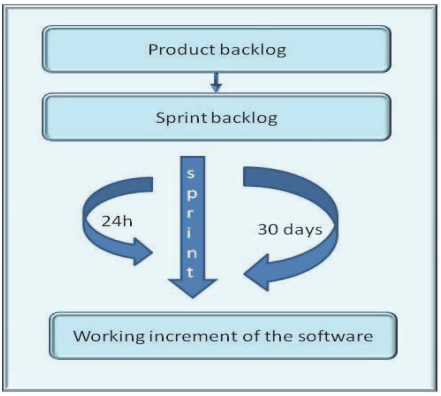


Figure 5. 24h and 30days sprint cycles

This 30 days sprint cycle contains a small part of the Solution backlog, which is called Sprint Backlog. Each requirement is split into small parts, no longer than 16 hours of development time, and connected to developers. During this every cycle, the project team does daily sprint meetings.

The goal of these meetings are the following: each team member shares what has been achieved since last meeting, what will be done until the next meeting and what issues affects the project.

The functionalities, which have been developed, are synchronized and built into daily builds. As soon as a requirement is reached, the sprint testing will be executed. If any change is needed after testing, this will be the task of next daily sprint cycle.

At the end of the sprint cycle, the Sprint Technical Preview is done, where the customer reviews, and signs the developed functionality. After the next sprint cycle is started, and iterated as many times as necessary, until the Solution Backlog is empty. At this point, the developer team does a detailed Solution Testing, to verify all the requirements are satisfied, and the whole solution is working perfectly according to the expectations. After the Solution Testing, the project resembles again to the standard project types, as the deployment status arrives and the whole solution goes into live.

4. Summary

An overview of Dynamics AX introduction methodology changes has been given in the previous sections. It is clear that speed of changing development and functional approaches depends on business. Therefore it is hard to estimate the success of business changes. A categorization of MS Dynamics change management is being built for core applications. Business issues and corresponding tasks required are registered and tracked for such cases. This tool will help decision-makers to make realistic estimations for necessary business changes.

It is clear now that the Agile Implementation opens a good framework for developments in organizations which requirements are volatile and fuzzy. The solution development is incremental, but the deployment is still done in one big step. There some

reasons, why Agile Implementation does this, but the strongest is the ERP data consistency.

Further questions will have to be answered, for example: why would the customers choose agile implementation over the well known waterfall methods? Are there enough benefits in using agile implementations over the waterfall methods, which has the advantage of the (almost) fix price and time? These issues have to be carefully answered before choosing implementation method. Despite of these questions, Agile Implementation method is completely different approach, which needs totally new way of thinking from the development team members. This implementation method has surely more chance to deliver a solution which suits the needs of the customer better.

References

[1] Gupta, A. K.: *Quality Assurance for Dynamics AX-Based ERP Solutions: Verifying Dynamics AX customization to the Microsoft IBI Standards*, Packt Publishing 2008 ISBN-10: 1847192912

[2] Pocius, M.: *Microsoft Dynamics Ax 2009 Development Cookbook,* 2009, ISBN: 1847199429

[3] Hamilton, S.: *Managing Lean Manufacturing using Microsoft Dynamics AX 2009*, 2009
ISBN-10: 097925521X

[4] Koop, R., Muris, E.: *Successfully Implementing Microsoft Dynamics(TM): By Using the Regatta® Approach for Microsoft Dynamics (TM),* Springer 2007
ISBN-10: 3540715924

[5] Andreasen, S.: *MORPHX IT, An introduction to Axapta X++ and the MorphX Development Suite*, 2007
ASIN: B001CJQU6G

[6] Filicetti, J. *PMO and Project Management Dictionary,* PM Hut. (August 20, 2007): Retrieved 16 November 2009
http://www.pmhut.com/pmo-and-project-management-d

[7] Greef, A., Pontoppidan, M. F., Olsen, S. D.: *Inside Microsoft Dynamics AX 4.0. Microsoft,* Microsoft Press E-Book, 2006
ISBN-10: 0-7356-2257-4

[8] Phillips, J. R.: *Enhancing the Effectiveness of Organizational Change Management,* Human Resource Management, vol. 22(iss. 1/2), pp. 183-199, 1983
Retrieved 12/21/11 from http://onlinelibrary.wiley.com
DOI: 10.1002/hrm.3930220125

[9] Shankar, C., Bellefroid, V.: *Microsoft Dynamics Sure Step 2010,* 2011
ISBN-10: 1849681104

[10] Muka L., Lencse, G.: *Developing a meta-methodology for efficient simulation of infocommunication systems and related processes,* Infocommunications Journal, vol. LXIII, no. 7. pp. 9-14, 2008

[11] Muka, L., Lencse, G.: *Cooperating Modelling Methods for Performance Evaluation of Interconnected Infocommunication and Business Process Systems,* Proceedings of the 2008 European Simulation and Modelling Conference, 2008

[12] Muka, L., Benkő, B. K.: *Meta-level performance management of simulation: The problemcontext retrieval approach,* Periodica Polytechnica, vol. 55, no. 1-2, pp. 53-64, 2011
DOI: 10.3311/pp.ee.2011-1-2.06

[13] Mátrai, R., Kosztyán, Zs. T.: *A New Method for the Characterization of the Perspicuity of User Interfaces,* Acta Polytechnica Hungarica, vol. 9, no. 1, pp. 139-156, 2012

[14] Ősz, R.: *Educational organization for new generation,* SAMI 2012, IEEE 10th International Symposium on Applied Machine Intelligence and Informatics, Herl'any (Szlovákia) Conference Proceedings, pp. 373-375, 2012
DOI: 10.1109/SAMI.2012.6208992

Acta
Technica
Jaurinensis

Vol. 7., No.2., pp. 193-206, 2014
DOI: 10.14513/actatechjaur.v7.n2.278
Available online at acta.sze.hu

Domain Decomposition Algorithms for Edge Element Based Parabolic Type Problems

D. Marcsa[1,2]

[1]Széchenyi István University, Department of Automation
Egyetem tér 1., H-9026, Győr, Hungary
E-mail: marcsad@sze.hu

[2]Research Center of Vehicle Industry
Egyetem tér 1., H-9026, Győr, Hungary

Abstract: Today, a huge amount of interest in new actuators and a correspondingly huge pressure on companies and research centres to develop the most efficient, cost-effective electric design of the actuators for electric vehicles. However, the design and analysis of actuators is very complex task, thus need the help of multiprocessor computers to run efficient computations. This work analyzes the use of two non-overlapping domain decomposition methods (DDMs) in order to improve the calculation behaviour of finite element method (FEM) with edge element approximation. In this case, the DDMs under investigation are the Schur complement method and the Lagrange multiplier based Finite Element Tearing and Interconnecting (FETI) method with their algorithms. The performances of these methods and solvers have been investigated in detail for two-dimensional parabolic type problems as case studies.

Keywords: Domain decomposition, Iterative algorithm, Finite element method, Edge element based parabolic type problem

1. Introduction

The numerical design of electromechanical actuators of electric vehicles is a very complex task, because a lot of different physical aspects should be considered. The performances of electrical equipments are not defined only by their electromagnetic field, because the electromagnetic field has strong interaction between the following quantities: electromagnetic field distribution, mechanical equation, external circuits, etc. But, the analysis of complex system, e.g. analysis of motors of electric vehicle is very resource-intensive and time consuming, wherein the resources and time of the simulation plays an important role for designers and researchers. Therefore, the solution of a complex system must be parallelised in order to speedup the numerical computations with less computer requirement. The parallelisation of computation can be approached with the use of a domain decomposition method (DDM) [1-8]. To solve

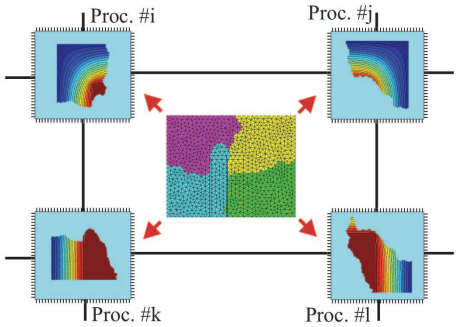


Figure 1. Demonstration of mesh partitioning and distributed computation.

large scale problems, a domain has been divided into sub-domains that are fit into the computer memory. While limited progress can be reached with improvement of numerical algorithms, a radical time reduction can be made with multiprocessor computation. In order to perform numerical design a computer with parallel processors, computations should be distributed across processors (see in fig. 1). The used DDMs are the Schur complement method and the Finite Element Tearing and Interconnecting (FETI) method.

The Schur complement method [1-3], as sequential algorithm was started to use many decades ago, when computer RAM was extremely small. Nevertheless, nowadays, this method is a very popular parallel domain decomposition technique among engineers [3].

In the last decade, the Finite Element Tearing and Interconnecting (FETI) method [4-6] has seemed as one of the most powerful and one of the most popular solvers for numerical computation. The FETI requires fewer interprocess communication, than the Schur complement method, while is still offers the same amount of parallelism [1].

The parallel finite element based numerical analysis on massively parallel computers or on clusters of Personal Computers (PCs) needs the efficient partitioning [5] of the finite element mesh. This is the first and the most important step of parallelization with the use of domain decomposition methods.

Many domain decomposition or graph-partitioning algorithms can be found in the literature. Gmsh [7] combined with METIS [8] algorithm has been used for the discretization of the domain of problem and for the mesh partitioning.

The finite element method [9, 10] is an important technique for the solution of a wide range of problems in science and engineering. Generally use vector and scalar potential functions to describe the field quantities. Finite element techniques using nodal based functions to approximate both scalar and vector potentials were first to emerge [10]. However, if the vector potential has two (in 2D) or three components (in 3D), the uniqueness of the vector potential is not so evident and it can be prescribed by implicit enforcement of Coulomb gauge [9, 10]. This problem is avoided if the vector potentials are approximated by edge finite elements and taking care about the representation of source current density [9, 10].

This paper presents an edge element based parallel approach for the solution of two-dimensional parabolic type problems by finite element method. The problems are benchmarks to show the steps of the edge element based finite element technique with DDMs. The comparison focused on the speedup and iteration of solvers of methods.

2. Problems Definition

The paper presents the steps of parallel finite element method through two 2D parabolic type problems. The studied parabolic type problems are separated into two parts: the conducting region (laminated core, steel plate) denoted by Ω_c, and the nonmagnetic and nonconducting domain (air, windings) denoted by Ω_n. The Γ_B boundary denotes that the normal component of the magnetic flux density is vanishing. The Γ_H boundary denotes that the tangential component of the magnetic field intensity is zero or a known surface current density $\boldsymbol{K}$ [9, 10].

The first problem is the quarter of a single-phase transformer (see in fig. 2), where the problem domain Ω is split into conducting Ω_c and nonconducting Ω_n part. The detailed description about the geometry of this problem you can find in [11]. The partial differential equations of this problem are [9, 10]

$$\left.\begin{aligned} \nabla \times \nabla \times \boldsymbol{T}_0 &= \nabla \times \boldsymbol{J}_0 \quad \text{in } \Omega \\ \boldsymbol{T_0} \times \boldsymbol{n} &= \boldsymbol{0} \quad \text{on } \Gamma_H \end{aligned}\right\} \rightarrow \mathbf{K}_E \mathbf{a}_E = \mathbf{b}_E, \tag{1}$$

$$\left.\begin{aligned} \nabla \times \left(\frac{1}{\mu_0} \nabla \times \boldsymbol{A}\right) &= \nabla \times \mathrm{T}_0, \quad \text{in } \Omega_n, \\ \nabla \times \left(\frac{1}{\mu} \nabla \times \boldsymbol{A}\right) + \sigma \frac{\partial \boldsymbol{A}}{\partial t} &= 0, \quad \text{in } \Omega_c, \\ \boldsymbol{A} \cdot \boldsymbol{n} &= 0, \quad \text{on } \Gamma_{B,} \end{aligned}\right\} \rightarrow \mathbf{K}_N \mathbf{a}_N = \mathbf{b}_N, \tag{2}$$

where $\boldsymbol{J}_0$ = 63902 A/m^2 is the source current density, $\boldsymbol{T}_0$ is the impressed current vector potential, $\boldsymbol{A}$ is the magnetic vector potential, $\mu = \mu_0 \mu_r$ is the permeability, $\mu_0 = 4\pi 10^{-7} \mathrm{H/m}$ is the permeability of vacuum, $\mu_r = 4500$ is the relative permeability,

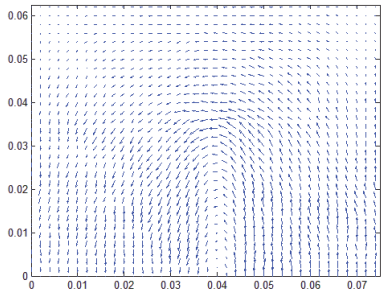


Figure 2. The solution of $\boldsymbol{T}_0$ impressed current vector potential at 10Hz.

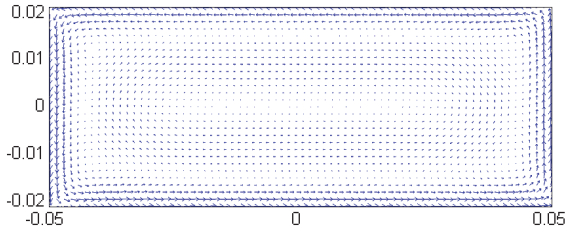


Figure 3. The solution of (3) when the frequency is 5Hz.

$\sigma = 1780$ S/m is the conductivity, and $\boldsymbol{n}$ is the outer normal unit vector. The subscript E and N denotes, the linear system of equation ($\mathbf{Ka} = \mathbf{b}$) from edge or nodal element approximation.

At this problem, the impressed current vector potential the $\boldsymbol{T}_0$ computation has been performed by edge finite element approximation, and the magnetic vector potential $\boldsymbol{A}$ has been approximated by nodal shape function, because it has only z-component, i.e. A_z. The (2) problem has been used to validate the results of parallel edge element based finite element technique.

The second problem is a steel plate (0.1 x 0.04 m) around a coil (see in fig. 3). In this problem, the conductivity and relative permeability are $\sigma = 2 \cdot 10^5$ S/m and $\mu_\mathrm{r} = 4000$. This problem has only eddy current region Ω_c, because the excitation is the surface current density $\boldsymbol{K}$ = 200 A/m as boundary condition with the appropriate direction. The partial differential equation and boundary condition of the steel plate problem are the following [9, 10],

$$\left.\begin{array}{r} \nabla \times \left(\frac{1}{\mu}\nabla \times \boldsymbol{A}\right) + \sigma \frac{\partial \boldsymbol{A}}{\partial \mathrm{t}} = 0, \quad \text{in } \Omega, \\ \boldsymbol{T_0} \times \boldsymbol{n} = \boldsymbol{K} \quad \text{on } \Gamma_\mathrm{H} \end{array}\right\} \rightarrow \mathbf{K}_\mathrm{E}\mathbf{a}_\mathrm{E} = \mathbf{b}_\mathrm{E}, \tag{3}$$

The above mentioned problems have been solved by finite element method, i.e. the $\Omega = \Omega_\mathrm{c} \cup \Omega_\mathrm{n}$ region has been discretized by FEM mesh. The weighted residual method with the Galerkin formulation has been applied to give the weak formulations of the (1), (2) and (3).

The weak formulation of the (1) and (2) are [9, 10]

$$\int_\Omega (\nabla \times \boldsymbol{W}_k) \cdot (\nabla \times \boldsymbol{T}_0^\kappa)\mathrm{d}\Omega = \int_{\Omega_\mathrm{n}} (\nabla \times \boldsymbol{W}_k) \cdot \boldsymbol{J}_0 \mathrm{d}\Omega, \tag{4}$$

and

$$\int_{\Omega_\mathrm{c} \cup \Omega_\mathrm{n}} \frac{1}{\mu}(\nabla N_k) \cdot (\nabla A_z^\kappa)\mathrm{d}\Omega + \int_{\Omega_\mathrm{c}} N_k \cdot \sigma \frac{\partial A_z^\kappa}{\partial t}\mathrm{d}\Omega = \int_{\Omega_\mathrm{n}} (\nabla \times \boldsymbol{W}_\mathrm{k}) \cdot \boldsymbol{T}_0^\kappa \mathrm{d}\Omega, \tag{5}$$

where $\boldsymbol{W}_k$ is the vector weighting function of the k^{th} element, N_k is the nodal weighting function of the k^{th} element. $\boldsymbol{T}_0^\kappa$, A_z^κ denote the approximated unknown potential functions, e.g. [9]

$$\begin{aligned} \boldsymbol{T}_0 &\cong \boldsymbol{T}_0^{\kappa} = \boldsymbol{T}_{0D} + \sum_{k=1}^{n} \boldsymbol{W}_k T_{0k}, \\ \mathrm{A}_z &\cong \mathrm{A}_z^{\kappa} = \mathrm{A}_{zD} + \sum_{k=1}^{n} N_k A_{zk}\,, \end{aligned} \tag{6}$$

where $\boldsymbol{T}_{0D}$ and A_{zD} prescribe the Dirichlet boundary condition.

The weak formulation of (3) is [9, 10]

$$\int_{\Omega_c} \frac{1}{\mu} (\nabla \times \boldsymbol{W}_k) \cdot (\nabla \times \mathbf{A}^{\kappa}) \mathrm{d}\Omega + \int_{\Omega_c} \boldsymbol{W}_k \cdot \sigma \frac{\partial \boldsymbol{A}^{\kappa}}{\partial t} \mathrm{d}\Omega = \int_{\Gamma_H} \boldsymbol{W}_k \cdot \boldsymbol{K}\, \mathrm{d}\Gamma. \tag{7}$$

3. Parallel Finite Element Method with Domain Decomposition

The main idea of domain decomposition method is to divide the domain Ω into several sub-domains in which the unknown potentials can be calculated simultaneously, i.e. in a parallel.

The general form of a linear algebraic problem arising from the discretization of a parabolic type problems defined on the domain Ω can be written as [12]

$$\mathbf{K}\mathbf{a} = \mathbf{b}, \tag{8}$$

where $\mathbf{K} \in R^{n \times n}$ is a positive definite mass matrix, $\mathbf{b} \in R^n$ on the right hand side of the equations represents the excitations, and $\mathbf{a} \in R^n$ contains the unknown potentials. Here n is a number of unknowns.

In the following, we introduced the extensive review of the Schur complement method and the FETI method for the parallel solution of edge element based problems. However, the presented techniques also useful for the solution of nodal element based problems.

3.1. Schur Complement Method

After the problem is partitioned into a set of N_S disconnected sub-domains, it can be seen in fig. 4, and equation (8) has been split into N_S particular blocks [1-3]

$$\begin{bmatrix} \mathbf{K}_{jj} & \mathbf{K}_{j\Gamma_j} \\ \mathbf{K}_{\Gamma_j j} & \mathbf{K}_{\Gamma_j \Gamma_j} \end{bmatrix} \begin{bmatrix} \mathbf{a}_j \\ \mathbf{a}_{\Gamma_j} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_j \\ \mathbf{b}_{\Gamma_j} \end{bmatrix}, \tag{9}$$

where $j = 1 \ldots N_S$, $\mathbf{K}_{jj}$ is the positive definite sub-matrix of the j^{th} sub-domain, $\mathbf{a}_j$ is the vector of the right hand side defined inside the sub-domain. The sub-matrix $\mathbf{K}_{j\Gamma_j} = \mathbf{K}_{\Gamma_j j}^{\mathrm{T}}$ contains the value of j^{th} sub-domain, which connect to the interface boundary unknowns of that region. The superscript T denotes the transpose. $\mathbf{K}_{\Gamma_j \Gamma_j}$, and $\mathbf{a}_{\Gamma_j}$ expresses the coupling of the interface unknowns.

Each sub-domain will be allocated to an independent processor core, because the sub-matrix $\mathbf{K}_{jj}$ with the $\mathbf{K}_{j\Gamma_j}$, $\mathbf{K}_{\Gamma_j j}$ and the right-hand side $\mathbf{b}_j$ are independent, i.e. they can be assembled in parallel on distributed memory. Only the $\mathbf{K}_{\Gamma_j \Gamma_j}$, and $\mathbf{b}_{\Gamma_j}$ are not independent. The matrix $\mathbf{K}_{\Gamma\Gamma}$ and the vector $\mathbf{b}_{\Gamma}$ are assembled after interprocess data

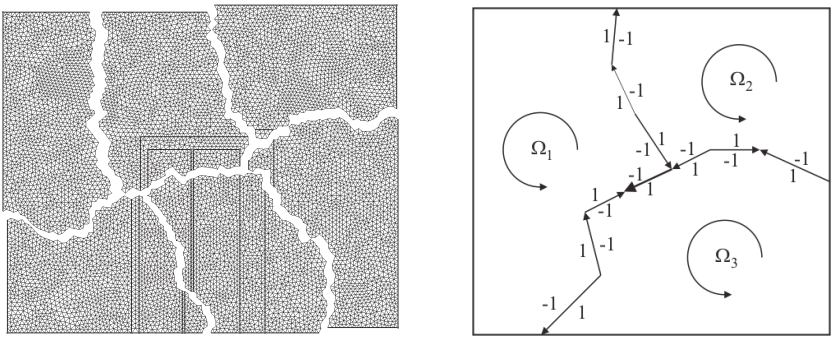


Figure 4. Example of partitioned of domain Ω, and we show the edge directions of the sub-domain interface boundaries, which is very important in the edge element based DDMs.

transfer, because they are the assembly of $\mathbf{K}_{\Gamma_j\Gamma_j}$ and $\mathbf{b}_{\Gamma_j}$, where j is the index of sub-domains.

The assembly and the solution of the sub-matrices can be performed parallel by independent processors. However, the solution requires exchange of interface values between the processes in charge of the various sub-domains. In many practical applications, the preconditioned conjugate gradient (PCG) method is used because of its simplicity and efficiency. The parallel implementation of the preconditioned conjugate gradient method can be presented algorithm 1 [2].

Algorithm 1. Parallel PCG Algorithm.

1 Initialization: $\mathbf{a_0} = \mathbf{0}$,
2 $\mathbf{r_0} = \mathbf{b}_s$
3 Assembly local $\mathbf{r_0}$ with $\mathbf{r}_{\Gamma_{\mathbf{int}}}$ entries → $\bar{\mathbf{r}}_\mathbf{0}$,
4 **for** i = 0, 1, … **do**
5 $\mathbf{w}_i = \mathbf{M}^{-1}\bar{\mathbf{r}}_i$,
6 $\gamma_i = \mathbf{r}_i^{\mathbf{T}}\mathbf{w}_i$,
7 Assembly $\boldsymbol{\gamma_i}$,
8 Assembly local $\mathbf{w}_i$ with $\mathbf{w}_{\Gamma_{\mathbf{int}}}$ entries → $\bar{\mathbf{w}}_i$,
9 **if** i = 0 **then**
10 $\bar{\mathbf{p}}_i = \bar{\mathbf{w}}_i$,
11 **else**
12 $\bar{\mathbf{p}}_i = \bar{\mathbf{w}}_i + (\gamma_i/\gamma_{i-1})\bar{\mathbf{p}}_{i-1}$,
13 $\mathbf{w}_i = \mathbf{K}_s\bar{\mathbf{p}}_i$,
14 $\beta_i = \mathbf{p}_i^{\mathbf{T}}\bar{\mathbf{w}}_i$,
15 Assembly β_i,
16 Assembly local $\mathbf{w}_i$ with $\mathbf{w}_{\Gamma_{\mathbf{int}}}$ entries → $\bar{\mathbf{w}}_i$,
17 $\bar{\mathbf{a}}_i = \bar{\mathbf{p}}_{i-1} + (\gamma_i/\beta_i)\bar{\mathbf{p}}_i$,
18 $\bar{\mathbf{r}}_i = \bar{\mathbf{r}}_{i-1} + (\gamma_i/\beta_i)\bar{\mathbf{w}}_i$,
19 **if** $\gamma_i/\gamma_0 < \varepsilon$ **then**
20 **return**

In the parallel PCG algorithm, $\mathbf{K}_S$ and $\mathbf{b}_S$ are the mass matrix and right-hand side of sub-domain, $\mathbf{a}$ is the unknown potentials, $\mathbf{r}$ is the residual vector, subscript Γ_{int} denotes the external interface entries from neighbouring sub-domains, $\mathbf{M} = \text{diag}(\mathbf{K}_S)$ is the diagonal preconditioning matrix [12], $\mathbf{w}$ and $\mathbf{p}$ are working vectors, and ε is the specified error tolerance.

3.2. Finite Element Tearing and Interconnecting Method

After mesh partitioning (see in fig. 4), the FETI method consists in transforming the original problem, equation (8) with the equivalent system of sub-domain equations [1, 4-6]

$$\mathbf{K}_j\mathbf{a}_j = \mathbf{b}_j - \mathbf{B}_j^{\mathrm{T}}\mathbf{\Lambda}, \tag{10}$$

with the compatibility of the magnetic vector potentials at the sub-domain interface [1, 4-6]

$$\sum_{j=1}^{N_S} \mathbf{B}_j\mathbf{a}_j = \mathbf{0}, \tag{11}$$

where $j = 1 \ldots N_S$, the number of sub-domains, $\mathbf{K}_j$, $\mathbf{b}_j$ and $\mathbf{a}_j$ are respectively the system matrix, the representation of the excitation and the unknown potentials of j^{th} sub-domain. The vector of Lagrange multipliers $\mathbf{\Lambda}$ introduced for enforcing the constraints (11) on the sub-domain interface, and $\mathbf{B}_j$ is a signed ($\pm$) Boolean mapping matrix, which is used to express the compatibility condition at the j^{th} sub-domain interface Γ_j

Usually, the partitioned problem may contain $N_f \leq N_S$ floating sub-domains, where sub-matrix $\mathbf{K}_f$ ($f = 1 \ldots N_f$) is singular. The floating sub-domain means a sub-domain without any constraints on boundary, e.g. Dirichlet boundary condition. To guarantee the solvability of these generally singular problems, we require that [4, 5]

$$\left(\mathbf{b}_j - \mathbf{B}_j^{\mathrm{T}}\mathbf{\Lambda}\right) \perp \mathbf{R}_j, \tag{12}$$

and compute the solution of equation in (10) as [4, 5]

$$\mathbf{a}_j = \mathbf{K}_j^{+}\left(\mathbf{b}_j - \mathbf{B}_j^{\mathrm{T}}\mathbf{\Lambda}\right) + \mathbf{R}_j\boldsymbol{\alpha}_j, \tag{13}$$

where $\mathbf{K}_j^{+}$ is a pseudo-inverse of $\mathbf{K}_j$ if $\mathbf{K}_j$ is singular, else $\mathbf{K}_j^{+} = \mathbf{K}_j^{-1}$. $\mathbf{R}_j = \text{Ker}\left(\mathbf{K}_j\right)$ is the null space of $\mathbf{K}_j$ [12], and $\boldsymbol{\alpha}$ is the set of amplitudes that specifies the contribution of the null space $\mathbf{R}_j$ to the solution$\mathbf{a}_j$. Instead of a pseudo-inverse of matrix, the Moore-Penrose matrix inverse has been used here [12]. The introduction of the $\boldsymbol{\alpha}_j$ is compensated by the additional equations resulting from equation (12) [4, 5]

$$\mathbf{R}_j^{\mathrm{T}}\left(\mathbf{b}_j - \mathbf{B}_j^{\mathrm{T}}\mathbf{\Lambda}\right) = \mathbf{0}. \tag{14}$$

Substituting equation (13) into equation (11), and exploiting the solvability condition (14) lead after some algebraic manipulations to the following interface problem [1, 4-6]:

$$\begin{bmatrix} \mathbf{F}_{\mathrm{I}} & -\mathbf{G}_{\mathrm{I}} \\ -\mathbf{G}_{\mathrm{I}}^{\mathrm{T}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda} \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ -\mathbf{e} \end{bmatrix}. \tag{15}$$

where

$$\begin{aligned} \boldsymbol{F}_I &= \sum_{j=1}^{N_S} \boldsymbol{B}_j \boldsymbol{K}_j^{+} \boldsymbol{B}_j^T, \\ \boldsymbol{G}_I &= [\boldsymbol{B}_1 \boldsymbol{R}_1 \quad \dots \quad \boldsymbol{B}_{N_S} \boldsymbol{R}_{N_S}], \\ \boldsymbol{d} &= \sum_{j=1}^{N_S} \boldsymbol{B}_j \boldsymbol{K}_j^{+} \boldsymbol{b}_j, \\ \boldsymbol{e} &= \left[\boldsymbol{b}_1^T \boldsymbol{R}_1 \quad \dots \quad \boldsymbol{b}_{N_s}^T \boldsymbol{R}_{N_s}\right]. \end{aligned} \tag{16}$$

In order to solve equation in (15) for the Lagrange multiplier vector$\boldsymbol{\Lambda}$, the following splitting of $\boldsymbol{\Lambda}$ is performed [4, 5]

$$\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_0 + \mathbf{P}(\mathbf{Q})\bar{\boldsymbol{\Lambda}}, \tag{17}$$

where $\boldsymbol{\Lambda}_0$ is a particular solution of $\mathbf{G}_\mathrm{I}^\mathrm{T}\boldsymbol{\Lambda} = \mathbf{e}$ and $\mathbf{P}(\mathbf{Q})$ is a projection operator [4], where $\mathbf{Q}$ is an arbitrary positive definite matrix, in this case a unit matrix. The projection matrix is $\mathbf{P}(\mathbf{Q}) = \mathbf{I} - \mathbf{Q}\mathbf{G}_\mathrm{I}\left(\mathbf{G}_\mathrm{I}^\mathrm{T}\mathbf{Q}\mathbf{G}_\mathrm{I}\right)\mathbf{G}_\mathrm{I}^\mathrm{T}$, which the following relationship hold $\mathbf{G}_\mathrm{I}^\mathrm{T}\mathbf{P}(\mathbf{Q}) = \mathbf{0}$.

The classical preconditioned conjugate gradient method cannot be used because it is derived for systems with symmetric positive definite matrices and the matrix $\boldsymbol{F}_I$ does not fulfil this requirement [1]. Therefore a modification which is based on the application of an additional condition, $\mathbf{G}_\mathrm{I}^\mathrm{T}\boldsymbol{\Lambda} = \mathbf{e}$ must be included. Additional condition must always be satisfied during the iteration process.

The algorithm of the preconditioned modified conjugate gradient method is summarized in algorithm 2 [1, 4].

Algorithm 2. PMCG Method.

1 Initialization: $\boldsymbol{\Lambda_0} = \mathbf{Q}\mathbf{G}_\mathbf{I}\left(\mathbf{G}_\mathbf{I}^\mathbf{T}\mathbf{Q}\mathbf{G}_\mathbf{I}\right)^{-\mathbf{1}}\mathbf{e}$,
2 $\mathbf{r_0} = \mathbf{d} - \mathbf{F}_\mathbf{I}\boldsymbol{\Lambda_0}$,
3 $\mathbf{w_0} = \mathbf{P}(\mathbf{Q})^\mathbf{T}\mathbf{r_0}$,
4 $\mathbf{h_0} = \mathbf{P}(\mathbf{Q})\mathbf{F}_\mathbf{I}^\mathbf{L}\mathbf{w_0}$,
5 $\mathbf{s_0} = \mathbf{h_0}$,
6 **while** $\|\mathbf{r}\|_2 < \varepsilon$ (k = 1, …) **do**
7 $\alpha_k = \frac{\mathbf{h}_k^\mathbf{T}\mathbf{s}_k}{\mathbf{s}_k^\mathbf{T}\mathbf{F}_\mathbf{I}\mathbf{s}_k}$,
8 $\boldsymbol{\Lambda_{k+1}} = \boldsymbol{\Lambda_k} - \alpha_k\mathbf{s_k}$,
9 $\mathbf{r_{k+1}} = \mathbf{r_k} - \alpha_k\mathbf{F_I}\mathbf{s_k}$,
10 $\mathbf{w_{k+1}} = \mathbf{P}(\mathbf{Q})^\mathbf{T}\mathbf{r_{k+1}}$,
11 $\mathbf{h_{k+1}} = \mathbf{P}(\mathbf{Q})\mathbf{F}_\mathbf{I}^\mathbf{L}\mathbf{w_{k+1}}$,
12 **for** $0 \le j \le k$, **do**
13 $\beta_k^j = -\frac{\mathbf{h}_{k+1}^\mathbf{T}\mathbf{F_I}\mathbf{s_j}}{\mathbf{s}_j^\mathbf{T}\mathbf{F_I}\mathbf{s_j}}$,
14 $\mathbf{s_{k+1}} = \mathbf{h_{k+1}} + \sum_{j=1}^{k}\beta_k^j\,\mathbf{s_j}$.

In this paper, the also called lumped preconditioner $\mathbf{F}_\mathrm{I}^\mathrm{L}$ has been used [1, 4]. In this case, each sub-domain mass matrix is partitioned as in equation (9),

$$\mathbf{K}_j = \begin{bmatrix} \mathbf{K}_j^{ii} & \mathbf{K}_j^{i\Gamma} \\ \mathbf{K}_j^{\Gamma i} & \mathbf{K}_j^{\Gamma\Gamma} \end{bmatrix}, \tag{18}$$

where the superscript i and Γ designate the sub-domain interior and interface boundary DoF, respectively. The lumped preconditioner is given by [4, 5]

$$\mathbf{F}_{\mathrm{I}}^{\mathrm{L}} = \sum_{j=1}^{N_S} \boldsymbol{B}_j \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_j^{\Gamma\Gamma} \end{bmatrix} \boldsymbol{B}_j^T. \tag{19}$$

The PMCG algorithm is run sequential to solve the equation (15) interface problem. After obtaining the solution of Lagrange multipliers, $\boldsymbol{\Lambda}$, $\boldsymbol{\alpha}$ can be calculated as [4]

$$\boldsymbol{\alpha} = \left(\mathbf{G}_{\mathrm{I}}^{\mathrm{T}}\mathbf{Q}\mathbf{G}_{\mathrm{I}}\right)^{-1}\mathbf{G}_{\mathrm{I}}^{\mathrm{T}}\mathbf{r}, \tag{20}$$

where $\mathbf{r}$ is the final residual term which is equivalent the jump of potentials between sub-domains, and the sub-solution can be calculated parallel by equation (13).

4. Results and Discussion

In order to compare the iteration counts and speedup of the methods, we have run a number of test cases using a research code that has been developed for that purpose on the Matlab computing environment. This code simulates the state of the art techniques used to implement the discussed DDMs in lower level programming languages (e.g. FORTRAN, C) for high performance application.

The computations have been carried out on a massively parallel computer (SUN Fire X2250). This computer works with a shared memory topology with two Quad-Core Intel® Xeon® processors. The parallel programs have been implemented under the operating system Linux.

The first parabolic type problem is a shell-type transformer (see in fig. 4), and it contains 120994 edge elements, which means 81232 3-node triangular elements and 39763 nodes. At this problem, the impressed current vector potential [9] computation has been performed by edge finite element approximation, and the magnetic vector potential [9, 10] has been approximated by nodal shape function, because it has only z-component. This problem has been used to validate the results of parallel edge element based finite element technique. The second parabolic type problem is the plate problem (see in fig. 5), and it contains 73984 edge elements, which means 49152 3-node triangular elements and 24833 nodes.

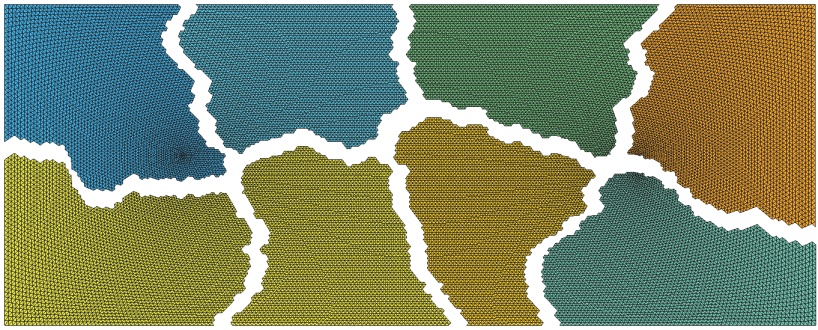


Figure 5. Irregular decomposition of mesh into 8 sub-domain.

In order to use the same stop criterion for the methods, $\varepsilon = 10^{-9}$. The speedup has been calculated by the following formula, $speedup = Time_1/Time_n$, where $Time_1$ is the running time of the sequential algorithm or the running time with least processor number, and $Time_n$ is the running time of the parallel algorithm executed on n processor [1, 13].

In the tables are summarized the number of applied processor cores, N_p, the number of degree of freedom of one sub-domain, DoF, the number of interface unknown, $C.DoF$, the computation time (wall clock time) in seconds, $Time$, and the number of iteration of the iterative solvers, $NoIt$.

The parallel performance results of Schur complement method and the FETI method for the first parabolic type problem summarized in fig. 6, when the number of processor cores is varied between 4 and 8. The speedups are computed using $N_p = 4$ as the reference point. The Schur complement method achieves good speedups, and the FETI

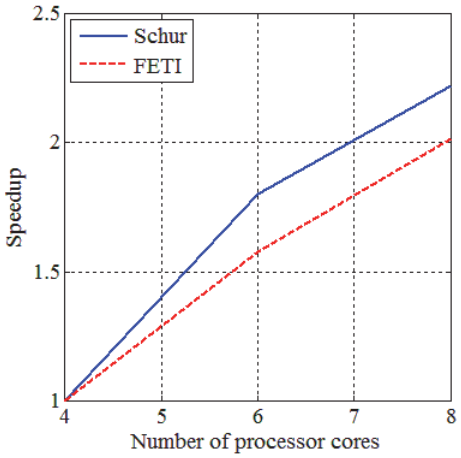


Figure 6. The speedup of the first parabolic type problem.

Table 1. Performance comparison – First parabolic type problem at 10Hz.

N_p		4	6	8
	DoF	30477	20355	15297
	$C.DoF$	424	650	827
Schur Complement Method	$Time$	152.332	84.555	68.495
	$NoIt$	1128	1134	1146
FETI Method	$Time$	801.086	508.481	397.266
	$NoIt$	204	192	201

method solver achieves reasonable ones. However, the total wall clock times ($Time$) reported in table 1 show the big difference of the running performance. The reason of this difference, the calculation of null space and pseudo-inverse are very time consuming in edge element case, because the mass matrices of sub-domains are singular.

For $1 \leq N_p \leq 8$, the performance results of the Schur complement method and FETI method are reported in fig. 7 for steel plate problem. The speedups are computed using the wall clock time of sequential calculation (197.004 sec) as the reference point. The Schur complement method achieves good speedups, and the FETI method solver achieves reasonable ones. Furthermore, the total wall clock times ($Time$) reported in table 2 show nearly same running performance, because this problem is relatively small as the transformer problem.

It seems to be, the Schur complement method solved the problem faster, than the

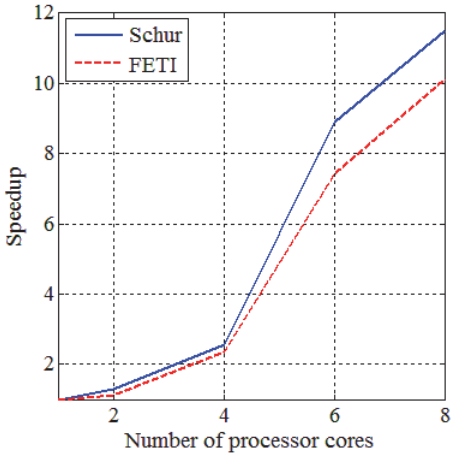


Figure 7. The speedup of the second parabolic type problem.

Table 2. Performance comparison – Second parabolic type problem at 5Hz.

N_p		2	4	6	8
	DoF	37060	18594	12436	9342
	$C.DoF$	135	416	592	736
Schur Complement Method	$Time$	151.319	77.416	22.102	17.118
	$NoIt$	171	174	176	176
FETI Method	$Time$	176.428	84.291	26.561	19.422
	$NoIt$	67	68	72	73

FETI method. However, the number of iteration reported in table 1 and 2 show the solver of FETI method faster convergence rate as the solver of Schur complement method. This conclusion is also supported by the convergence curves gives in fig. 8 and 9. These figures show the convergence curves of iterative solvers at steel plate problem when using $N_p = 8$ and three different frequencies.

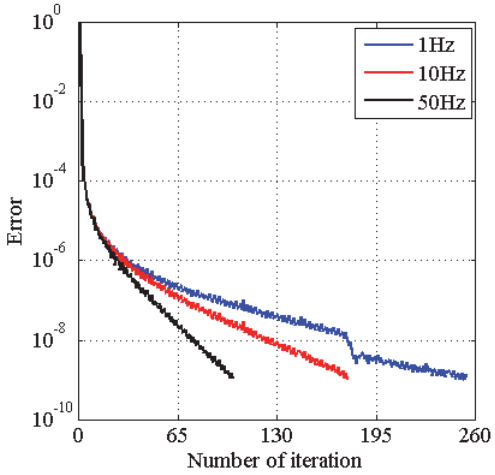


Figure 8. Convergence behaviour of PCG algorithm at different frequencies.

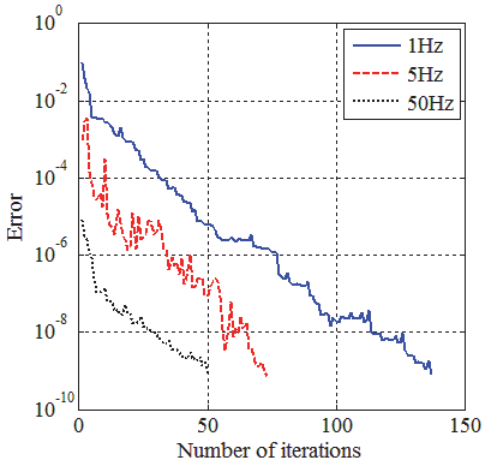


Figure 9. Convergence behaviour of PMCG algorithm at different frequencies.

5. Conclusions

In this paper, we have presented the description of Schur complement method and FETI method for edge element finite element method. We have illustrated the speedup of the methods and the convergence behaviour of solvers with the solution of two-dimensional parabolic type problems on massively parallel computer.

In all cases, the Schur complement method outperforms the FETI method. The computation of the null space and pseudo-inverse of the sub-domain mass matrices can become the Achille's heel for the FETI method in edge element case. The computation costs and the memory requirement are very high of these calculations. This is the reason, why the total times of Schur complement method are smaller. However the PMCG method convergence rate is better as the PCG method for Schur complement method.

We have to note that only two benchmarks have been used for the numerical tests. The tests with more complex three-dimensional problems will be the subject of a forthcoming work.

Acknowledgement

This paper is sponsored by "TÁMOP-4.2.2.A-11/1/KONV-2012-0012: Basic research for the development of hybrid and electric vehicles - The Project is supported by the Hungarian Government and co-financed by the European Social Fund".

References

[1] Kruis, J.: *Domain Decomposition Methods for Distributed Computing*, Saxe-Coburg Publications, Kippen, Stirling, 2006

[2] Nikishkov, G. P.: *Basics of the domain decomposition method for finite element analysis*, in Mesh Partitioning Techniques and Domain Decomposition Methods, Editor: Magoulés, F., Saxe-Coburg Publications, Kippen, Stirling, pp. 119-142, 2007

[3] Marcsa, D., Kuczmann, M.: *Comparison of domain decomposition methods for elliptic partial differential problems with unstructured mesh*, Przeglad Elektrotechniczny, vol. 2012, no. 12b, pp. 1-4, 2012

[4] Farhat, C., Pierson, K., Lesoinne, M.: *The second generation FETI methods and their application to the parallel solution of large-scale linear and geometrically non-linear structural analysis problems*, Computer Methods in Applied Mechanics and Engineering, vol. 184, no. 2-4, pp. 333-374, 2000
DOI: 10.1016/S0045-7825(99)00234-0

[5] Farhat, C., Roux, F. X.: *Method of finite element tearing and interconnecting and its parallel solution algorithm*, International Journal for Numerical Methods in Engineering, vol. 32, no. 6, pp. 1205-1227, 1991
DOI: 10.1002/nme.1620320604

[6] Toselli, A., Vasseur, X.*: Robust and efficient FETI domain decomposition algorithms for edge element approximations*, COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, vol. 24, no. 2, pp. 396-407, 2005
DOI: 10.1108/03321640510586033

[7] Geuzaine, C., Remacle. J. F.*: Gmsh: A Three-Dimensional Finite Element Mesh Generator with Built-in Pre- and Post-Processing Facilities*, International Journal for Numerical Methods in Engineering, vol. 79, no. 11, pp. 1309-1331, 2009
DOI: 10.1002/nme.2579

[8] *METIS - Serial Graph Partitioning and Fill-reducing Matrix Ordering*, available at: http://glaros.dtc.umn.edu/gkhome/views/metis (accessed 18 January 2014)
[9] Kuczmann M., Iványi, A.: *The Finite Element Method in Magnetics*. Akadémiai Kiadó, Budapest, 2008
[10] Bíró, O.: *Edge element formulation of eddy current problems*, Computer Methods in Applied Mechanics and Engineering, vol. 169, no. 3-4, pp. 391-405, 1999 DOI: 10.1016/S0045-7825(98)00165-0
[11] Bianchi N.: *Electrical Machine Analysis Using Finite Elements*, Taylor & Francis, Boca Rotan, FL, USA, 2005
[12] Saad, Y.: *Iterative Methods for Sparse Linear Systems*, 2nd edition, SIAM, 2003
[13] Molnárka G., Varjasi N.: *A Simultaneous Solution for General Linear Equations on a Ring or Hierarchical Cluster*, Acta Technica Jaurinensis, vol. 3, no. 1, pp. 65-73, 2010

Acta
Technica
Jaurinensis

Vol. 7., No.2., pp.207-223, 2014
DOI: 10.14513/actatechjaur.v7.n2.283
Available online at acta.sze.hu

Effective End-user Interfaces for Various Business Needs

A. Selmeci, T. Orosz

Óbuda University, Székesfehérvár, Hungary
e-mail:selmeci.attila@arek.uni-obuda.hu, orosz.tamas@arek.uni-obuda.hu

Abstract: SAP, the leading Enterprise Resources Planning System in the World, has been providing different effective solutions for business requirements for more than 40 years. Therefore SAP tried to follow the changes of the end-user requirements with different user interface solutions. The technology novelties drive the internal developments to support new protocols and offer more visual components. Our paper discovers the possibilities lying on the desk and gives directions to select the most effective technique for a given business need. In the last years the end-user want to have efficient UIs, which bring more potent to the daily work. Beyond offering effective and efficient options we uncovered the power of the different solutions in changing environments as well.

Keywords: user interface, SAP, front-end, efficient surface, sustainable UI

1. Introduction

SAP has a client-server architecture, which is enhanced according the market needs. This brings the changes and extension of user interface techniques as well [1-3]. This paper analyzes the different available approaches by showing the effectiveness of them. It describes the usability from end-user and project perspectives as well. To understand the mechanism of the different UI techniques first we explain the SAP architecture from UI point of view. In the paper we should distinguish between front-end based and Web-based UI surfaces also.

2. Basic Architecture of the SAP Systems

SAP solutions are based on the SAP Web Application Server technology. This technology provides open, scalable and robust infrastructure for running and developing dynamic applications. This is the core element of the SAP ERP (successor of R/3) system as well. The earlier R/2 releases had two layers, where the business logic and the storage as a monolith element took place on a mainframe, and the presentation layer appeared on a terminal. The communication capabilities were limited to LU 6.2 for other systems. R/3 systems inherit the real-time capability as we can see in the 'R' letter, but is redesigned for three-tier client-server architecture (see Fig. 1).

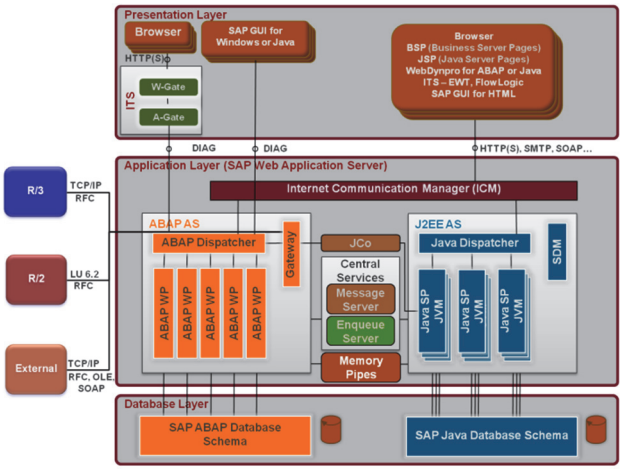


Figure 1. WebAS high-level technical architecture

This architecture means that the application is divided into three layers:

- Database layer: this is a single component (except some supported grid database systems), where the business data, and the whole repository are stored.
- Application layer: it is responsible for controlling and processing the applications. The business logic is running in this layer providing system-oriented services and ensures connectivity to presentation and other software components. For Web application this layer can be separated to more parts determining the connectivity, presentation, business logic itself, integration and persistent layers. (See details on Fig. 2. and later.)
- Presentation layer: it is the front-end, the Graphical User Interface, that runs on PCs, workstations, Web browsers or even on mobile devices.

Fig. 1. shows above the today's high-level technical architecture of the SAP Web Application Server.

The above figure refers some till now not detailed component and protocols used in communications. In the early R/3 system, before WebAS SAP does not provided direct HTTP protocol, but only DIAG (Dynamic Information and Action Gateway) and RFC (Remote Function Call) were provided. The DIAG is SAP's own protocol to communicate between the application layer and the SAP front-end software, called SAP GUI. For real data exchange SAP uses the CPI-C (Common Program Interface for Communication) based RFC protocol. This is a program-to-program communication protocol, which was used and implemented in the R/2 and R/3 world. The figure mentions the R/2 connection via LU 6.2 protocol. CPI-C and RFC communications could use TCP/IP and LU 6.2 communication as well. The IBM's Logical Unit (LU) 6.2 is part of the System Network Architecture (SNA) protocol. SAP provided by RFC a possible communication layer where other programs, system or even own developed front-end applications could use the system services.

SAP introduced the SAP NetWeaver Application and Integration platform around in 2003. This initiative was designed to give a comprehensive platform for many SAP applications and provide integration capabilities not only on technical level, but from business-oriented view as well. SAP NetWeaver integration platform is open and manages heterogeneous environments by the integration layers (containing several solutions, applications) from bottom up:

- Process integration (or the newer name: Business Process Integration): this integration layer guarantees the capability to integrate systems, technologies by interoperability using EAI (Enterprise Application Integration) tools as part of NetWeaver: Process Integration (PI), Business Process Management (BPM), Integration Broker.
- Information integration: As the process integration layer connected technically the implemented solutions in a company and in case of BPM on business level as well, higher abstraction level integration is also available. This integration manages data harmonization, consolidation through the company and beyond. Different data can be integrated in a company and different tools could solve the integration tasks. The documents and other contents should be organized around taxonomies and the unstructured objects should be accessed in a structured and role based manner. For this purpose SAP provides the Knowledge Management solution. The business data can be collected in data warehouse (BW, Business Warehouse) system for consolidation and business intelligence functions like reporting, dashboards and so on. SAP provides master data harmonization, mapping and full management capabilities within the information integration solution by Master Data Management (MDM).

People integration: as we integrated our best of bread (not only SAP solutions) through process and information integration it is useful to support user level integration as well. Main function of people integration is the collaboration. On the other hand the user interfaces from the different systems are integrated on the SAP NetWeaver Portal providing the right information to the right person on right time in right place. Beyond the portal the NetWeaver SOA (Service Oriented Architecture) capabilities offer a unified development and run-time tool for cross applications. The SAP Composite Environment (CE) supports the development of composite application providing web user interfaces for the new applications calling services from the different underlying application (even if they are beyond the company's boundaries).

The NetWeaver implements and applies open standards and can interoperate with other technologies like IBM WebSphere, J2EE or Microsoft .NET. As we mentioned, the NetWeaver is not only an integration platform, but an application platform as well. The application platform as it is shown on the above picture has two personalities: ABAP and J2EE. The ABAP is the original SAP owned application server, which is a process based server communicating with the database layer via process-to-process channel. The standard clients communicating via RFC or DIAG are connecting to the so-called dispatcher process, which dispatches the tasks to an available (empty) dialog process after queuing it for execution of the request [4-5]. The clients can log into an SAP ABAP system via server selection or group selection. In the first case the client is directly communicates with an SAP application instance immediately, but in the second

case the client asks the so-called message server to provide login information from a corresponding application instance. Corresponding mean in this case a good performing instance, so the message server helps in the logon load balancing as well. (The message server communicates short technical messages among the available application instances to provide an overall good performance from the response time point of view and collects data about the available services and load as well [6-8].) ABAP is the name of the internal SAP programming language referring to Advance Business Application Programming. Each of the SAP business applications are developed in ABAP, which is an open 4th generation event controlled language enhanced with object-orientation as well. The ABAP language has a so-called Open-SQL part for database communication as well. It provides (as described above) RFC protocol, which is an SAP implementation of RPC (Remote Process Call). SAP delivers and makes available to develop function modules (compilation units), which are remote enabled, and let other systems, programs, and applications to call services of the SAP systems. Main functionalities are implemented as function modules and many of them are RFC functions, so the SAP systems can be controlled and updated from outside via RFCs as well.

The other parts of the drown architecture will be discussed later to make easier the understanding of the expansion of the application server to Web Application Server.

As of NetWeaver releases the SAP Application Server is expanded with Web server functionality as well, and the whole SAP Web Application Server architecture can be separated as we mentioned above into five sub-layers according to the technical areas (see Fig. 2).

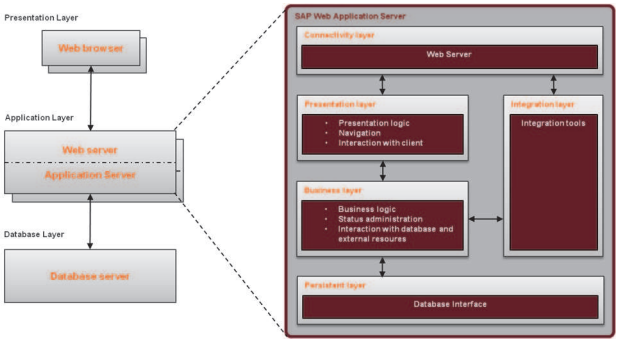


Figure 2. Sub-components of the Web Application Server

The 5 sub-components of the WebAS architecture:

- Presentation layer: presentation logic (view or screen definitions and switches), navigation, and the interaction with the client are defined here. The user interface can be developed on J2EE platform with Java Server Pages (JSP) or WebDynpro technology; on ABAP platform with Business Server Pages (BSP), WebDynpro for ABAP technology, or ITS (Internet Transaction Server) IAC (Internet Application Component, see later). For these kinds of surfaces the underlying business layer provides the business login and content.

- Business layer: the business logic is running here, so the status of the applications are administered as well on this layer, and of course the service call like database connection or external resource usage are defined and maintained here. This layer provides run-time environment for the business logic by processing the requests passed from the ICM (Internet Communication Manager, see below), and the responses are generated also here. The business logic (encapsulating the persistent layer) is implemented in J2EE environment by Enterprise JavaBeans (EJB), or in the ABAP environment they are developed in the ABAP Workbench using local objects and persistence.

- Integration layer: this layer opens the Web AS via integration engine to communicate via standard interfaces with other systems and solutions. Message service is provided for communication between the Web AS and through SAP XI (Exchange Infrastructure, basic EAI solution by SAP) connected systems. It enables other systems to integrate their functionalities and services to SAP Web AS for usage in Web applications.

- Connectivity layer: this is the first layer facing with client or connected systems, application. The main service, which provides the Web server functionality and other standard protocols, is the Internet Communication Manager (ICM). Through the Internet Communication Framework (ICF) many standard protocols are available, like Hypertext Transfer Protocol (HTTP), HTTPS (extension of HTTP running under the Secure Socket Layer (SSL)), Web Distributed Authoring and Versioning (WebDAV), Simple Object Access Protocol (SOAP), Fast Common Gateway Interface (FastCGI), and Simple Mail Transfer Protocol (SMTP), etc. ICM dispatches UI requests to the above mentioned presentation layer, and the ICF is used for connectivity using the listed, various communication protocols.

- Persistence layer: as we described above the Web AS ABAP environment guarantees database independent programming by Open SQL. The database interface (DBIF) handles the SQL requests and translates them to the underlying native database SQL language in an optimized way. SAP developed the Open SQL capabilities to the Java world as well and offers a variety of standard Application Programming Interfaces (APIs) to Java programmers, such as SQLJ and other Java technologies.

As we have seen the SAP Web AS with its two (ABAP and J2EE) personalities provides not only thick client capabilities (see it soon), but also many different Web UI options are offered out of the box.

3. Frontend Possibilities of SAP Systems

SAP provides a universal front-end solution called SAP GUI (SAP Graphical User Interface). SAP at the beginning provided stand-alone SAP GUI solutions for Mac, OS/2, Windows and so on, but later decided to support three flavors according the available platforms. The SAP GUI family contains

- SAP GUI for Windows supporting MS Windows operating systems based on OLE and ActiveX controls
- SAP GUI for Java supporting all environments running Java runtime environment. This GUI is generally called platin GUI as well, because it is a platform independent solution.
- SAP GUI for HTML is a web based GUI frequently called as WebGUI because of the name of the service, which provides it.

SAP communicates with the SAP GUI client via DIAG protocol and in some cases RFC is used as well. The SAP GUI is a thick client; it must be installed onto the front-end workstation. The standard GUI for SAP is the SAP GUI and it was designed for end-users working with numbers, calculations and so on (as the Fig. 3. shows on the right side).

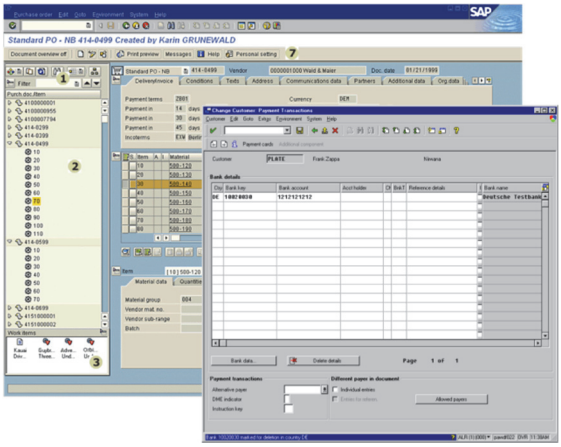


Figure 3. SAP GUI with and w/o Enjoy controls

The figure above contains two screen shots. The right one refers the original R/3 SAP GUI for Windows look. Each SAP application looks like the other, because the client software retrieves the screen element type, position, size, and of course the content and some other attributes, but the final design takes place on the front-end machine. Each of the screens has a standard tool bar containing standard, almost always available functions (like save, print, page up and down, etc.), a menu up to 8 items, but the last two items are always accessible having standard functions. If the application requires another, so-called customer tool bar can be defined holding pushbuttons with special functions. The surface itself can contain many different screen elements, like text field, input field, tab strip, table control, checkbox, radio button, frame, pushbutton, etc. The designs of these elements are predefined in the client software.

There are two main look and processing options of SAP programs: reports and screens. The two kinds of program are different because the reports are generally reading programs, which collect data and present them in a list. (In some cases report can be used for mass changes as well.) The screen or transaction programs are designed to maintain data in the SAP systems. An SAP transaction (Logical Unit of Work) leads

through more screens (or dynpro-s as the original name refer to dynamic program) collecting different data for a special purpose before booking it into the database. Behind the screens the transaction main program holds the status of the transaction and handles the dataflow and navigation between the screens. Each screen contains beyond the element list and attributes, the so-called flow logic, which has (at least) two main parts the Process Before Output (PBO) and Process After Input (PAI) event blocks. These event blocks are executed for each screen at the time as their names describe, so PBO can contain reading customizing, checking authorization, etc. and with these information we can set default values and modify the attributes (e.g. required or read-only) of the screen elements before it is sent to the client. When the client executes a function, which starts a request to the Application Server the client sends the information and the PAI event is raised. The transaction executes the PAI event block, where the entered data is checked against conditions and if everything is correct navigation to another screen or database updatecould take place. The above figure shows a screenshot of a transaction dynpro.

As we described the SAP application look similar. It is because SAP recommends to use a standard style guide not only for standard SAP delivered programs and applications, but for customer developed once as well. For that SAP owns a Web site http://www.sapdesignguild.org/ containing the relevant information for programmers or designers.

SAP realized that the users want to have a bit easier, colorful, manageable surface, so as of release R/3 4.6 SAP implemented the so-called Enjoy controls to make the end-user of SAP's software a V.I.P. meaning new visual design, new interaction design and personalization options. As it is shown on the left side of Fig. 3. SAP implemented many new screen elements and the whole design was changed a bit as well. The Enjoy controls are implemented as ActiveX controls in case of SAP GUI for Windows, but for JavaGUI SAP used JavaBeans instead.

SAP implemented Text Editor, HTML and picture viewer, tool bar, hierarchical tree controls and the so-called ALV (ABAP List Viewer) control, among others. The programming of controls is a bit different, because in case of controls a huge data amount is sent from the Application Server to the client to store and provide to the corresponding control. Automation handling and OLE or Java means (depending on SAPGUI for Windows or for Java) do this data exchange. In the server side embedding proxy objects handle the request and manages the data exchange through the control framework. If we use Enjoy controls we should keep in mind the data amount and the performance and use the available optimization techniques (e.g. incremental tree construction) not to put too much load to the network or to the front-end PC at a time. As an example the Tree controls make it possible not to send the whole tree in one step but only some levels (e.g. the first only) of the tree. If the end-user wants to open a tree node defined as a folder without having data under it, the client sends back an event for the child node (`EXPAND_NO_CHILDREN`) to bring to corresponding data. If the node is not opened the data will not be collected and sent to the client, so application processing, network and client side storage resources are saved.

SAP GUI gives enhanced opportunities implemented on client side as well. In this case depending on the client PC setting the same SAP transaction, screen can look different. Two options are available currently:

- GuiXT
- GUI Scripting.

For the GuiXT implementation on the front-end PC the GuiXT component should be installed as well (as part of the standard SAP GUI product CD). For each screen referred by transaction main program (module pool) name and the screen number a separate text file can be created for each language containing the script for the GuiXT. With the GuiXT screen elements can be hidden, moved, and new one pasted on the screen. This program runs with the SAP GUI process and modifies the look according to the script definition if available for the current screen. As an example an image can be placed to the screen be the following script row:

```
Image (20,30) "C:\Images\maci.jpg"
```

The local image file**C:\Images\maci.jpg** is displayed in row **20**, column**30**. This capability was delivered with the SAP GUI for earlier releases as well, than the Enjoy controls handle any pictures. The solution was designed by Synactive GmbH, and the whole documentation and technology of GuiXT is written on the http://www.synactive.com site.

The GUI Scripting is a different scripting technology and it was introduced as of Web AS 6.20. SAP GUI Scripting do not change the screen layout as GuiXT does. An object model is delivered, where the screen is represented with its controls at runtime and an API can be used to modify the object itself. The API can be used from the VBScript and JavaScript languages. The main different is that the GUI scripting will not change the surface, but a script can be created manually or even by recording a SAPGUI execution and the script content can be change to execute again with other data. It is useful to automate such tasks, which should be executed repeatedly. It is a very good tool for testing a transaction.

SAP developed for non-standard display environments the SAPConsole client to support character-cell terminals, including radio frequency (RF) devices or even web-equipped devices. This is not a SAP GUI, but a different approach to handle special UI-s. Two possible architecture are available: telnet client with a telnet server running the console displayer, and a browser based solution where an IIS Web Server runs the Web Displayer program, which communicate with the device. Fig. 4. shows an example.

Figure 4. SAP Console example

4. External, Developed Frontend options

As a communication layer SAP provides RFC enabled function to the outside world. These functions can be called from own developed programs or applications. The architecture figure shows above that the SAP communication through the gateway process, which handles the program-to-program communication to partner applications. SAP offers for external developers connectors to make easier the communication with the SAP:

- RFC Library: this is the classic RFC connector, which offers with its RFC API C-routines to create external RFC capability for own developments. RFC library makes it possible to create server or client RFC capable programs as well is C/C++ language. This library will be replaced by the enhanced NetWeaver RFC Library, which provides more features and improvement, but is not compatible to the classic RFC Library.
- SAP Java Connector: with the SAP JCo Java application can communicate with SAP systems. The connector contains toolkit for providing and consuming RFC services as well. As on the Fig. 1. is drown, SAP uses this connector for inside communication as well. (The ABAP stack of the Web AS is process based, but the Java stack is thread based, so different connections are needed for communications.)

SAP Connector for Microsoft .NET (SAP NCo): It makes possible to call SAP RFCs or BAPIs directly from .NET applications even it is written in Visual Basic, C++ or C#. The provided Proxy Wizard makes possible the integration of business data and processes of SAP systems into .NET by generating proxy classes, which contain the required methods, attributes (properties) and some table structure definitions. The connector gives easy connection with pre-generated content (e.g. logon) in the Visual Studio .NET environment. SAP NCo manages SAP SOAP connections as well.

With these connectors it is possible to develop any required UI for SAP forgetting the standard SAP GUI and its possibilities. The basis of the developed UI is the RFC modules, which could provide data from and to the SAP system.

SAP introduced the Business Framework Architecture (BFA) to provide an object-oriented layer on the SAP application functionality. The BFA collects the Business Objects, like Employee, Sales Order in so-called Business Components. Business

Objects have attributes, methods, events, as any object can have in an object-oriented environment. These Business Object components can be used within an SAP system, but some of the methods are public for even external programs as well. These public methods are the BAPIs (Business Application Programming Interfaces). Technically the BAPIs are implemented as remote-enabled function modules in the SAP system, but some of the external codes (using corresponding connectors) can refer to them not only as RFC functions, but also as methods of objects. It means real object oriented object are created on the caller side and the methods (only the public ones, the BAPIs) can be invoked. For sustainability SAP freezes the interface definitions of the BAPIs. The older programs and external UIs can still use the old BAPIs. If interface modification needed, SAP implements a new method, new BAPI with the new signature.

SAP can be used as OLE (Object Linking and Embedding) server or as OLE client. For UI development we can use SAP as an OLE server. Visual Basic or even Excel can be used directly to call up SAP, log in and use according the login authorization the (remote) services of the system. With this feature simple Excel applications can be developed where behind an SAP system provides (or stores) the data, so Excel is a special UI for SAP.

SAP itself uses this as well for Business Warehouse (SAP BW) UI, because the SAP Business Explorer (BEx) is an Excel based solution (delivered as a front-end component with SAP GUI as well). The end-user can evaluate old and current data to varying degrees of detail and from different perspectives on the Web and also in MS Excel.

5. SAP Web Frontend Options

SAP implemented around middle 90's a special Web enablement for the systems. During that time SAP had no Web Application Server technology as Fig. 2. shows, but a separated component was introduce to convert the SAP DIAG protocol to the Web HTTP protocol. This converter component is the Internet Transaction Server (ITS). The ITS contains two separated sub-component: application gate (A-gate) for application server communication and a Web gate (W-gate) running on a Web server providing mime object and rendering the converted information with additional date to provide a usable html page for the end-user. SAP offers three programming models as shown on Fig. 5.

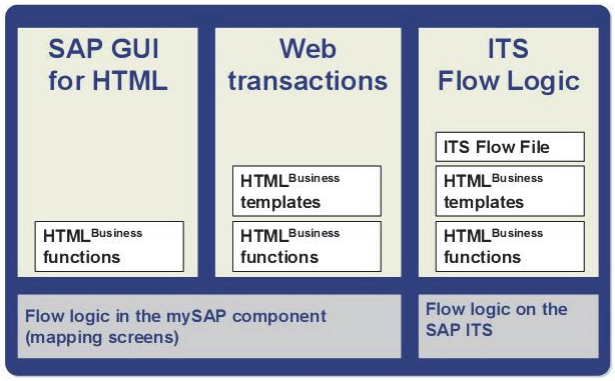


Figure 5. SAP ITS Programming Models

Earlier we mentioned that SAP GUI family has an SAP GUI for HTML, the so-called WebGUI. This is provided as a service by the ITS, which converts the standard SAP content to HTML Business functions. This UI is a standard front-end like each of the elements of the SAP GUI family.

The EWT (Easy Web Transaction) is a transaction, where the status of it takes place in the SAP system, because the basis of an EWT is a standard SAP transaction. Each of the screens (Dynpros) of the SAP transaction is translated in design time to so called HTML Business templates, which contain the screen element definitions using HTML Business functions. These templates are stored on the ITS server, and can be modified and enhanced according to the Web requirements (e.g. images can be placed instead of buttons, etc.). When the end-user executes the transaction on the Web, SAP starts it in the backend (using the PBI of the first screen) and sends the data (via DIAG protocol) to the ITS, which puts the data into the html page using the pre-generated HTML templates and the W-gate adds mime objects if needed. When the end-user executes a function (e.g. presses a pushbutton) the ITS receives the request and translates to the SAP "language" of the SAP protocol and sends forward to the SAP application server. The SAP executes the PAI processing block of the current screen and sets the status of the transaction. This UI serves as an easy way of converting existing SAP transaction to Web format.

SAP ITS offers a third programming model, where the status of the transaction is stored on the ITS side. The ITS Flo Logic manages event flows, where the ITS can call back to the SAP system to retrieve or store data if the process flow requires it, but the screen definition and the flow logic is totally defined on the ITS side. ITS can use only remote enabled function for this purpose, so only RFCs and BAPIs can be used for this task.

As the Fig. 2. shows the new SAP Web Application Server itself contains a built in Web server and with it offers beyond the old DIAG and RFC protocols many new a standard protocols like HTTP, SOAP, SMTP, etc. The main component of the Web enablement is the Internet Communication Manager (ICM) and the framework (ICF) around it. The services offered by a Web Application Server are listed in the ICF. As a special service SAP offers the so-called Integrated ITS as a service as well. So the Fig.

1. has the ITS offerings in the presentation layer because of the service. The `webgui` service is also offers as a standard service without installing any other components for it.

In the ABAP Web AS SAP introduced the page based server side scripting Web programming tool, the Business Server Pages (BSP). This is a real internal development tool and environment for Web UIs using local (e.g. database, file system) or remote (other SAP or non-SAP systems remote enabled functionalities) resources for data retrieval and storage. A BSP application is a complete functional web application containing everything like a classic SAP transaction. BSP defines a Web UI containing server-side scripting, arbitrary additional files (Images, background pictures, button GIFs,), Style-Sheets, etc. The technical structure of a BSP Application is shown on Fig. 6. Essentially a BSP application is divided to user interface and business logic. The user interface itself includes the static Web pages, dynamic, generated Web sites, which can be templates containing server-side scripting, where as scripting language JavaScript or ABAP can be used.

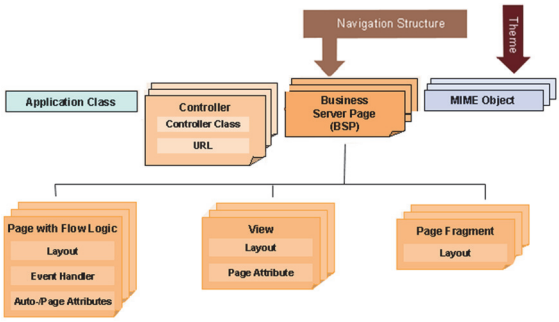


Figure 6. Structure of an SAP BSP Application

The scripts are executed dynamically during runtime. The generated static Web sites can be defined as pages with flow logic or even views. Depending on the requirements and the developing style pages can be defined with flow logic, simple views or only page fragments as the structure hierarchy of the BSP application shows on Fig. 6. The pages with flow logic are the original BSPs defining a page layout with scripts, attributes holding the state of the page or the application (depending on the definition), and event handlers to process end-user actions. The code snippet quoted below show a very simple example of page with flow logic using ABAP scripting. There are page attributes (`airlines, connections`) defined separately as global fields for the whole page, so the event handlers can use their content as well. The example creates a simple tree displaying connections by airlines. The code shows special and unusual tags. The `htmlb` tags as a library are defined according to the HTML for business definition, just like with ITS applications. There are many delivered libraries to use in BSP pages and the developer can define his own as well. Such library can be used as a rendering family, and makes it possible to use high-level programming constructs. The developer can achieve faster rendering than by hand. These tag libraries give one of the strengths of BSP applications.

```
<%@page language="abap" %>
<%@extension name="htmlb" prefix="htmlb" %>
<htmlb:content design="design2003" >
<htmlb:page title="Displaying a tree " >
<htmlb:form>
<htmlb:treeid    = "forest"
title = "This is a tree " >
<%
data: airline like line of airlines.
loop at airlines into airline.
%>
<htmlb:treeNode
id= "<%= airline-carrid %>"
text   = "<%= airline-carrname %>"
isOpen = "false" >
<%
data: connection
like line of connections.
read table connections
with key carrid= airline-carrid
transporting no fields.
ifsy-subrceq0.
loop at connections into connection
where carrid= airline-carrid.
%>
<htmlb:treeNode
id= "<%= connection-connid %>"
text = "<%= connection-connid %>"
isOpen = "false" />
<%
endloop.
endif.
%>
</htmlb:treeNode>
<%
endloop.
%>
</htmlb:tree>
<htmlb:textViewtext   = "Hello World!"
design = "EMPHASIZED" />
<htmlb:buttononClick= "myClickHandler"
text= "Press Me"/>
</htmlb:form>
</htmlb:page>
</htmlb:content>
```

As we referred above, not only special, but unusual tags are also in the html like code: `<%`, `%>`, `<%= airline-carrname %>`. These and some other (not mentioned) ones are the so-called BSP-directives. Using these directives we can use direct ABAP scripting in the code. E.g. in the code above a loop is executed on the `airlines` internal table to define a tree node for each airline. The internal table declarations are not part of the code, but they are defined among the page attributes (see Fig. 7. below).

The data of the internal tables are not read in the code, but separately in an event handler (`OnInitialization`) as the Fig. 7. shows below.

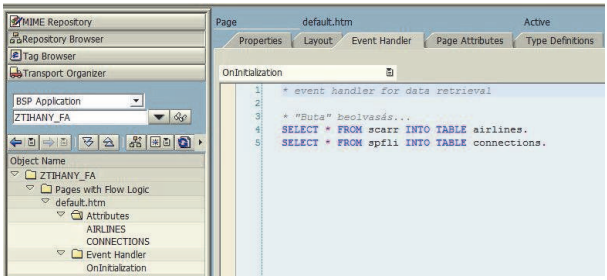


Figure 7. BSP development environment - events

There are more predefined events, which are triggered by the runtime environment except the input processing event, because this is triggered by the end-user action. As we have seen the BSP applications give much more capability, power and flexibility to the developer even with the standard pages with flow logic. In the later releases of BSP applications SAP introduced the model-view-controller design pattern. To define BSP application using the method SAP implemented views and controllers as shown on the Fig. 6. This kind of BSP applications are much more complex, but still gives the opportunity to define anything the end-user wants giving the capability for creating sustainable applications by separation of the user interface from the business logic.

Because of the ABAP scripting capability this Web tool has much higher interest than ITS ever (the old, experienced ABAP programmer can learn and program it easily) had. In the last releases SAP finished the development of standard Web transaction using BSP, but SAP still recommends using it for free-style Web programming for the customers.

SAP implemented first in J2EE Web AS and later in the ABAP Web AS as well the WebDynpro technology. The WebDynpro name come from the old, standard SAP screen name: dynpro. It uses the ModelView-Controller paradigm, and the definitions are stored as meta data and the system generates the corresponding code from this meta-data to create the real WebDynpro components. The WebDynpro components are reusable and according to the M-V-C paradigm they are sustainable because the look, the navigation, control and the business data management (model) are separated. The best way to keep the component unchanged during modification is to defin the model as methods of a class holding the state of the WebDynpro component. The methods are embedding tools for real services calls. A service call can be local real service, database extraction, managing files, or any remote service (RFC, BAPI from other SAP system, or Web service) consuming. WebDynpro cannot be used so freely as the BSP, because of the meta-data concept. Meta-data serves as the definition of the layout, screen elements, navigation, data exchange via hierarchical data store called context. The developer writes ABAP code only in special cases, like the real business logic takes place, or entry checks are needed. This makes easier to use different UI device for WebDynpro, because from the meta-data different, device dependent code can be generated by the SAP system. On the other hand it leads to restricted surface option, only pre-defined elements can be used and the outlook is also pre-defined (though if SAP NetWeaver Portal is used, the WebDynpro elements can take over the portal design for having same style).

The above Web application development techniques are code based ones. As the new requirements aroused SAP implemented a special, model-driven Web application technique as well. The WebDynpro applications give the opportunity to separate the service development from the real user interface design by using M-V-C design pattern, and meta data for element definitions. The SAP NetWeaver Visual Composer (NW VC) goes forward and gives a fully web based modeling tool for UI definition. By modeling we have an abstract layer of definition, so the result can be technology independent. With new technology only compilation must be done again. It increases the efficiency of the web application because the sustainability of it cost much less than with other techniques. NW VC implements WebDynpro and Adobe Flash capabilities as well. The modeling is lifted to a higher level here, because not only the developer can model, but also the business expert without development knowledge can create web base applications, like with ABAP Query any user can create a report in the classic ABAP environment. The design tool makes it possible to define data flow and event handling using business coherency. The service developer as in usual case defines the building blocks for NW VC and WebDynpros as well. In case of Visual Composer not only the reusable service blocks are predefined, but also the visual elements can be developed separately. This kind of programming model distinguishes between content admin, business expert, business application or service developer, and view developer (J2EE, ABAP or .NET). The main focus is on the business expert, because he is responsible for the business process, and he know it much better than any developer, so the developers just provide the view elements (like table, chart, etc.) and the data retrieval services, and the business experts puts these together by the modeling tool. It speeds up the development, makes the parts reusable and helps to modify the applications quickly according to the changing business needs.

6. How to Create Sustainable UIs

As we worked out the technical capabilities we can declare that the basis for creating sustainable backend functions for UI-s are the well defined RFCs and BAPIs, which can be called as WebServices as well. The main point of the sustainability in time when we have changes in the system is to have encapsulating, remote enable proxy services. These services can build the model level of the M-V-C paradigm to make a reusable and embedding layer for the UI-s. This makes possible not to change anything in the developed UI if any modification occurs behind (like upgrade, service modification), because the only the content of the embedding service should be modified if needed. This guarantees a more sustainable environment, because the end-user does note fill any changes on the UI, even if the back-end functionality is modified.

The opposite side also remains unmodified against the changes, if we consider a service or functionality on the back-end side stable, but we switch from a UI option to another one. As we learned above the almost all internal, standard SAP UIs can use any Workbench object, which provides functionality as model. But we can use remote-enabled function modules as well to implement model functionalities to guarantee the independence from the UIs (even external UIs). If we think about not RFCs, the standard Dynpro, ITS EWT, BSP or WebDynpro can be used as well. If we want to switch from one to the other the control and view layers should be redesigned, but the model can stay as it is. Using RFCs we have the opportunity to switch to ITS

FlowLogic or use separated Web surface, which can be an external one or another SAP Web AS based UI (like ABAP or Java WebDynpro). This changeability on UI side leads us beyond the UIs, because it tries to uncover the power of design and development methodologies advised for sustainable and maintainable solution in a changing environment [9-12]. These innovation for development strategies help us the generate better change management in a heterogeneous system landscape as well.

The Visual Composer would be a great tool but it needs NetWeaver Portal as basis element, which is not always offered at an implementation.

7. Conclusion

SAP and Microsoft ERP and other platform-based solution have many options to build a well-defined UI surface using difference data source and storage. Both solutions can provide services to consume by other application. With this property it is easy to create in a heterogeneous environment smoothly embedded new transactions, applications, which use services from other solutions for back-end functionalities. SAP can offer RFCs, BAPIs and Web Services for external use, so MS AX is a good candidate to profit from this offering for its developed UIs or UI extensions. The available RFCs and BAPIs services can be consumed only if the SAP NCo (SAP Connector for Microsoft .NET) is available for the developer. If no SAP connector is applied, Web Services (embedded RFCs and BAPIs as well) offered by SAP are consumable by AX.

SAP is not prepared for consuming services offered on any protocols. SAP can consume only provided Web Services and RFCs (and some other specialties not mentioned here). Definitely RFCs are provided mainly by SAP systems, but if a corresponding connector is in use, other applications, programs can also provide remotely callable functions for SAP. MS AX generally provides Web Services, which can be consumed by SAP applications based on SAP Web AS. But using the SAP NCo for Microsoft AX, not only Web Services, but also RFCs can be provided as well. When SAP consumes a Web Service or RFC provided by MS AX, different UIs, even standard ABAP Dnypro (screen based transaction) or WebDynpro can be based on them.

The interoperability and the service oriented new architectures and techniques make it easier to create application-wide transactions using UIs required. These UIs are maintainable and sustainable if the services behind are embedded to proxy classes or internal services.

References

[1] Sík-Lányi, C., Brown, J. D., Standen, P., Lewis, J., Butkute, V.: *Results of User Interface Evaluation of Serious Games for Students with Intellectual Disability*, Acta Polytechnica Hungarica vol. 9, no. 1, pp. 225-245, 2012

[2] Ősz, R., Pálmai, O.,: *A képernyő az interkulturális folyamatokban,* XI. Dunaújvárosi Nemzetközi Alkalmazott Nyelvészeti és Kommunikációs Konferencia, 2009

[3] Mátrai, R., Kosztyán, Zs.: *A New Method for the Characterization of the Perspicuity of User Interfaces,* Acta Polytechnica Hungarica vol. 9, no. 1, pp.139-156, 2012

[4] Barbaric, I.: *Design Patterns in Object-Oriented ABAP,* Galileo Press Bonn-Boston, 2010

[5] Heilman, R., Jung, T.: *Next Generation ABAP Development,* Galileo Press Bonn-Boston, 2007

[6] Muka; L., Lencse, G.: *Developing a meta-methodology for efficient simulation of infocommunication systems and related processes,* Infocommunications Journal, vol. LXIII, no. 7. pp. 9-14, 2008

[7] Muka, L., Lencse, G.: *Cooperating Modelling Methods for Performance Evaluation of Interconnected Infocommunication and Business Process Systems,* Proceedings of the 2008 European Simulation and Modelling Conference, 2008

[8] Muka, L., Benkő, B. K.: *Meta-level performance management of simulation: The problem context retrieval approach*, Periodica Polytechnica vol. 55, no. 1-2, pp. 53-64, 2011
DOI: 10.3311/pp.ee.2011-1-2.06

[9] Fodor, J., Ősz R.: *Possible applications of fuzzy methodology in the educational process,* IEEE 11th International Symposium on Applied Machine Intelligence and Informatics (SAMI), pp. 37-40, 2013
DOI: 10.1109/SAMI.2013.6480992

[10] Fodor, J., Ősz, R.: *Possible connecting areas of education and intelligent systems,* 2013 IEEE 9th International Conference on Computational Cybernetics (ICCC), pp. 51 – 56, 2013
DOI: 10.1109/ICCCyb.2013.6617560

[11] Ősz, R.: *Educational organization for new generation,* SAMI 2012, IEEE 10th Jubilee International Symposium on Applied Machine Intelligence and Informatics, Herl'any (Szlovákia) Conference Proceedings, pp. 373 – 375, 2012
DOI: 10.1109/SAMI.2012.6208992

[12] Ősz, R.*: New Technologies Mean New Methods of Learning?,* in Recent Advances in Modern Educational Technologies (editors: Hamido Fujita, Jun Sasaki), Iwate, WSEAS Press, ISBN:978-1-61804-180-7, pp. 59-64, 2013