

OPTIMIZATION OF SIGNAL TIMING BY EQUILIBRIUM TECHNIQUE

Ning WU

Institute for Traffic Engineering, Ruhr University, D-44780 Bochum, Germany;
PH: +49 (234) 3226557; FAX: +49 (234) 3214151; e-mail: ning.wu@rub.de;
<http://homepage.rub.de/ning.wu>

Abstract: Optimization of signal timing plans is normally based on operation research methods. In this paper, a new optimization technique is introduced. As a basic idea, the signal timing plan is considered as an analogy to a construction structure. Therefore, all of the basic concepts and methods of the structural analysis can be used for solving the optimal timing of the signal timing plans. The only difference compared to the conventional structural mechanics is the non-linear "elasticity" of the signal timing plans. As an objective function of the optimization, either the sum of delays or the fuel consumption or other variables like air pollutants can be used. For the solution of this equilibrium problem, algorithms from structural engineering can be applied. These algorithms are very efficient in term of computation converge very quickly. Therefore, rather complex intersection structures can be optimised with quite short computer times. Also an on-line rolling optimization of signal timing plans with measured traffic flows can easily be realised. This technique is proven to be very effective for practical applications. The new method is capable of extending the optimization from single intersections (including optimization of the cycle time) to network of traffic signals and to traffic-actuated controllers.

Keywords: *Signalized intersection, Optimization of signal timing plans, delays at traffic signals*

1 Introduction

Optimization of signal timing plans is a frequently discussed topic in traffic engineering. Optimization means, above all, minimization of the sum of delays. Webster was the first one to formulate an analytical solution for minimization of the sum of delays for a two-stage traffic signal in 1958 (Webster, 1958). Using some simplifications, he was able to give the optimum cycle time and the optimum stage green times. Since then, numerous papers have been published on theoretical analyses and practical applications of signal timing optimizations. In the field of theoretical researches, publications from Allsop (1971, 1992) and Tully (1976) should be mentioned at the first place. Concerning the practical applications of signal timing optimization, different models and computer programs were developed. The most among them are e.g. the AMPEL program package in German-speaking countries and the SIGSIGN package (Silcock and Sang, 1990) from Great Britain.

The existing procedures for signal timing optimization can be distinguished into two groups: stage-oriented optimization and signal-group-oriented optimization. Most of the previous authors first treated the stage-oriented optimization due to its simple applicability. This simple applicability was necessary and made sense, since traffic signals were stage-oriented themselves until the seventies. Due to development of modern signal controllers, signal groups can nowadays be controlled individually. Thus, an optimization which takes into account the green time and traffic characteristics of each individual signalized movement at the intersection, becomes necessary.

Most of the existing models for signal-group-oriented optimization were based on the theory of "Operations Research" (OR method) or on "Trial and Error" methods. Since the maximization of capacity is a linear optimization problem, it can easily be handled with OR methods. The minimization of the sum of delays, however, is a non-linear problem. Here, most of the current methods are based on numeric solutions. These numeric solutions are normally quite time-consuming and can, therefore, only be used for off-line calculations of signal timing plans.

This paper presents a new optimization model in which the signal timing plan is treated as a mechanical system and its optimization is achieved by determining the state of equilibrium of the mechanical system. The optimization, i.e. the state of equilibrium, is determined using the so-called moment distribution method from Cross (cf. Beaufait, 1972) in classical structure mechanics. In this mechanism the connections (connection between two or more signal groups) of signal groups (=constraints of the structure) are used as the variables for the optimization in place of the green times. Thus, the free degrees of the system can be drastically reduced. Since this procedure is very efficient (because of less free degrees) and since all characteristic information of the signal timing plan is passed on with every further step of optimization, this optimization procedure is particularly suitable for a dynamic optimization, i.e. for an on-line optimization (adaptive control) of signal timing plans. The principle of the new model can also be transferred to optimizing large scaled, coordinated networks.

This paper is the partial result of a project sponsored by the German Research Community for a two year period.

2 The Signal Timing Plan as a Mechanical System

In the theory of elasto-mechanics (cf. Lehmann 1979), Dirichlet formulated the theorem on the minimum of total potential in stable equilibrium:

If a system is in a state of stable equilibrium, the total potential (or total potential energy) of this system has its minimum value

Normally, one of several possible local minima of the total potential in the mechanical system can be determined if the corresponding state of equilibrium is ascertained. If one compares the objective function (e.g. sum of delays) for the

optimization of a signal timing plan with the total potential of a mechanical system, the minimization of the objective function can be achieved by transferring the signal timing plan into the "state of equilibrium".

Figure 1 shows a basic element of a signal timing plan and a basic element of a mechanical spring system. It demonstrates that, in analogy to mechanics, the green time can be described as an elastic spring.

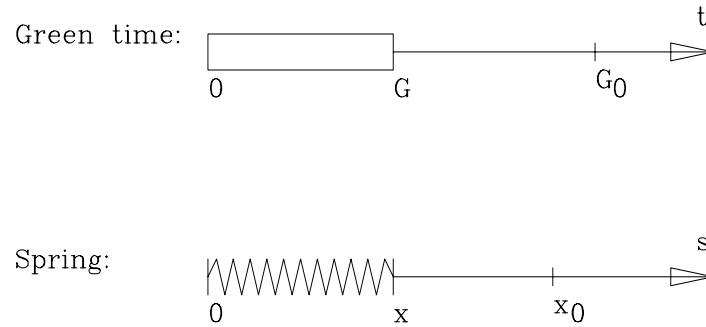


Figure 1. Comparison between a green time and a spring

Denote B the demand for more green time in a signal group. With increasing green time, the capacity is increased and the delay W is reduced. Therefore, B can be defined as the conservative force of the potential function W . Thus, B is the derivative of W towards G , i.e. $B=W'(G)$.

Therefore, the following analogies between the two elements can be established:

Green time		Spring system	
Duration of the green time	G	Length of the spring	x
Cycle time	C	Length in original state	l
Traffic volume	q	Specific elasticity	k
Delay	W	Potential	U
Green time demand	B	Spring force	F
Slope of B	M	Stiffness coefficient	K
Point of reference	G_0	Point of reference	x_0

G_0 or x_0 are the points of reference, at which the potential (W or U) is defined as zero. They are chosen so that the force F and the "green time demand" B is zero. That is,

$$G_0 = C \quad (\text{or } G_0 = \infty) \qquad x_0 = l$$

It is true

$$\begin{aligned} B &= f(q, G, C) & F &= f(k, x, l) \\ W &= f(q, G, C) & U &= f(k, x, l) \\ M &= f(q, G, C) & K &= f(k, x, l) \end{aligned}$$

where f is the abbreviation for "function". It can be seen that

$$B = \frac{dW}{dG} \qquad F = \frac{dU}{dx} \qquad (1a)$$

$$M = \frac{d^2W}{dG^2} \qquad K = \frac{d^2U}{dx^2} \qquad (1b)$$

or

$$W = \int_{G_0}^G B \cdot dt = - \int_G^{G_0} B \cdot dt \qquad U = \int_{x_0}^x F \cdot ds = - \int_x^{x_0} F \cdot ds \qquad (2)$$

$$B = \int_G^G M \cdot dt = - \int_G^{G_0} M \cdot dt \qquad F = \int_{x_0}^x K \cdot ds = - \int_x^{x_0} K \cdot ds \qquad (2b)$$

With the Eq. (1) and (2) we are able to transform a problem of investigating the maximum (or minimum) of the potential function (objective function) into a problem of investigating the equilibrium of the corresponding virtual forces and vice versa. Under some conditions we want to investigate the maximum (or minimum) of the potential function for achieving the equilibrium of the virtual forces and under other conditions we want to investigate the equilibrium of the virtual forces for achieving the maximum (or minimum) of the potential function.

Figure 2 schematically shows the typical curves of the delay W and the green time demand B of a signal group in analogy to the potential U and the spring force F of a spring system. Both have the same characteristics as following:

- (1) B and F are monotonically falling, i.e.: $f'(G) < 0$
- (2) W and U are strictly convex, i.e.: $f''(x) > 0$

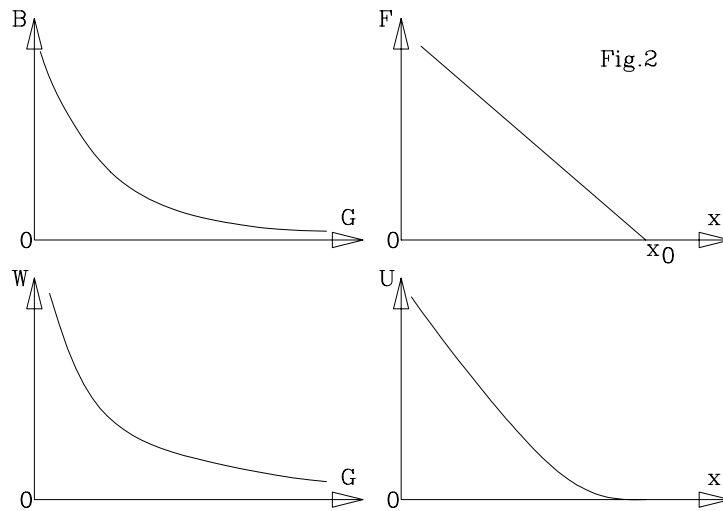


Figure 2. Characteristics of a green time and a spring

The second characteristic is sufficient but not necessary for the existence and uniqueness of an optimum solution and for the convergence of the optimization procedure.

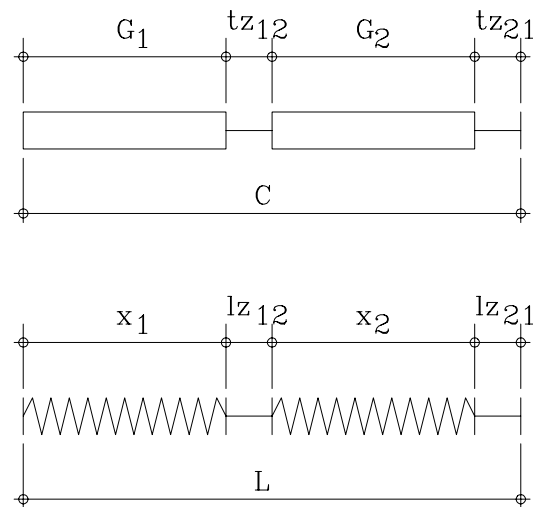


Figure 3. Signal timing plan with two incompatible and competing signal groups

According to the analogy defined above, every signal timing plan can be considered as a mechanical system of springs. The springs can be linked in a serial way, one after another, or in a parallel way, side by side. The arrangement of such spring systems (analogy to signal groups in a signal timing plan) can be defined by the sequence of stages and by the restrictions (intergreens etc.)

The simplest example in this analogy is a signal timing plan with two incompatible and competing signal groups. Figure 3 shows this example. Here G_1 and G_2 are the green times of the two signal groups, $t_{z_{12}}$ and $t_{z_{21}}$ are the intergreens between the signal groups, and C is the cycle time. Respectively, the spring system can be described by x_1, x_2 (length of the springs), l_{12}, l_{21} (inflexible connections between the springs) and L (total length of the spring system).

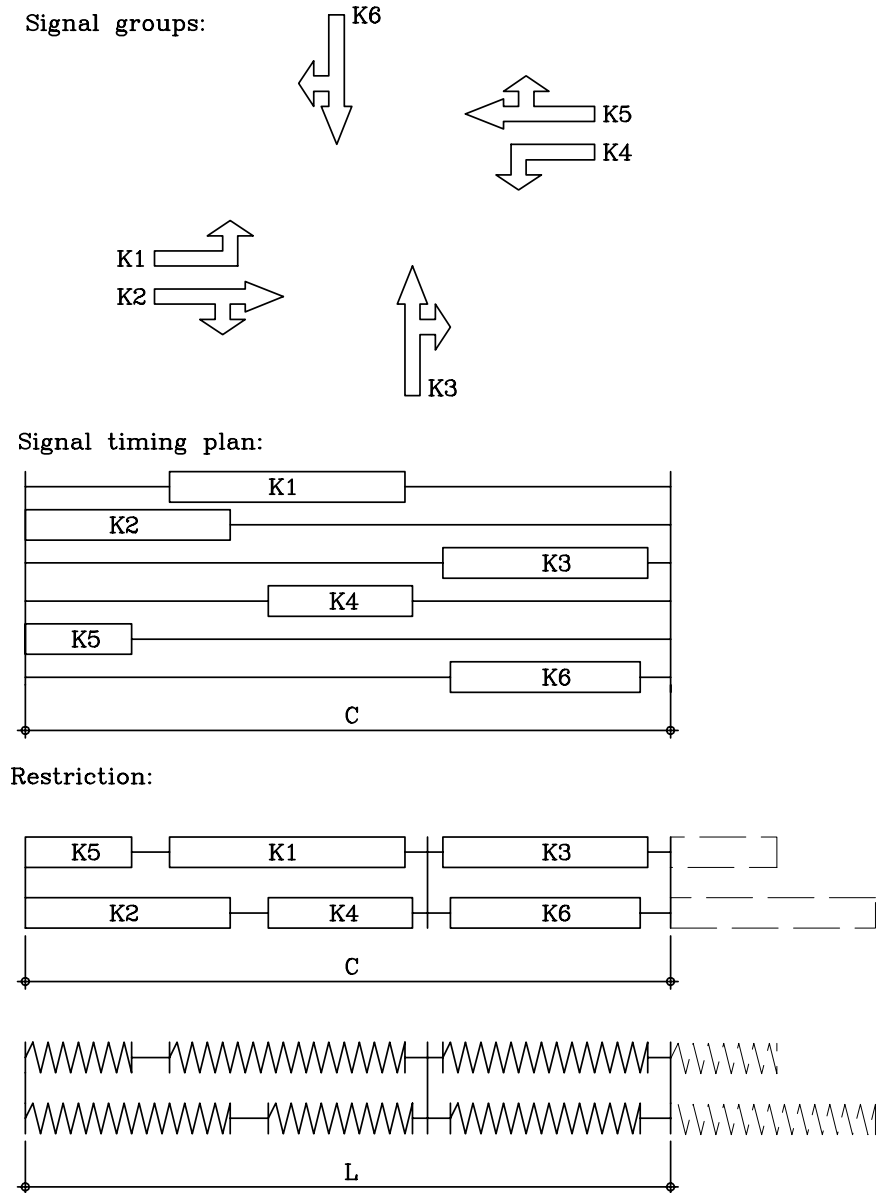


Figure 4. Simplified intersection with 6 signal groups

Figure 4 depicts a simplified intersection with 6 signal groups as a realistic example. The signal timing plan shows that the signal groups in a stage have not always the same length. For example, the signal group K1 has extended beyond the stage area and has displaced the green time of signal group K5. However, the signal timing plan is defined by the restriction of intergreens. Due to those restrictions, this signal plan can be described as a mechanical system with 6 springs.

This spring system contains serial as well as parallel connections between the springs (signal groups). The force of the springs makes the mechanical system strive for the equilibrium by itself, and thus achieves the minimum total potential (sum of delays). The task of optimizing the signal timing plan has therefore become a task of determining the state of equilibrium of the corresponding mechanical system.

The determination of the state of equilibrium requires the knowledge of the given structure of the signal plan. The optimum signal timing plan can be found by checking all possible signal plan structures and finding out the optimum structures (stages and stage orders) with respect to the objective function. The possible signal timing plan structures can be listed e.g. according to the procedure by Tully (1976). In this paper, this procedure is not discussed in further detail. The structure of the signal timing plan is considered as predefined.

It can be proven mathematically that the objective function (the potential function U or the delay function W) has exact one minimum, if it is strictly convex in the considered interval (Allsop, 1992). Since the functions which are defined as the potential (delay etc.) during the optimization of the signal timing plan fulfil this condition, it can be assumed that a unique solution for the optimization of the signal timing plan can be found. Also by some non-convex functions the unique optimization can be found (Wu, 1999).

3 Objective Functions and Equilibrium Conditions for Minimization of the Sum of delays

With given traffic volumes and boundary conditions, the objective of the optimization is to obtain the cycle time and the corresponding green times by minimizing the sum of delays over all signal groups. The function for the delay calculation is used as the potential function. The objective function for the minimization of the sum of delays is

$$U_g = \sum_{i=1}^n W(q_i, G_i, C) \quad (3)$$

The force function of the signal group F_i (=green time demand B_i), of which the equilibrium is to be found, is

$$F_i = \frac{\partial W(q_i, G_i, C)_i}{\partial G_i} \quad (3b)$$

The stiffness coefficient of the signal group $K_i(=M_i)$ is respectively

$$K_i = \frac{\partial B_i}{\partial G_i} = \frac{\partial^2 W(q_i, G_i, C)_i}{\partial G_i^2} \quad (3c)$$

The parameters in Eq. (3) are

- n = number of signal groups
- G_i = green time of the signal group i
- C = cycle time
- $W(q_i, G_i, C)$ = $w(q_i, G_i, C) \cdot q_i$
= delay of signal group i
- $w(q_i, G_i, C)$ = average delay per vehicle of signal group i
- q_i = traffic volume of signal group i

The boundary conditions for the optimization are

- a) $G_i > G_{i,\min}$
- b) $C_{\min} < C < C_{\max}$

with

- $G_{i,\min}$ = predefined minimum green time for signal group i
- $C_{\min}$ = predefined minimum cycle time
- $C_{\max}$ = predefined maximum cycle time

and in case of stationary traffic

- c) $s_i \cdot G_i > q_i \cdot C$

with s_i = saturation flow of signal group i

If a delay formula defined for temporary over-saturations (e.g. Akcelik, 1980 or Wu, 1990) is used, the boundary condition c) can be omitted.

Furthermore,

- d) all boundary conditions which are necessary to ensure safe traffic operations, e.g. intergreens, restricted overlap of green time for permitted left turns, parallel pedestrian crosswalks with permitted right or left turners etc. (for German conditions cf. FGSV, 1992, 2003)

must be hold.

All common delay formulae (Webster, 1958; Miller, 1968; Akcelik, 1980; Kimber and Hollis, 1979; Wu, 1990 etc.) for $W(q_i, G_i, C)$ fulfil the convex condition over the interval $(0, C)$ because

$$\begin{aligned}
(1) \quad & \frac{\partial W_i(q_i, G_i, C)}{\partial G_i} < 0 \\
(2) \quad & \frac{\partial^2 W_i(q_i, G_i, C)}{\partial G_i^2} > 0 \\
(3) \quad & \frac{\partial^2 W_i(q_i, G_i, C)}{\partial C^2} > 0 .
\end{aligned}$$

The optimization of the objective function (Eq. (3)) can be carried out for the optimum distribution of green times $G_{i,opt}$ with a fixed cycle time C . In addition, also the cycle time C can be optimized.

4 Determination of the Equilibrium for the Optimization Procedure

According to the principle of virtual work from Dirichlet (cf. Lehmann, 1979)

A mechanical system is in its equilibrium only if the virtual change of the total potential for any virtual displacements is equal to zero, that is

$$\delta U_g = \sum \delta U_i = 0$$

For a mechanical system with n degrees of freedom, Dirichlet's theorem corresponds to

$$\frac{\partial U_g(x_i)}{\partial x_i} = 0 \quad i = 1 \text{ to } n \quad (7)$$

with x_i = coordinates of the i -th degree of freedom

Eq. (7) is nothing else but the condition of equilibrium of the conservative forces. If all forces F_i of this mechanical system are proportional to the coordinate x_i - as in a real spring system - a solution for the equation system Eq. (7) can be found for every force F_i and all coordinates. Unfortunately, the virtual forces (green time demand B_i) of a signal timing plan with Eq. (3), (4) and (5) as the potential function (objective) do not fulfil this condition of linearity. Most of the delay formulae are functions of higher orders. Some even contain transcendental functions. An analytical solution of Eq. (7) is therefore very difficult or even impossible to find. Up to now, optimization procedures with the sum of delays as the objective function have been carried out only numerically with an enormous calculation effort.

The state of equilibrium of a mechanical system can also be ascertained iteratively. One of the well-known procedures is the moment distribution method from Cross (cf. Beaufait, 1972) which is based on the stiffness method. This procedure can be explained by Figure 5a.

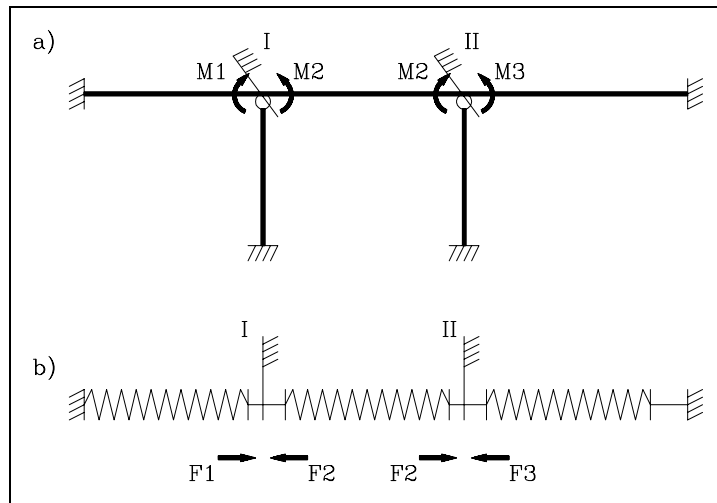


Figure 5. Principle of the moment and force distribution method

Figure 5a shows the structure of a bridge. The bending moments of the beam near the columns are to be determined. The degrees of freedom of this mechanical system are the slope deflections of the beam at both columns. First, these two slope deflections are fastened in their original state by fictitious constraints. Putting weight on the beam causes a surplus of bending moments ($M_2 - M_1$, $M_3 - M_2$), which is fictitious as well. The bending moments are in a state of imbalance. To determine the state of equilibrium, the constraints are released alternately, i.e.:

- (1) Constraint *I* is released first while constraint *II* stays fastened. Constraint *I* turns as an effect of the surplus of the bending moments $M_2 - M_1$. When this surplus is distributed, the bending moments at constraint *I* have achieved the state of equilibrium again. The constraint *I* is fastened again. As a consequence of the turning movement at constraint *I*, constraint *II* gets a new surplus of bending moments.
- (2) Then constraint *II* is released while constraint *I* stays fastened. Constraint *II* turns as an effect of the surplus of the bending moments $M_3 - M_2$. When this surplus is distributed, the bending moments at constraint *II* achieve their state of equilibrium. Constraint *II* is fastened again. As a consequence of the turning movement of constraint *II*, constraint *I* gets a new surplus of bending moments.
- (3) Steps 1 and 2 are repeated.

The bending moments at the constraints are alternately set into the state of equilibrium. The surplus of the bending moments decreases as the number of iterations increases. The iteration is stopped when the surplus of the bending moments has become small enough for practical application at all constraints. This method is also called the moment-distribution procedure. It can be used for a system with any number of constraints. All constraints are repeatedly released and fastened until all of them have achieved the state of equilibrium.

The state of equilibrium of a mechanical spring system can be determined analogously (cf. Figure 5b). Instead of the surplus of the bending moments M_i , the surplus of the spring forces at the connections F_i is distributed. The connections are first fastened and then released alternately (which causes horizontal displacements). Then they are fastened until the surplus of the spring forces (F_2-F_1 , F_3-F_2) falls below a certain minimum at the connections. This procedure can be called the force-distribution procedure.

The force-distribution procedure does not start with the determination of the state of equilibrium for the total system, but with that of the individual constraints. In a signal timing plan as an analogy to a system of springs, the constraints are clearly defined by the restrictions of the signal timing plan (intergreens, permitted left turners or minimum green times etc.). Thus, a system with n springs (analogue to n signal groups) is simplified as a system with m constraints. In most of cases, m is always smaller than n . The example in Figure 4 has $n = 6$ signal groups, but only $m = 3$ constraints if the cycle time is fixed. If the cycle time C is regarded as a variable as well, this example has $m = 4$ constraints.

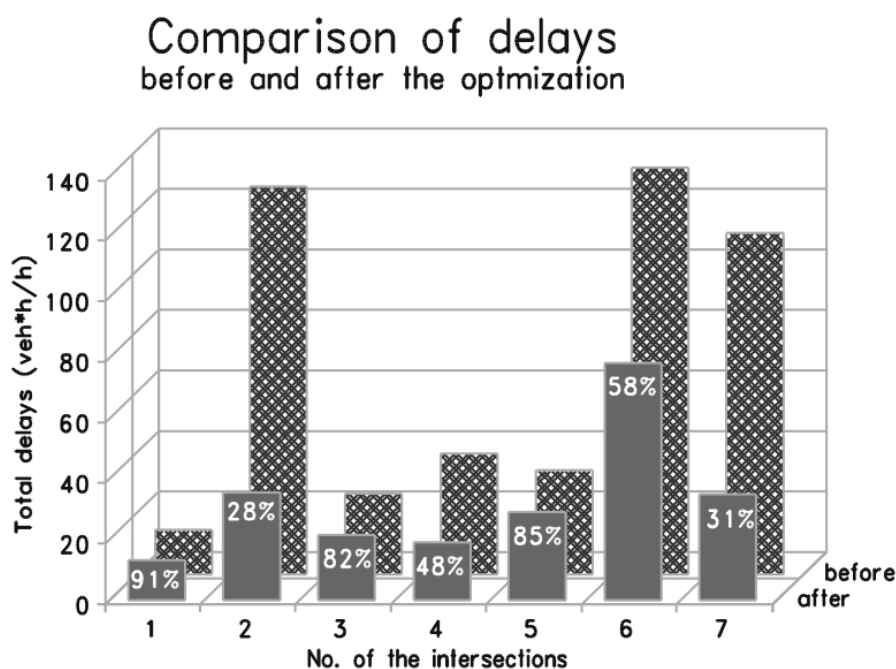


Figure 6. Optimization results for 7 intersections

The new procedure was tested at 7 intersections in Düsseldorf, Germany. The local authorities provided the existing signal timing plans. The existing signal times were used as a basis for the optimization. The result is presented in Figure 6. It

shows that the signal timing plans, some of which were several years old and for which the traffic volumes have undoubtedly changed in the meantime, need to be improved. The intersections 2 and 7 were controlled by so-called advanced signals. Here, the green times of some movements were intentionally restricted to limited capacity ($\text{capacity} < \text{demand}$) with the purpose of keeping congestion out of the city area. Therefore, the before/after comparison is irrelevant for these two intersections. For the other 5 intersections, improvements could be achieved. The significant improvements at intersections 4 and 6 are results of overloaded left turning traffic streams.

5 Conclusion and outlook

The optimization procedure for a signal timing plan according to the principle of equilibrium has the following advantages compared to the existing procedures:

- Fast computation due to reduction of variables (in place of number of signal groups n the number of constraints m is used). An optimization procedure for a fixed-time controlled standard intersection normally takes less than 1s.
- Minimum work before the optimization (only the signal timing plan and the corresponding intergreens and the traffic volumes are needed).
- Applicable for any convex objective functions.
- Continuity of the signal timing plan during the optimization (the signal timing plan is not produced totally anew but is only developed in an innovative way).
- Dynamic character (adaptation to changed traffic volumes, minimum green times etc. is possible during optimization).
- The procedure can be influenced manually (new definition of green times, reduction of cycle time etc.).

Regarding the general theoretical background, this procedure can also be easily transferred to the following traffic control strategies:

- Signal timing control with consideration of pedestrian delays
- Signal timing control with special stages for public transport
- Adaptive signal timing control
- Coordinated signal timing control in networks

In the future, the new model will be formulated in details.

References

- Akcelik, R. (1980). Time-dependent expressions for delay, stop rate and queue length at traffic signals. Australian Road Research Board, No. 361-1, 10/1980.
- Allsop, R.E. (1971). Delay-minimising settings for fixed-time traffic signals at a single road junction. Journal of the Institute for Mathematics and its Applications, 8 (2), 164-185.

- Allsop, R.E. (1992). Evolving Application of Mathematical Optimisation in Design and Operation of Individual Signal-Controlled Road Junctions. In: Mathematics in Transport Planning and Control, edited by J.D. Griffith, Clarendon Press, Oxford.
- Beaufait, F.W. (1972). Basic Concepts of Structural Analysis. Prentice-Hall, Inc. Englewood Cliffs, New Jersey 07632.
- Brilon, W. and Wu, N. (1990). Delays at fixed-time traffic signals under time-dependent traffic conditions. Traffic Engineering and Control. 12/1990.
- FGSV (1992, 2003). RiLSA - Guidelines for Traffic Signals - English Version of with minor Modifications. FGSV-Verlag, Köln.
- Kimber, R.M. and Hollis, E.M. (1979). Traffic queue and delays at road junctions. TRRL Laboratory Report, LR 909.
- Lehmann, Th. (1979). Elemente der Mechanik IV: Schwingungen, Variationsprinzip. Friedr. Vieweg & Sohn, Braunschweig/Wiesbaden 1979.
- Miller, A.J. (1968). Signalised intersections - capacity quite. Australian Research Board Bull, No. 4/1968. Reprinted as ARRB Research Report ARR No. 79/1978.
- Möller, K. (1987). Signalgruppenorientiertes Modell zur Optimierung von Festzeitprogrammen an Einzelknotenpunkten. Schriftenreihe des Instituts für Verkehrswesen der Universität Karlsruhe, Heft 37.
- Rose, R. (1972). Foundations of Mathematical Biology. ACADEMIC PRESS, New York and London.
- Silcock, J.P.; Sang, A. (1990). SIGSIGN: a phase-based optimisation program for individual signal-controlled junctions. Traffic Engineering and Control, 5/1990.
- Tully, I.M.S.N.Z. (1976). Synthesis of sequences for traffic signal controllers using techniques of the theory of graphs. OUEL Report 1189/77, University of Oxford.
- Webster, F.V. (1958), Traffic signal settings. RRL Technical Paper No. 39, HMSO, London.
- Wu, N. (1990). Wartezeit und Leistungsfähigkeit von Lichtsignalanlagen unter Berücksichtigung von Instationarität und Teilgebundenheit des Verkehrs (Delay and Capacity at Signalised Intersections under non-stationary and partly-bunched traffic conditions). Schriftenreihe des Lehrstuhls für Verkehrswesen der Ruhr-Universität, Heft 8 (A short English version is given in Brilon, Wu, 1990).
- Wu, N. (1999). Optimierung von Signalzeitenplänen nach dem Gleichgewichtsprinzip (Optimization of Signal Timing by Equilibrium Technique). Schriftenreihe des Lehrstuhls für Verkehrswesen, Heft 20, Ruhr-Universität Bochum.