27th ITS World Congress, Hamburg, Germany, 11-15 October 2021

Paper ID #551

MOZART
Traffic management through traffic signal control by Quantum-Inspired optimization

Fritz Schinkel[1*], Isabel Schwende[1], Rando Schade[2], Eva Cerny[3], Martin Fellendorf[3]

1. Fujitsu, Germany, fritz.schinkel@fujitsu.com

2. Hamburg Port Authority AöR, Germany

3. Graz University of Technology, Austria

Abstract

Road traffic in cities such as Hamburg is controlled by traffic actuated signals, which, however, do not adapt sufficiently well to highly variable traffic situations. The performance of traffic signals, and hence the whole traffic flow, depends on the quality of their rule-based control logics. Despite extensive efforts by traffic engineers to design each traffic control logic individually, no overall network optimization is performed. The Hamburg Port Authority (HPA) sought new technical ways to improve traffic control throughout the port area by starting to implement a comprehensive traffic management system incorporating new sensor technology. Responding to all detected traffic situations leads to a huge combination of measures over the whole network. With quantum-inspired optimization, HPA is taking a novel approach to continuously finding optimal solutions to problems with a large number of possible scenarios - all within seconds from sensor reading to actuating the signal sequence. This makes it the world's first real-time road network control using quantum-inspired technology.

Keywords:

1. Quantum computing
2. Real-time traffic flow optimization

Introduction

The Hamburg Port Authority (HPA) initiated a traffic innovation project called MOZART to improve the traffic flow throughout the entire port region. This road network faces the challenges of a steady increase in heavy traffic of goods that are transported on trucks at unpredictable temporal patterns as well as commuter traffic by passenger cars. The road network cannot be expanded due to cost reasons, limited space, and environmental considerations. Therefore, the existing network must be utilized more efficiently by means of traffic control as part of the port's new traffic management system.

Currently, the majority of the 30 traffic signals in the harbor area are operated in an actuated mode (traffic responsive); some are fixed time with various pretimed signal settings depending on time of day. According to the German guidelines RiLSA [4] the traffic responsive control units contain a rule-based implementation at each signalized intersection. Signal timings are computed at each intersection individually considering traffic measured by inductive loops or infrared detectors. The control logic has the ability to extend green times, skip a phase in case of the lack of demand, or alter the phase sequence. In practice, phase skipping is not used, especially not at coordinated intersections in the harbor.

Coordinated traffic light circuits often don’t work as desired and only coordinate the main direction ("green wave"). In the case of coordination along an arterial, the permitted green times are limited to time frames within a cycle. The planning process of the rule-based control logics has to be performed by experienced traffic engineers. Nevertheless, on changing conditions the result is not always satisfactorily. This planning method does not optimize traffic flow in a mathematical sense according to optimization criteria such as overall delay or travel time. Adaptive traffic control systems such as SCATS [8] and SCOOT [5] are based on traffic models and optimize traffic according to performance indicators such as delay and number of stops. However, these adaptive control systems are not implemented in Germany due to limited support for public transport priority. The complexity of adaptive systems and requirements on sensing technology are further drawbacks limiting the introduction of adaptive control systems in Germany. Therefore, the port authority HPA was searching for new technology optimizing the particular port traffic.

HPA has updated all traffic controllers to the most recent OCIT-standard. The network is being equipped with a dense mesh of thermal imaging cameras, induction loops and Bluetooth sensors. In addition to the existing presence detectors for actuated signal control, detectors are installed at major cross-sections along the roads counting cars and trucks separately. Bluetooth sensors at strategic locations in the network trace detected Bluetooth devices and monitor travel times along individual routes. Thermal imaging cameras supplement the sensor system to provide traffic-monitoring across the entire harbor road network. The variety of different traffic sensors monitoring the current traffic state and the need to optimize the traffic flow holistically by traffic signal control in real-time requires a new approach of computing. The MOZART initiative is in-line with other international initiatives to intensify the use of modern computing technology in traffic signal control [13].

Traffic light optimization Concept

This solution intends to identify the most suitable signal timings to maximize traffic throughput in total. Minimization of travel time is taken as optimization indicator while considering numerous constraints imposed by the guidelines on traffic signals [4]. Besides travel time minimization, which corresponds strongly with delay minimization, other criteria such as environmental issues can be addressed. The aim is not only to reduce standstill times with the engine running, but also to avoid harmful emissions - from unnecessary acceleration and braking indirectly through the design of the optimization problem. Traffic safety is also improved by minimizing stops and consequently the probability of rear-end collisions.

To achieve this, no more pre-planned signal programs are to be used, but only the safety-relevant rules. Red and green time distributions are selected freely. This means that signal controllers no longer execute fixed programs, but receive phase durations and sequences from a central command in a very short frequency. However, this also means that the optimal switching times, green and red times must be found for each traffic light from a huge number of possible combinations. In addition, it must be possible to take the large number of existing heterogeneous sources into account and process them for real-time traffic light control. Due to the importance of highly variable truck traffic within the port area, truck platoons are prioritized by this solution to meet the global optimization goal. The unique advantage of overall network optimization is that not only the arriving vehicles per node are prioritized, but also the effects on the following nodes are considered. This allows emerging truck platoons to travel through the network without stopping.

To achieve the challenging requirements, a new concept for traffic modeling and computational optimization was realized in the MOZART project by making use of classical high-performance computing (HPC) and the new quantum inspired Digital Annealer technology (DA).

For a single directed road between two intersections, the influx and efflux of vehicles depend directly on the green times of the respective traffic lights. It is possible to shrink and stretch the length of green phases, drop certain phases, or even change the order of phases stitched together by approved transitions. Starting at a detected traffic situation for each combination of alternative phase sequences and durations, the traffic on the road is predicted. Based on this the quality contribution e.g. in terms of mean velocity is calculated. This is the classical part of the hybrid concept. Assuming a number of p alternative phase sequences / durations for each intersection and a network with n intersections and $4n$ directed roads results in $4p^2n$ predictions. The compute effort for this task grows linearly with the number of intersections and can easily be executed in parallel on a classical HPC cluster.

These quality contributions are used to construct a global quality function f, which calculates the traffic quality of the whole network for each combination of phase sequences / durations. Computation of single values of f is fast but it is hard to test all combinations in order to find the optimum, since the number of combinations is p^n and grows exponentially with the number of intersections. The harbor network with 30 intersections yields $10^{30} = 1{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000$ possibilities. Hence this quality function is implemented on non-classical hardware. A quantum computer, more precisely a quantum annealer, can find the optimum within seconds. Since currently available quantum machines support only a small number of intersections, we use the Digital Annealer from Fujitsu. This technology finds very good solutions in seconds and the DA can support networks with $n = 800$ intersections. This easily covers the harbor of Hamburg but is suitable for real-time traffic light optimization of the entire metropolitan region as well.

The variety of optimization criteria and the large set of feasible but not optimal solutions in traffic signal setting across multiple intersections requires a large amount of computing power. Classical high-performance computing (HPC) can accelerate computation by massive parallel processing, but it fails to solve the non-parallelizable combinatorial optimization for bigger networks, which is a main challenge in NP-hard problems. With the Digital Annealer, complex calculations are supported for a high number of possible scenarios in just a few seconds and hence enable operations in real-time. This allows the optimization to be applied onto the entire traffic system, resulting in a holistic solution that can map temporary fluctuations like planned events and local peculiarities in the traffic flows.

The innovative control concept is realized as a Quantum-Inspired Optimization Service (QIOS) with a

Digital Annealer component. It is comprising the currently implemented simulation and the envisioned operational real-time traffic flow optimization.

Figure 1 shows the high-level architecture of the current test system and the envisioned operative version forthe solution. As central element the traffic light optimizer (TLO) realizes the concept outlined above. The detailed Static Road Map is imported to the TLO and the data is foundation for the Digital Twin to enable realistic traffic simulations. Based on this, the Local Scenario Prediction simulates in traffic on directed roads for alternative phase sequence / durations of the

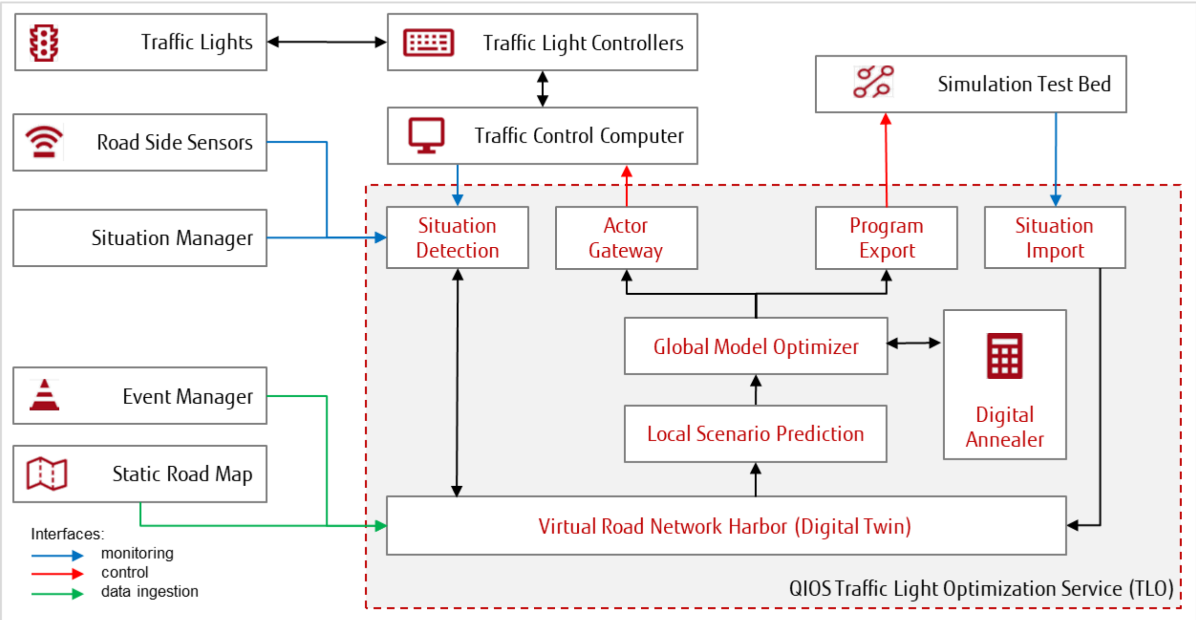

Figure 1– High level architecture of QIOS service for traffic light optimization

traffic lights at the influx and efflux intersections. The Global Model Optimizer creates the global target function and solves the optimization task on the Digital Annealer hardware. Details about the mathematical model are outlined in the chapter "TLO - Applying Quadratic Optimization for Traffic Light Control". The system uses a Simulation Testbed as counterpart for an evaluation under realistic traffic conditions and dynamics. This testbed is implemented on top of the software PTV Vissim [3] as described in the chapter "Simulation Testbed". Situation Import reads vehicle types, positions and velocities as starting point for the optimization from the testbed. After optimization, the selected phase sequence and durations for all intersections are applied to the simulation via Program Export.

The architecture developed as part of the MOZART project is designed to enable the seamless collaboration of sensors, traffic control computers at the traffic management center, the traffic light optimization service, and traffic light controllers to implement an operational traffic light control system. Additional components provide the necessary information about the current traffic and communicate the optimization results to the TLO control system. The Situation Detection derives an estimated traffic situation as vehicle positions and velocities from the event history and most recent information from the Road Side Sensors. In addition, real-time information on the general traffic situation and status updates from the traffic control computer are processed by the

`Situation Detection` for usage within the Digital Twin of the road network and traffic situation. It is also envisioned to include information on planned events such as construction sites from an event manager component. This input data is equivalent to the simulation testbed data and is used by the TLO to determine globally best traffic light control decisions. Once an optimization computation phase is complete, the selected best phase sequences and durations are communicated via `Actuator Gateway` to the `Traffic Control Computer`, which then updates the `Traffic Light Controllers` in a coordinated manner.

The concept of this software is designed so that additional sensors can be added with ease by using a flexible data model and potentially a local data hub for data management and analysis. Another important consideration for the design is openness for future extension into service ecosystems such as IDS and GAIA X, as such a highly complex software solution can only be developed following an agile approach by cross-functional teams in collaboration with partners in the relevant ecosystems.

As the MOZART project is in a proof-of-concept phase, it is currently geographically restricted to a small area within the authority of the Hamburg Port Authority. In order to fully leverage the computational power of the Digital Annealer at its core, the software solution is built in a scalable manner to be extended to a much larger geographical area, potentially operating traffic flow optimization for hundreds of traffic lights in real-time.

Quantum Computing and Digital Annealer

The exponential increase of compute power over the last 60 years is approaching a barrier: physical laws limit miniaturization of chip structures as it reaches the atomic scale. Further improvements require new approaches like specialized, purpose-built hardware [1]. Quantum computers are considered a breakthrough technology to provide a significant leap for more computational power and overcome the runtime issues of many NP-hard problems. This is the disruptive improvement that is necessary to realize complex optimizations within the short time span between sensing and predicting the traffic and the rollout of the globally best control sequences to all the traffic light controllers in the network.

Problems like risk mitigation, cost reduction, task scheduling or– as in this work – traffic management are combinatorial optimization tasks. These problems often turn out to be NP-hard, thus economically reasonable solutions on classical hardware are out of reach. Many of these NP-hard problems can be mapped to energy minimization of Ising models [9]. Quantum Annealers can solve such Ising models in seconds. The theoretical concept by Kadowaki and Nishimori [3] was realized and commercialized as Quantum Annealer by D-Wave ever since 2011 [2]. D-Wave machines can solve Ising models, but the size of precise models is still small. This limits the application of Quantum Annealing in operative scenarios today.

In optimization applications such as traffic management, good solutions and fast execution is more important than the guaranteed global optimality of the solution. For those applications, the Digital Annealer, a purpose- built machine by Fujitsu [12], is a good choice. It handles the same problem class as Quantum Annealers, but it supports fully connected high precision Ising models with up to 8192 bits.

The input for Quantum Annealing and Digital Annealing can be specified as a polynomial E of quadratic order in binary variables x_i. It is called Quadratic Unconstrained Binary Optimization (QUBO), which contains all information about the optimization problem. For a search space $X = \{0,1\}^N$ with a natural dimension N a real valued energy function $E: X \rightarrow \mathbb{R}$ is defined as

$$E(x_1, \dots, x_N) = \sum_{i=1}^{N} b_i x_i + \sum_{i=1}^{N} \sum_{j=i+1}^{N} w_{i,j} x_i x_j \qquad \textbf{(1)}$$

for a real valued bias vector

$$\boldsymbol{b} = (b_1, \dots, b_N) \in \mathbb{R}^N \qquad \textbf{(2)}$$

and an upper triangular weight matrix with zero diagonal

$$\boldsymbol{w} = \left(w_{i,j}\right)_{\substack{i=1,2,\dots,N \\ j=1,2,\dots,N}} \in \mathbb{R}^{N \times N}. \qquad \textbf{(3)}$$

The Quantum Annealing searches for a solution $\boldsymbol{x_{min}} \in X$ with minimum energy, i.e.

$$\forall \boldsymbol{x} \in X: E(\boldsymbol{x}) \geq E(\boldsymbol{x_{min}}). \qquad \textbf{(4)}$$

The Digital Annealer receives a QUBO as input and therefore supports exactly the same problems and problem formulation as Quantum Annealers do. It can handle QUBOs defined on up to 8192 bits and the bias and weight values can be represented in 16 to 64 bit integer precision. This allows much bigger models than possible on quantum technologies today. For real time optimization of traffic lights, the technology scales for scenarios with over 800 intersections, which meets operative requirements of metropolitan areas and is far beyond the size of laboratory examples.

TLO: Applying Quadratic Optimization for Traffic Light Control

As outlined in the chapter "Traffic light optimization Concept" a global optimization function is constructed from the results of local simulations under various conditions. This function will be described as a QUBO. It is a hybrid approach, where the simulations and the construction of the QUBO run on classical hardware and the optimization with this function is executed on a Digital Annealer quantum inspired hardware. In this chapter the mathematical model for this QUBO is explained.

For the formulation of the model let L be the set of directed road links and C the set of intersections. Every link connects two intersections, the inbound c_{in} and the outbound c_{out}, so we can write the link $l \in L$ as a tuple $l = (c_{in}, c_{out})$. For every $c \in C$ let P_c be the set of alternative traffic light phase sequences. For every link $l = (c_{in}, c_{out}) \in L$ and every phase sequences $p \in P_{c_{in}}$ and $q \in P_{c_{out}}$ the traffic on the main link l can be predicted by simulating the in and out flowing traffic with respect to the selected programs p and q; the quality of the predicted traffic (e.g. the average speed) is denoted as $Q(p,q)$. The figure below shows an example of a map of regular 4-way intersections. The road links are shown as arrows. The colored links are considered to simulate the traffic on the main link l.

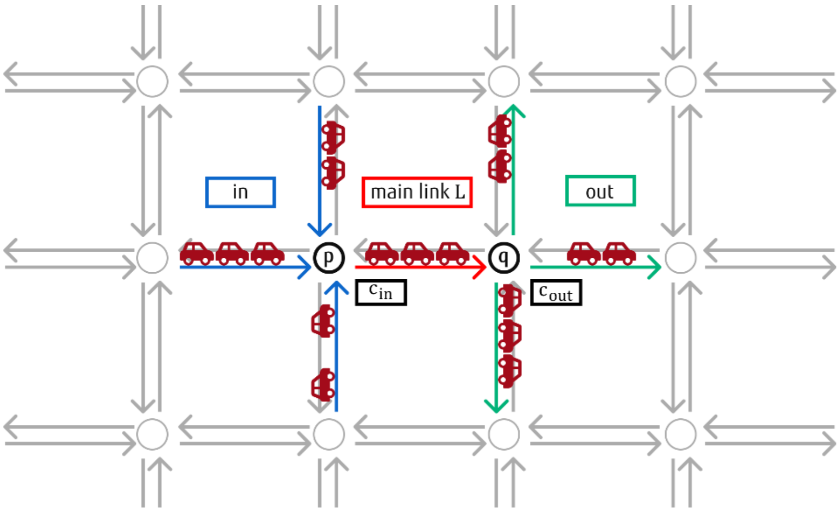

Figure 2 – Sub network for local prediction of traffic light control caused effects

For every intersection $c \in C$ and phase sequence alternative $p \in P_c$ let $x_{c,p}$ be a decision bit, which is 1 if at intersection c the phase sequence p has been selected and 0 if not. For each bit vector $\boldsymbol{x} = \left(x_{c,p}\right)_{c \in C, p \in P_c}$ now the energy function for the optimization problem can be written as

$$E(\boldsymbol{x}) = -\sum_{(c_{in},c_{out}) \in L} \sum_{p \in P_{c_{in}}} \sum_{q \in P_{c_{out}}} Q(p,q)\, x_{c_{in},p} x_{c_{out},q} + \sum_{c \in C} \left(1 - \sum_{p \in P_c} x_{c,p}\right)^2 \quad (5)$$

The first part of (5) is the optimization target, which sums up the qualities for all selected pairs of inbound and outbound phase sequences. The algorithm is designed to find a combination with maximum quality; therefore the negative quality sum is used. The second term of (5) enforces the logical constraint, that for every intersection exactly one phase sequence has to be selected. The minimum of this function is a state, where for each intersection exactly one phase sequence is selected and under all states with that property the negative sum of quality terms $Q(p,q)$ is minimal, i.e. the sum of quality metrics over the complete network reaches a maximum.

Simulation testbed

For testing purposes, this traffic flow optimization system relies on simulated traffic flows to verify the functionality of the TLO. The simulation testbed is based on a microscopic traffic flow simulation model in PTV Vissim to replicate current and optimized traffic flow of the survey area including five signalized intersections. The traffic volume for three time windows was matched with cross sectional counts, which were used to modify origin-destination matrices of a travel demand model. Matrix estimation as implemented in PTV Visum was applied to separately generate vehicle routing information for trucks and passenger cars. As truck traffic experiences several restrictions (e.g. no passing on the Köhlbrand-Bridge) and the variable but large contribution of trucks within the harbor region of up to 50% of the total volume, the distinction of truck traffic is important. Traffic volumes per route and the road network including the geometric layout at each intersection were exported to Vissim. The traffic signals were replicated precisely including pedestrians crossing. As a starting point, the fixed time (pre-timed) signal settings for morning, daytime and evening peak were implemented according to the plans in the regional control center.

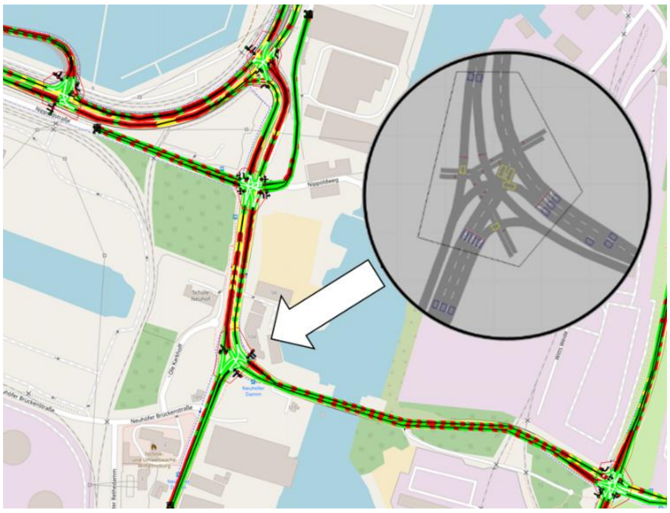

Figure 3 – Visum road network imported to Traffic Light Optimization (TLO)

The microscopic model was calibrated comparing observed and modeled queue length and traffic volume per lane. Especially for truck traffic, the calibration of lane-specific traffic was necessary. The calibrated microscopic traffic model is used as the base case to compare the impact of the optimized signal programs. Queue length, travel time and number of stops are used as performance indicators.
The structural Visum model is imported to the TLO. It is the foundation for fast short-term simulations that are used as input for the global optimization. The figure above is a screenshot of the TLO traffic density map. The grey circle indicates the level of detail as modeled in the microsimulation Vissim. The structural consistency is reached by import from the common Visum source.

Evaluation and results

A part of the harbor road network implemented on Vissim is used as a proven external test bed to compare the status quo against a TLO improved traffic light control. The TLO imports the map structure as foundation for the what-if analysis on local topologies focused on single roads. Roads connect intersections by single or multi-lane paths. Driving dynamic on the links is modeled using the cellular automaton approach of Nagel and Schreckenberg [11]. The aim is to produce a reliable short-term prediction for real world scenarios. For these computations as well as the complete optimization process, there are tough compute performance requirements. The optimization has to complete as fast as possible, so that the result can be applied for a significant part of the prediction span: 60 seconds prediction and the optimization in that period should complete within 10 seconds. This means the first 10 seconds cannot be changed since traffic observations cannot be updated within that time frame, but the next 50 seconds are controlled optimally with respect to the selectable phase sequences. First

results show that these requirements can be fulfilled. The annealing process for the current network of five crossings takes approximately 1 second. It is not expected that this time will grow substantially when more crossings are included in the global optimization. The prediction of 5 or 10 alternatives per crossing takes 3 or 6 seconds. This effort will linearly grow with the number of crossings. Since predictions of disjoint sub networks are independent from each other, the elapsed time can be kept at a level of 5 seconds or even be further reduced by parallelization of the tasks. The compute performance for real time optimization of traffic light control will be achievable also for larger networks with some hundreds of crossings, since the non-divisible overall optimization does fit onto a Quantum-Inspired specialized hardware today and the classical preparation tasks can be parallelized on CPUs or GPUs.
The effect of the optimization was evaluated using the testbed on Vissim. For comparison, the traffic was controlled by periodic programs and movement of all vehicles recorded as baseline. For evaluation first 5 minutes were defined as warm-up phase and controlled by the same periodic programs as in the baseline recording. Then the TLO was activated for 10 minutes, and the optimized phase sequences / durations were imported into the test bed. As before the movement of all vehicles was recorded. Figure 4 shows the increase of average speed relative to the baseline recording.

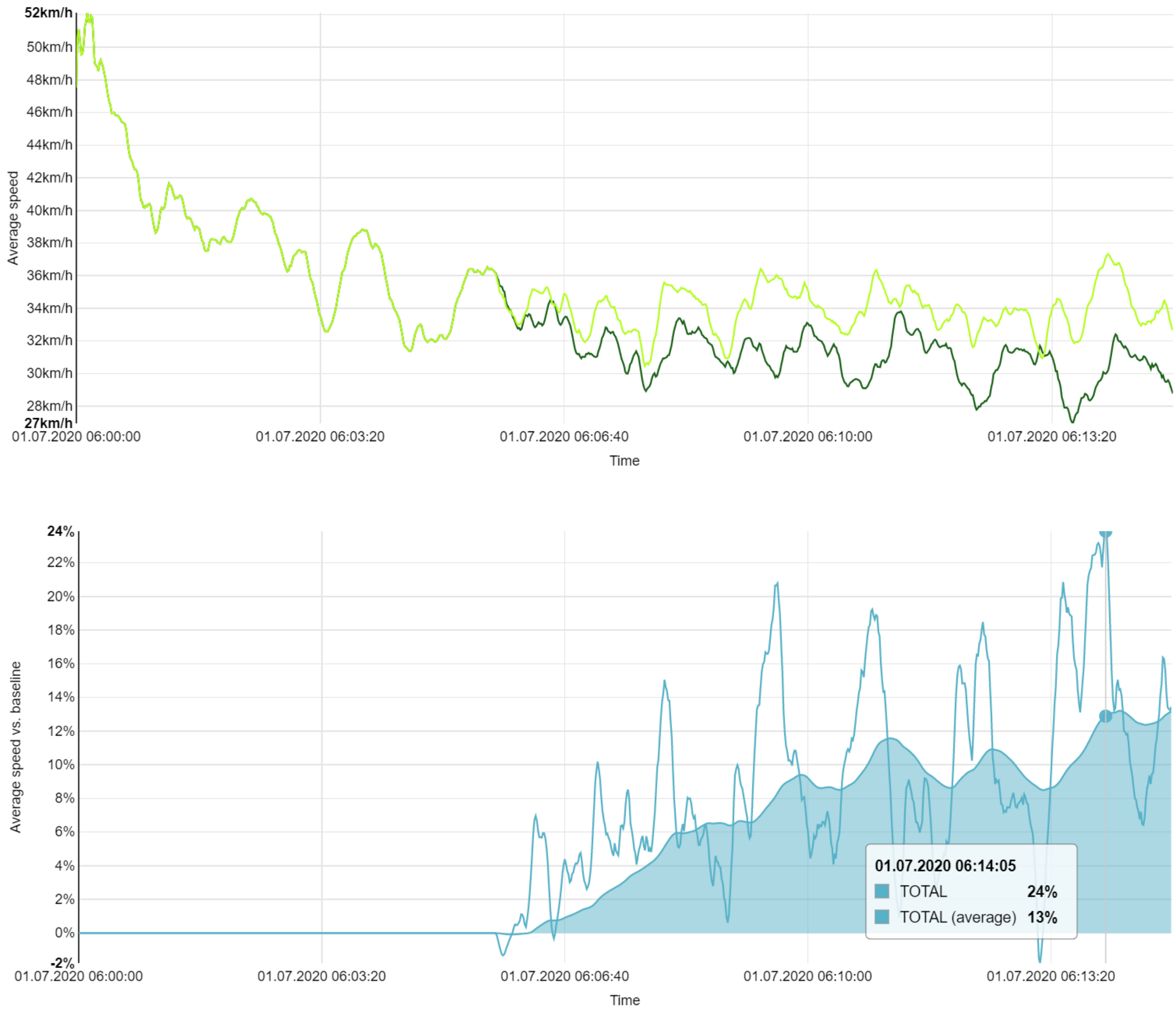

Figure 4 – Average speed of base line vs. optimized

The upper plot shows in the warmup-phase the two identical curves for baseline (black) and optimized (green). Then the baseline degrades while the optimized curve holds the level. In the second chart the relative increase of average speed due to optimization is plotted. The filled curve shows the development smoothed by rolling average over 120 seconds. After switching on the TLO the relative improvement curve steadily grows and reaches a relative improvement of more than 10%. The TLO found excessive green time and could use that for traffic in other directions. The solution managed to keep the density of vehicles and average speed to remain on a constant level, which gave significant advantages compared to the non-optimized baseline.

Conclusion and outlook

The potential of this solution for traffic optimization is proven by our experimental results obtained within the MOZART project. Parallelizable simulation and optimization using Quantum-Inspired annealing for the overall network optimization is possible in a scalable and computationally efficient manner. The first measurements show promising improvements of average speed in heavy load situations; optimization could prevent a degrading of speed which results in significant relative improvements of over 10%. It is expected that lower traffic load will reduce the effects while higher load will even emphasize the benefits of the solution. It is worth to mention, that our current test field is more an arterial constellation with a main route and some low frequent crossing roads. A network with more parallel and higher density intersections can make even more use of the global optimization concept presented in this paper.

The development of the operative QIOS solution for the "Port of Hamburg" and a detailed reference architecture are work-in-progress. The next steps will be to feed the model not only with real historic data, but also with real- time data, directly from the HPA's Port Road Management Centre, and to enable real-time calculation for traffic signal optimization. The aim is to test the data flow in a sub-network simulation and to extend it to other sub-networks of the harbor area based on the results. Additional features, such as green time prioritization for emergency vehicles or public transport are considered in the design as well. Finally, the prototype will stimulate the strategic real-time traffic management in the port.

The project MOZART provides an innovative solution to an urgent traffic management challenge in the harbor area. Nevertheless the TLO and the underlying technologies have the potential to be a defining success factor, not only for ports but also for smart cities in the nearer future – making our cities significantly more livable and sustainable.

References

1. Colwell, R., 2013. *The chip design game at the end of Moore's law*. 2013 IEEE Hot Chips 25 Symposium (HCS), Stanford, CA, p.1-16. https://doi.org/10.1109/HOTCHIPS.2013.7478302.
2. D-Wave, 2011: *D-Wave Systems sells its first Quantum Computing System to Lockheed Martin Corporation*. https://www.dwavesys.com/news/d-wave-systems-sells-its-first-quantum-computing-system-lockheed-martin-corporation.
3. Fellendorf, M., P. Vortisch (2010). Microscopic Traffic Flow Simulator VISSIM. In: Barcelo, J. (ed) Fundamentals of Traffic Simulation. Springer Science + Business Media, 145, p. 64-99. ISBN 978-1-4419-6141-
4. FGSV (German Road and Transport Research Association, 2021) RiLSA Guidelines for Traffic Signals, Köln, www.fgsv-verlag.de/pub/media/pdf/321_E.v.pdf
5. Hunt P., D. Robertson, R. Bretherton, and M. Cr Royle. (1982) *The SCOOT on-line traffic signal optimisation technique*. Traffic Engineering & Control 23, 4
6. Kadowaki, T., H. Nishimori (1998). *Quantum annealing in the transverse Ising model*. Phys. Rev. E 58(5), 5355-5363. https://doi.org/10.1103/PhysRevE.58.5355.
7. Kirkpatrick, S., C.D. Gelatt Jr, M.P. Vecchi, (1983). Optimization by Simulated Annealing. Sci. 220(4598), 671–680. https://doi.org/10.1126/science.220.4598.671.
8. Lowrie P. (1992) SCATS–a traffic responsive method of controlling urban traffic. Roads and traffic authority. NSW, Australia
9. Lucas, A., (2014. Ising formulations of many NP problems. Front. Phys. 2(5), 1-27. https://doi.org/10.3389/fphy.2014.00005.
10. Matsubara, S., M. Takatsu, T. Miyazawa, T. Shibasaki, Y. Watanabe, K. Takemoto, H. Tamura, (2020). Digital Annealer for High-Speed Solving of Combinatorial optimization Problems and Its Applications. 2020 25th Asia and South Pacific Design Automation Conference (ASP-DAC), Beijing, China, pp.667-672. https://doi.org/10.1109/ASP-DAC47756.2020.9045100.
11. Nagel K., M. Schreckenberg (1992). A cellular automaton model for freeway traffic. In: J. Phys. I France, 2, 1992, S. 2221–2229.
12. Tsukamoto, S., M. Takatsu, S. Matsubara, H. Tamura, (2017). An accelerator architecture for combinatorial optimization problems. Fujitsu Sci. Tech. J. 53(5), 8-13.
13. Wei H., G. Zheng, V. Gayah, Z. Li (2020). A survey on traffic control methods. arXiv:1904.08117v3